Using discrete choice experiments to value benefits and risks in primary care

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

2015

Caroline Mary Vass

School of Medicine

Volume I of II

<u>Volume I</u>

Contents

List of Tables
List of Figures10
List of Appendices11
Glossary14
List of Abbreviations
Abstract
Declaration27
Copyright statement
Acknowledgements
The Author
Chapter One
Introduction
1.1 Thesis outline
1.2 Key concepts
1.2.1 Risk in healthcare
1.2.2 Valuing health and healthcare
1.3 Research questions
1.4 Evolution of research
1.5 Study Design
1.6 Methods
1.6.1 Quantitative choice models
1.6.2 Qualitative research methods
1.6.3 Eye-tracking methods
1.7 The case study: breast screening
1.8 Thesis outline
Chapter Two
Why and how to elicit preferences for benefit-risk trade-offs in healthcare?

2.1 Introduction	
2.2 The role of health economics	
2.2.1 The role of public and patient preferences	
2.3 Valuation methods	40
2.3.1 Revealed preferences	40
2.3.2 Stated preference methods	42
2.4 Theoretical foundation of DCEs	49
2.4.1 Lancaster's theory	49
2.4.2 Random utility theory	50
2.5 Risk in healthcare valuation	52
2.5.1 Definition of risk	52
2.5.2 Risk in healthcare	53
2.5.3 Risk in a DCE	53
2.5.4 Decisions with risk	55
2.5.4.1 Expected utility theory	56
2.5.4.2 Risk literacy	57
2.5.5 Communicating risk information	57
2.6 Conclusion	59
Chapter Three	
Valuing risks and benefits: a case study of population-based scr	eening for breast
cancer	60
3.1 Introduction	60
3.2 Risk communication formats	60
3.3 Selecting a case study	61
3.3.1 Case study selection criteria	62
3.3.2 Identification and appraisal of case studies	63
3.4 Screening for breast cancer	67
3.4.1 Background to breast screening	67
3.4.2 Facts, figures and debates	67
3.4.3 Benefit-risks trade-offs in breast screening	71

3.4.4 Healthcare setting	71
3.4.5 Potential outputs of the DCE	71
3.4.6 Existing DCEs relevant to cancer screening: a rapid review.	72
3.5 Breast screening: designing the DCE	77
3.5.1 Research questions	77
3.5.2 Sampling frame	77
3.5.3 Identification of attributes and levels	
3.5.4 Choice format	83
3.5.5 Experimental design	83
3.5.6 Survey materials	86
3.5.7 Ethical approval	
3.5.8 The piloting process	
3.5.9 DCE design: main study	90
3.6 Conclusion	93
Chapter Four	
Benefit-risk trade-offs for breast screening: results of the DCE	94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94 94 94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94 94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94 94 94 95 95 97
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	94
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	
Benefit-risk trade-offs for breast screening: results of the DCE 4.1 Introduction	

4.5 Discussion	121
4.5.1 Key findings	121
4.5.2 Key implications	123
4.5.3 Comparisons with existing literature	124
4.6 Conclusion	127
Chapter Five Benefit-risk trade-offs for breast screening: a qualitative study	
5.1 Introduction	128
5.1.1 Background to qualitative research methods	128
5.2 Qualitative research methods and DCEs	129
5.2.1 Systematic review of qualitative research methods alongside DCEs	129
5.2.2 A survey to authors	131
5.2.3 Discussion of review and survey findings	133
5.3 Research questions	134
5.4 Introduction to qualitative research methods	134
5.5 Methods	136
5.5.1 Qualitative research methods	136
5.5.2 Data generation	141
5.5.3 Process of analysis	142
5.5.4 Qualitative boundaries and analytical limitations	143
5.5.5 Summary of the analysis and methods	144
5.6 Results	145
5.6.1 Interviewees	146
5.6.2 Coding tree	148
5.6.3 Experience of completing a DCE	151
5.6.4 Evidence of economic phenomena	153
5.6.5 Experience of completing a DCE with risk attributes	165
5.7 Discussion	172
5.7.1 Key findings	172
5.7.2 Strengths and limitations	175

5.8 Conclusion	178
Chapter Six	
Benefit-risk trade-offs for breast screening: an eye-tracking study	179
6.1 Introduction	179
6.2 Research questions	179
6.3 Introduction to eye-tracking as a method	180
6.3.1 Eye physiology and key concepts	180
6.3.2 Technology and equipment	181
6.3.3 Psychological foundations	184
6.3.4 Research to date	185
6.4 Hypotheses	
6.5 Methods	
6.5.1 Recruitment	186
6.5.2 Apparatus	187
6.5.3 Data collection	188
6.5.4 Analysis	
6.6 Results	197
6.6.1 Participants	197
6.6.2 Experimental validity	197
6.6.3 Visual attention: fixation and dwell times	200
6.6.4 Attribute non-attendance	201
6.6.5 Saccades	202
6.6.6 Pupil dilation	203
6.7 Discussion	204
6.7.1 Key findings	204
6.7.2 Strengths and limitations	206
6.8 Conclusion	208
Chapter Seven	
Discussion	210
7.1 Introduction	210

7.2 Key results	210
7.3 Limitations	213
7.4 Comparisons with existing literature	217
7.5 Key implications	217
7.6 Future research	219
7.7 Conclusion	220

List of Tables

Table 1.1: Chapter titles, topics covered and methods employed
Table 2.1: Definitions of methods of economic evaluation 39
Table 2.2 Components of risk 53
Table 2.3: Examples of risk in types of healthcare 53
Table 3.1: Potential case studies for a benefit-risk DCE 64
Table 3. 2: Risks to women of breast cancer
Table 3.3: Breast cancer and screening DCEs 74
Table 3.4: Attributes and levels used in pilot DCE
Table 3.5: Attributes and levels used in final DCE
Table 4.1: Attributes, levels and coding
Table 4.2: Dummy variable definitions
Table 4.3: WTP for hypothetical screening programmes
Table 4.4: Sample characteristics
Table 4.5: Validity test failure and self-reported task difficulty
Table 4.6: Results of the conditional logit Models 1A, 1B and 1C108
Table 4.7: Results of the conditional logit Model 1D
Table 4.8: Heteroscedastic conditional logit results 111
Table 4.9: Information criteria for varying preference-class numbers in Model 3112
Table 4.10: Information criteria for varying class numbers in latent-class and scale-
adjusted latent-class models113
Table 4.11: Preference-classes in scale-class one of Model 4
Table 4.12: MWTP for breast screening attributes 117
Table 4.13: WTP for a lifetime breast screening programme 118
Table 4.14: WTA risk to detect an additional cancer
Table 4.15: Probability of participating in a screening programme over 'No Screening'.119
Table 5.1: Schools of thought in research methodology
Table 5.2: Interviewee characteristics 147

Table 6. 1: Outcomes of interest in the eye-tracking data
Table 6. 2: Mean response time in milliseconds from starting the choice set to pressing a
button, and differences between risk formats
Table 6.3: Mean number of fixations to white space per choice set, and differences
between risk formats
Table 6. 4: Visual attention to each attribute in a trial (hypothesis one)201
Table 6.5: The effect of self-reported ANA on total dwell-time per trial in milliseconds
(hypothesis two)
Table 6.6: Number of choice-sets with no attention to an attribute for each risk format202
Table 6.7: Number and direction of saccades in a trial (hypothesis three), and difference
between risk formats
Table 6.8: Measurements (in pixel units) of cognitive pupillometry (hypothesis four)204

List of Figures

Figure 2.1: Preference valuation methods
Figure 3.1: Progression of women through screening70
Figure 3.2: Example choice questions
Figure 4.1: Functional form of the risk attribute for DCE respondents
Figure 4.2: Functional form of the detection attribute for DCE respondents110
Figure 4 3: Functional form of the cost attribute for DCE respondents110
Figure 4.4: Relative importance of attributes by preference-class115
Figure 4.5: The effect of increasing risk on uptake for breast screening by preference-class
Figure 4.6: The effect of increasing detection rates on uptake for breast screening by
preference-class
Figure 4. 7: The effect of increasing cost on uptake for breast screening by preference-class
Figure 5.1: Conversational qualitative research methods
Figure 5.2: Qualitative data collection processes145
Figure 5.3: Intermediate coding tree demonstrating identified codes and themes149
Figure 5.4: Final coding tree identifying key themes and inter-theme relationships150
Figure 6.1: Measurement of visual degrees
Figure 6.2: Configuration of the eye-tracker
Figure 6.3: Pre-defined AOI for each choice set
Figure 6.4: Definition of direction from saccade angles191
Figure 6.5: Example of eye-tracking fixation data193
Figure 6.6: Example of collapsed fixation data to key AOI194
Figure 6.7: Example scan path from a respondent completing the DCE with icon arrays and
percentages
Figure 6. 8: Example scan path from a respondent completing the DCE with percentages
only199

<u>Volume II</u>

List of Appendices

Appendix 3.1: How to frame risk in a DCE: an exploration and structured review246
Appendix 3.2: List of studies included in the structured review of risk communication
formats
Appendix 3.3: The published paper containing the results of the structured review274
Appendix 3.4: List of DCEs with a focus on breast cancer and screening from the
systematic review
Appendix 3.5: The design of the DCE used in the pilot study297
Appendix 3.6: Review of questions used to elicit individual characteristics in commonly-
used surveys
Appendix 3.7: Confirmation letter of University Research Ethics approval
Appendix 3.8: The DCE survey used for the pilot study
Appendix 3.9: Online advertisement for study participants
Appendix 3.10: Preliminary interview schedule for pilot study
Appendix 3.11: Information sheet for interviewees for the qualitative interview study341
Appendix 3.12: Consent form for interviewees for the qualitative interview study
Appendix 3.13: NHS Choices video approval
Appendix 3.14: Pilot quantitative results of the DCE from the internet panel
Appendix 3.15: The design of the DCE used in the final empirical studies
Appendix 3.16: The DCE survey used for the final empirical studies
Appendix 3.17: Quantitative results of the DCE from the internet panel in the second pilot
study
Appendix 4.1: ResearchNow® response rate in the large internet panel DCE 388
Appendix 4.2: Sample characteristics of respondents who completed the DCE in the final
study
Appendix 4.3: Results of a simple logit regression to determine if any covariates were a
significant predictor of risk communication format

Appendix 4.4: Latent class analysis results from 6-preference-class (no scale-classes)
model for the large internet panel DCE
Appendix 5.1: Systematic review of qualitative research alongside healthcare DCEs395
Appendix 5.2: Survey to authors of published healthcare DCEs420
Appendix 5.3: Flow of papers through the systematic review
Appendix 5.4: Comparison of existing appraisal tools for assessing the quality of
qualitative research
Appendix 5.5: Final critical appraisal tool for assessing the quality of qualitative research
alongside DCEs435
Appendix 5.6: Studies included in the systematic review of qualitative research conducted
alongside DCEs436
Appendix 5.7: Tabulated results from the systematic review of DCEs453
Appendix 5.8: List of studies not included in the Clark et al. (2014) systematic review455
Appendix 5.9: The submitted paper describing the development of the appraisal tool456
Appendix 5.10: The questionnaire sent to authors of DCE studies to assess their experience
and views of using qualitative research methods471
Appendix 5.11: The invitation to authors of DCE studies to complete the questionnaire to
assess their experience and views of qualitative research methods alongside DCEs478
Appendix 5.12: The reminder invitation to authors of DCE studies to complete the
questionnaire to assess their experience and views of qualitative research methods
alongside DCEs
Appendix 5.13: Author responses to the questionnaire of their experience and views of
conducting qualitative research alongside DCEs
Appendix 5.14: Paper advertisement to the general public for participation in the think-
aloud interviews
Appendix 5.15: Email advertisements to The University of Manchester staff for
participation in the think-aloud interviews

Appendix 5.16: Think-aloud interview schedule
Appendix 5.17: Flow of recruited study participants for the think-aloud interviews and the
eye-tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye-
tracking study
Appendix 6.2: Email advertisement to The University of Manchester staff for participation
in the eye-tracking study
Appendix 6.3: Information sheet for the eye-tracking study participants
Appendix 6.4: A description of the eye-tracking pilot study
Appendix 6.5: Consent form for participants in the eye-tracking study
Appendix 6.6: iPad survey for the eye-tracking study
Appendix 6.7: Heteroscedastic conditional logit results for the eye-tracking study

<u>Glossary</u>

Term	Definition
Alternative specific constant	A regression term representing the mean of the distribution of the unobserved effects in the random component of utility.
Anchoring effect	A phenomenon seen in surveys where respondents base their answers on previous responses or on the initial information provided.
Area of interest	Defined segments of a stimulus which can be sub-analysed by the researcher when using eye-tracking.
Attribute non-attendance	When respondents ignore the information contained in one attribute, violating the assumption of continuity of preferences which is key to the validation of discrete choice experiment data.
Attributes	The properties or characteristics of the goods from which utility may be derived and included in a discrete choice experiment.
Averting behaviour	Actions taken to reduce risk exposure.
Best-worst scaling	A type of choice experiment where respondents choose the option (alternative, attributes or levels) which yield the highest (best) and the lowest (worst) utility, revealing the components which are furthest apart on their individual utility spectrum.
Cognitive pupillometry	A hypothesised relationship between the size of the pupil of the eye of a respondent and their cognitive activity.
Choice set	The alternatives presented to the respondent from which they choose in a discrete choice experiment.
Coding tree	An arrangement of codes demonstrating links and potential relationships through branches from qualitative data.

Conditional logit model	A regression model with a categorical dependent variable where the values of the variables (usually choice characteristics) vary across the choices but parameters are common across the choices.
Conjoint analysis	A type of stated preference method where respondents are asked to order or score alternatives according to their preferences.
Conjoint analysis - adaptive	A type of stated preference method when the design is continuously updated so the choices presented to respondents depend on their previous responses.
Contingent valuation	A stated preference method used to elicit willingness-to- pay (or accept) values through a direct question.
Corneal reflection	A throwback of light from the eye which can be compared to the centre of the pupil to give an approximation of gaze.
Cost-benefit analysis	An assessment of the costs and the benefits of a health intervention where the benefits are measured in monetary terms.
Cost-effectiveness analysis	An assessment of the costs and the benefits of a health intervention where the benefits are measured in a clinical or health-related metric.
Cost-utility analysis	An assessment of the costs and the benefits of a health intervention where the benefits are measured in a quality- adjusted outcome, such as quality-adjusted life years.
Data reduction	The process of drawing meaning from large sets of qualitative data.
Data saturation	A principle point reached in the collection of data (often qualitative) where additional data does not contribute anything new.

D-efficiency	A type of discrete choice experiment design criteria with minimal D-error satisfying, or attempting to satisfy, orthogonality, level balance, minimal overlap, and utility balance.
Design efficiency	Design efficiency refers to a combination of attributes within the choice sets in a discrete choice experiment, so that the maximum amount of information is obtained from respondents.
Drift	In eye-tracking, a gradual misalignment from a correct calibration from adjustment in the head position or a miscalculation by the tracker.
Drift correction	A validation procedure with an eye-tracking experiment to confirm calibration.
Dummy coding	A variable that takes the value of 0 or 1 to indicate the absence or presence of a category compared to a base case (the dropped variable) in a regression model.
Dwell time	The total time of all fixations (see definition below) of an eye to an area in an eye-tracking study.
Effects coding	Alternative approach to dummy variables that allow for estimation of the effects of the dropped variable.
Epistemology	A branch of philosophy concerned with knowledge and its limitations.
Ethnography	An approach, method or analysis in qualitative research that involves observing the subject.
Experimental design	A sample from all possible combinations of attribute levels used to construct choice alternatives in a discrete choice experiment.
Extra-welfarist	An evaluative framework where something other than, or in addition to, utility is maximised (for example, health).

EyeLink®	A brand of eye-tracking device used by research
	institutions due to their high recording speeds.
Familiar good effect	When the valuation of a good or service is influenced positively because it is familiar to the respondents.
Fixation	A period when the eye is 'relatively' still and information is processed. Typically fixations in eye-tracking are thought to last over 75 milliseconds.
Fractional factorial design	A sample from a full factorial design which can estimate effects of interest through interactions in a discrete choice experiment.
Framework analysis	An approach, method or analytical technique that uses a pre-defined set of topics to identify codes or themes from qualitative data.
Full factorial design	A design using the complete set of all attribute and level combinations in the discrete choice experiment.
Gaze	The location of focus of an eye.
Generic set	Where the labels of alternatives convey no information beyond that provided by the attributes and levels.
Heteroscedastic conditional logit model	A conditional logit regression model accounting for differences in the variance of the error term through estimation of the scale parameter.
Icon array	Also known as a risk grid, icon array is a risk communication format using individual graphics to reflect proportions most commonly out of 10, 100 or 1000.
Independence of Irrelevant Alternatives	When adding another alternative or changing the characteristics of a third alternative does not affect the relative odds between the two original alternatives in a discrete choice experiment.

Information asymmetry	A situation where a consumer has incomplete information
	assess the utility associated from a good or service.
Interpretism	A subjective understanding of phenomena which are
	deviant or unique.
Lancaster's theory	A hypothesis which suggests that individuals do not value
	a good or service per se, but instead value the
	characteristics or attributes of which it is made.
LatentGold®	A latent-class and finite mixture software programme used
	to estimate complex regression models.
Latent-class analysis	A regression modelling technique which identifies subsets
	of respondents with similar preferences.
Level balance	When the design of a discrete choice experiment results in
	the level of each attribute occurring with equal frequency
	in survey choice sets.
Lexicographic preferences	When the good providing the most of X is always
	preferred, no matter what the amount of Y.
Logit model	A probability regression model which take a logistic
	distribution as opposed to a normal distribution (which
	would be a probit).
Mammography	The use of X-rays of the breast(s) to create an image that
	can be used to diagnose cancer.
Marginal rates of	The willingness to exchange a unit of one good for another
substitution	to maintain the same level of utility.
Mixed logit model	A regression model which allows for random taste
	variation, unrestricted substitution patterns, and correlation
	in unobserved factors utilising any distribution for the
	random coefficients.

Monotonicity	If A is preferred to B then A should be preferred to C if C is as good an alternative as B in all attributes.
Multinomial logit model	A conditional logit regression model where the values of the variables are common across choices for the same person (usually individual characteristics) but the parameters vary across choices.
Ngene®	A software package used to efficiently design and test the statistical properties of discrete choice experiments.
Non-use values	Utility derived from a good or service that is not related to current consumption.
Nonveridicality	When a study task or experiment is not a true reflection of real life behaviour.
NVivo®	A software package used in the collection, generation and analysis of qualitative data.
Ontological	A philosophy of study of the nature of being.
Open-coding	A process of identifying patterns in the qualitative data based on no prior ideas or assumptions.
Opportunity cost	The cost of an alternative that is forgone in making a particular choice within a fixed budget.
Optimism bias	A cognitive belief where an individual perceives themselves to be at less risk.
Ordering effect	A phenomenon which occurs when changing the arrangement of questions in a survey affects responses.
Orthogonality	A property of discrete choice experiments to remove collinearity and estimate level effects independently.
Payment vehicle effect	A phenomenon which occurs when describing a cost differently affects the values given by respondents – for example, the discrepancies between a 'tax' or a 'donation'.

Positivism	An objective assessment of measurable absolutes.
Primacy effect	A response order effect where the consequence of placing a single question or choice set first. This may be because respondents use their answer to this as an anchor for responses to later sets in a discrete choice experiment.
Profile	A combination of attributes at varying levels which make up an option in a choice set of a discrete choice experiment.
Protest bids	Responses given in stated preference methods to express an opinion or to change the results of the study.
Pupillometry	Measure of the size of the pupil of the eye.
Qualitative research methods	Exploratory research to understand motivations and generate ideas or hypotheses.
Random utility theory	A choice theory where decisions are deterministic and utility has a random component.
Reactivity	Challenges for respondents to verbalise their thoughts in a cognitive interview setting.
Reflexivity (qualitative research)	Acknowledging the effect of a researcher on qualitative data collection.
Researcher bias	When social norms or the presence of someone means true thoughts/feelings are not revealed.
ResearchNow®	An internet panel provider who recruits samples of respondents to complete online surveys.
Revealed preference	Data collected through observations of behaviour in real markets.
Risk	The probability of a hazard (usually a negative event) occurring.
Romanticism	Influence of society and culture on study results.

Saccades	Very rapid eye movements shifting gaze to acquire information.
Sawtooth®	A software package, used for hosting online surveys, which can also be used to design and analyse choice experiments.
Scale-adjusted latent-class analysis	An extension of the latent-class regression model which identifies subsets of respondents with similar 'randomness' within each preference class.
Scale parameter	An inverse of the error variance which is a measure of systematic randomness incorporated in choice data generated from different sources. It can also be thought of as a measure of choice consistency.
Stability (preferences)	If A is preferred to B at one point, A should still be preferred to B later.
Starting-point bias	A phenomenon in contingent valuation where respondents are influenced by the first question or bid.
STATA®	A software package used for the statistical analysis of quantitative data.
Stated preference	Data collected through surveying individuals to attain information about how they would behave in the hypothetical scenario presented.
Stimulus	In eye-tracking, this is the object that the participant is likely to react to or see.
Structuring	Grouping themes in qualitative analysis to make sense of the data.
Trial	A test in an experiment. In the context of eye-tracking, each trial refers to each recorded screen an individual sees.
Thematic analysis	The generation of clusters of codes from qualitative data related to similar topics.

Think-aloud	Technique used in qualitative research and cognitive
(verbal protocol analysis)	interviewing to attain information on individuals' thought processes. Individuals are asked to verbalise their thoughts
	as they complete a task.
Trade-off	A compromise which minimises the opportunity cost to maximise benefit or utility.
Transferability	A qualitative measure of generalisability.
Transitivity (preferences)	If A is preferred to B, and B to C, then A is preferred to C.
Uncertainty	Where the probability of an event occurring is unknown
	because it is ambiguous or because there are surrounding confidence intervals.
Utility	Term used in economics to describe the satisfaction gained
	from the consumption of goods or services.
Utility balance	A design criterion satisfied when the estimated utility of each alternative within a choice set in a discrete choice experiment is the same.
Visual attention	Eye-movements which result in fixations that indicate an object is being considered by an individual.
Visual degrees	A measure of distance in eye-tracking which allows for varying proximity between the object and pupil.
Warm-glow theory	Utility acquired from altruistic or charitable behaviour.
Welfarist	An evaluative framework that maximises total utility for society.
Yea-say bias	A tendency to agree with what was being asked (either in a survey or interview).
Zero bids	A phenomenon present in contingent valuation where a bid of nil is given for something the respondent does actually value.

List of Abbreviations

Abbreviation	Term
ACA	Adaptive Conjoint Analysis
AIC	Akaike Information Criterion
AMD	Age-Related Macular Degeneration
ANA	Attribute Non-Attendance
AOI	Areas of Interest
ASC	Alternative Specific Constant
BIC	Bayesian Information Criterion
BMA	British Medical Association
BMI	Body Mass Index
BWS	Best-Worst Scaling
CAIC	Consistent Akaike Information Criterion
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CJA	Conjoint Analysis
CRD	Centre for Reviews and Dissemination
CSV	Comma-Separated Value
CUA	Cost-Utility Analysis
DC	Dummy Coding
DCE	Discrete Choice Experiment
DT	Dwell Time
EC	Effects Coding

EQ-5D	EuroQol Five Dimension
EUT	Expected Utility Theory
GP	General Practitioner
Hz	Hertz
IAP	Icon Arrays and Percentages
ID	Identity
IIA	Independence of Irrelevant Alternatives
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
LL	Log Likelihood
LR	Likelihood Ratio
MF	Mean Fixation Duration
MLE	Maximum Likelihood Estimation
MNL	Multinomial Logit
MRS	Marginal Rate of Substitution
ms	Millisecond
MWTP	Marginal Willingness-to-Pay
MXL	Mixed Logit
NHS	National Health Service
NHSBSP	National Health Service Breast Screening Programme
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NNT	Number Needed to Treat

NOF	Number of Fixations
OLS	Ordinary Least Squares
ONS	Office for National Statistics
Pc	Preference-class
РО	Percentages Only
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality-Adjusted Life Year
RCT	Randomised Controlled Trial
RT	Response Time
RUT	Random Utility Theory
Sc	Scale-class
SPCR	School for Primary Care Research
UK	United Kingdom
USA	United Sates of America
VAS	Visual Analogue Scale
WTA	Willingness-to-Accept
WTP	Willingness-to-Pay

Abstract

The University of Manchester

Caroline Mary Vass

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

Using discrete choice experiments to value benefits and risks in primary care

2015

Discrete choice experiments (DCEs) are a stated preference valuation method. As a ubiquitous component of healthcare delivery, risk is increasingly used as an attribute in DCEs. Risk is a complex concept that is open to misinterpretation; potentially undermining the robustness of DCEs as a valuation method. This thesis employed quantitative, qualitative and eye-tracking methods to understand if and how risk communication formats affected individuals' choices when completing a DCE and the valuations derived.

This thesis used a case study focussing on the elicitation of women's preferences for a national breast screening programme. Breast screening was chosen because of its relevance to primary care and potential contribution to the ongoing debate about the benefits and harms of mammograms. A DCE containing three attributes (probability of detecting a cancer; risk of unnecessary follow-up; and cost of screening) was designed. Women were randomised to one of two risk communication formats: i) percentages only; or ii) icon arrays and percentages (identified from a structured review of risk communication literature in health).

Traditional quantitative analysis of the discrete choices made by 1,000 women recruited via an internet panel revealed the risk communication format made no difference in terms of either preferences or the consistency of choices. However, latent class analysis indicated that women's preferences for breast screening were highly heterogeneous; with some women acquiring large non-health benefits from screening, regardless of the risks, and others expressing complete intolerance for unnecessary follow-ups, regardless of the benefits. The think-aloud method, identified as a potential method from a systematic review of qualitative research alongside DCEs, was used to reveal more about DCE respondents' decision-making. Nineteen face-to-face cognitive interviews identified that respondents felt more engaged with the task when risk was presented with an additional icon array. Eye-tracking methods were used to understand respondents' choice making behaviour and attention to attributes. The method was successfully used alongside a DCE and provided valid data. The results of the eye-tracking study found attributes were visually attended to by respondents most of the time.

For researchers seeking to use DCEs for eliciting individuals' preferences for benefit-risk trade-offs, respondents were more receptive to risk communicated via an icon array suggesting this format is preferable. Policy-makers should acknowledge preference heterogeneity, and its drivers, in their appraisal of the benefits of breast screening programmes. Future research is required to test alternative risk communication formats and explore the robustness of eye-tracking and qualitative research methods alongside DCEs.

Declaration

This dissertation is entirely the result of my own research. No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

- The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy

(seehttp://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see

http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Acknowledgements

I would first like to express my gratitude to my supervisory team: Katherine Payne, Dan Rigby and Stephen Campbell for their intellectual support and guidance throughout this PhD. I would like to thank Dan for his assistance with the development of my quantitative skills. I would particularly like to thank Katherine for her extensive feedback, constructive criticism, patience, time and moral support whilst completing this thesis.

I would like to thank the National Institute for Health Research School for Primary Care Research for providing the funding for this PhD. The monies gave me the opportunity to attend conferences to acquire feedback on my work and develop networks within the field of health economics. Funding for training and courses was key in developing the mixed – methods skills required to complete this thesis.

I would also like to thank my research participants for taking time out of their lives to provide data for this thesis.

I would like to thank my family: Mum, Dad, Pete, Tom, Olive and EJ for only being a phone call away, and cheering me up constantly.

I would finally like to thank Alex Turner for being there for me whenever I needed him, and keeping me fed, watered and loved.

The Author

The author, Caroline, was awarded a Bachelor of Science in Economics degree (Class 2 Division I (Hons)) from the University of Birmingham in July 2010. She was awarded a Master of Science in Health Economics and Health Policy (Merit) also from the University of Birmingham in December 2011.

In October 2011, she was awarded a National Institute for Health Research School for Primary Care Research doctoral trainee studentship to complete this PhD at the Manchester Centre for Health Economics, The University of Manchester. In October 2014, Caroline won the Lee B Lusted Student Prize at the 36th Annual Meeting of the Society for Medical Decision Making after presenting pilot work conducted in this PhD thesis.

<u>Chapter One</u> Introduction

1.1 Thesis outline

The overall aim of this thesis was to identify and quantify benefit-risk trade-offs when eliciting preferences for healthcare interventions with discrete choice experiments (DCEs). This PhD used a combination of quantitative (DCEs), qualitative (think-aloud) and eye-tracking methods to understand individuals' preferences for benefit-risk trade-offs. A 'benefit-risk trade-off' was defined as the balance between a gain (to health) and the probability of incurring harm. The primary outcome of interest was how DCE respondents' choices and decision strategies were affected by the format of the risk attributes.

This chapter introduces the broad topic of risk in healthcare and briefly describes the selected valuation method, DCEs, before outlining the research questions in sections 1.2 and 1.3, respectively. The evolution of the research programme for this thesis is explained in section 1.4 and in section 1.5 the studies used in the thesis are outlined. The methods used in the thesis are introduced in section 1.6. In section 1.7, the case study of the United Kingdom's (UK) National Health Service (NHS) breast screening programme (NHSBSP) is described. Section 1.8 outlines the structure of the thesis.

1.2 Key concepts

This section introduces the key terms and concepts used in this PhD. All definitions are explained in more detail in subsequent chapters. In section 1.2.1, risk and its role in healthcare is briefly introduced. In section 1.2.2, the valuation method used in the thesis, DCEs, is briefly described.

1.2.1 Risk in healthcare

Risk is the probability of an event (often negative) occurring. Most choices in life involve an element of risk; the probability of a traffic accident on the way to work, or risking food poisoning for a 'cheap-eat'. Similarly, choices in healthcare often involve an element of risk; this could be in the form of a side-effect from an interaction or the risk of misdiagnosis. Many of the choices made in real-life are carried out subconsciously without explicit statements of the probability of particular events occurring. For this, and other reasons, it is well established that people find the concept of risk and probability difficult to understand (Fagerlin et al., 2007). It is this problem that marks the starting point for this thesis. It is known that risk is a poorly understood concept by many individuals (Lipkus et al., 2001), but this thesis is particularly concerned with how this impacts on the valuations of healthcare goods and services derived by health economists using DCEs.

1.2.2 Valuing health and healthcare

Health economics, in general, aims to use economic theories and methods to inform the efficient allocation of finite healthcare resources (Drummond & McGuire, 2001). Therefore decision-makers need to consider which intervention or service generates the most benefit for populations for a given cost. DCEs are an increasingly popular valuation method used to measure the benefit of a healthcare good or service (de Bekker-Grob et al., 2012; Clark et al., 2014). In a DCE, individuals are presented with a survey comprising hypothetical goods or services, described by their attributes, and asked to select their preferred alternative in a series of choice sets. The valuation method estimates the marginal rates of substitution (see glossary for a definition) from the choices made, and can, therefore, be useful in eliciting people's preferences for benefit-risk trade-offs (Johnson et al., 2010; Hauber et al., 2010).

1.3 Research questions

Four research questions are specifically related to the stated aim of this thesis (see p.31):

- 1. To what extent is qualitative research reported in health DCEs and how useful are these research methods?
- 2. What is the most appropriate risk communication format to use in public surveys?
- 3. How sensitive are valuations from a DCE to the methods used in communicating risk?
- 4. How effective are alternatives to qualitative methods in evaluating respondents' understanding of attributes and their levels?

The overall aim and research questions were developed in response to the increased use of DCEs for eliciting preferences for benefit-risk trade-offs. The research questions were devised to robustly investigate the overall aim by utilising, and assessing the usefulness, of different research methods.

1.4 Evolution of research

The approaches and methods conducted as part of this PhD were developed concurrently with the research. From the outset it was unclear which methods would be appropriate for understanding the effects of alternative risk formats on choice behaviour in a DCE. During the course of this PhD, the research questions and research methods evolved substantially. Whilst exploratory reviews were conducted in the early phases of the PhD, a subsequent systematic review of DCEs informed the appropriate qualitative methods and analytical perspective, in addition to aiding the selection of an original application of the DCE task (breast screening), and identifying a format for the framing of risk attributes.

1.5 Study Design

To investigate the impact of the framing of risk attributes in a DCE, the thesis was split into two components: 1) a synthesis of the relevant literatures; and 2) empirical experiments. The first component involved: 1) a literature review of valuation methods in health economics and economics more generally; 2) a systematic review of published DCEs with a particular focus on the qualitative methods employed by researchers; and 3) a structured review of risk communication formats with the aim of understanding the goldstandard of risk communication formats. These literature reviews were combined to inform the case study and DCE design. Three empirical experiments were then conducted: 1) a large sample DCE with discrete choice modelling and quantitative interpretation of the results; 2) a think-aloud study involving face-to-face interviews with DCE respondents and qualitative data analysis; and 3) an eye-tracking study where DCE respondents' choice making behaviour was measured through their visual attention and attendance. These three empirical pieces of work were compared with one another and then bought together in a discussion of the implications of their collective findings (see Chapter Seven).

1.6 Methods

This PhD used mixed methods to answer the four research questions (described in section 1.3). The specific objectives for each component of the study are stated at the start of the relevant chapter. To achieve the stated objectives, different research methods were employed to provide a mechanism for subsequent triangulation of the results. The three broad methods used were: 1) estimation of random utility choice models to derive marginal utilities and the marginal rates of substitution between different attributes; 2) qualitative research methods (think-aloud) to acquire a self-reported account of the cognitive processes involved in the completion of a DCE; and 3) eye-tracking methods to monitor respondents' attention and attendance to attributes in choice sets in the DCE.

1.6.1 Quantitative choice models

It is generally assumed that DCE respondents make choices which would maximise their utility. Random utility theory (RUT) is the economic theory which supports this assumption, but also allows for a random component which could be due to these measurement errors, latent attributes or unobservable taste heterogeneity. It is impossible to say with certainty which alternative a respondent will choose, therefore, probabilistic discrete choice modelling is used to explain the uncertainty around predicting respondents' choices. People's preferences for healthcare have been shown to be highly heterogeneous (Hole, 2008). Latent-class models aim to estimate groups of respondents with similar decision rules, which may provide a more useful interpretation for decision-makers (Louviere 2006). The empirical quantitative chapter (see Chapter Four) of the thesis estimates models with different numbers of latent-classes.

1.6.2 Qualitative research methods

In general, health economists continue to use the quantitative methods used in mainstream economics. This is despite calls for the use of qualitative research methods in health economics (Coast, 1999; Coast et al., 2004). The potential usefulness of qualitative research methods, particularly in the area of DCEs, has been acknowledged (Kløjgaard et al., 2012; Coast & Horrocks, 2007; Coast et al., 2012) and is somewhat incorporated in guidelines for the design, conduct and interpretation of DCEs (Lancsar & Louviere, 2008; Bridges et al., 2011a). In general, the focus in the literature to date has been to use qualitative research methods to design the DCE. In contrast, this PhD seeks to use qualitative research methods to understand more about how a DCE was completed and how the formatting of risk attributes may aid, or hinder, choice making. The qualitative empirical study in this thesis (see Chapter Five) used cognitive 'think-aloud' interviews (a form of verbal protocol analysis). The qualitative data were then analysed using an approach derived from a form of open-coding and framework analysis.

1.6.3 Eye-tracking methods

For many years researchers in psychology have used eye-tracking to investigate cognitive processes (Rayner, 1998). The method is well established in psychology and there have also been published studies in the fields of marketing and consumer choice (Bialkova & van Trijp, 2011). The method is supported by the 'eye-mind hypothesis' suggested by Just & Carpenter (1980) which provides the underpinning to most psychological analyses of eye-tracking data. However, there are limited examples of eye-tracking being used alongside DCEs. In this thesis, the use of eye-tracking as a method was exploratory and aimed to understand if the technique and results could be potentially useful to interpret choice making and impact of the formatting of risk in DCEs.

1.7 The case study: breast screening

The case study used in this thesis involved valuing preferences for a national breast screening programme. Screening for breast cancer using mammograms has been proven to detect cases of breast cancer earlier (NHS Information Centre Screening and Immunisations 2012). Women who participate in screening programmes have been shown to have improved mortality rates because of earlier subsequent intervention (Independent UK Panel on Breast Cancer Screening, 2012). However, there is a risk that a mammogram will result in a woman being recalled for biopsies and unnecessary tests because of a false-positive (Welch & Black, 2010). The unnecessary follow-ups are not only painful; an incorrect test result can cause undue worry for the woman and her family (Johnston et al., 1998).

In England, the NHSBSP invites all women to be screened between the ages of 50 to 70 years. The breast screening programme prevents an estimated 1500 deaths a year and costs an estimated £96 million annually (Pharoah et al., 2013; Public Health England, 2013a). However, the benefits of breast screening, given the risks of unnecessary follow-up and the over-diagnosis of slow-growing cancer, have been extensively debated in leading journals such as the British Medical Journal and The Lancet (Gøtzsche & Nielsen, 2009; Gøtzsche & Olsen, 2000; Baum, 2013). As a consequence of the debate, a complete review of the UK's screening policy was undertaken in 2010 by the Independent UK Panel on Breast Cancer Screening who published their report in 2012. Quantification of women's preferences for breast screening could make a useful contribution to the discussion of the potential balance between risks and benefits.

1.8 Thesis outline

This chapter has introduced the key concepts and methods used in the thesis. The subsequent chapters in this thesis present reviews, methods (a DCE for breast screening), and empirical studies. An outline of the topics and methods used in each chapter of the thesis are described in Table 1.1. In addition, a glossary and a list of abbreviations are presented at the start of the thesis to aid understanding of the economic, clinical and psychological terminology.

Chapter	Key Methods	Title
Two	Literature review	Why and how to elicit preferences for benefit-risk trade-offs in healthcare
Three	Efficient experimental design; Pilot interviews	Valuing risks and benefits: a case study of population- based screening for breast cancer
Four	Discrete choice modelling	Benefit-risk trade-offs for breast screening: results of the DCE
Five	Qualitative interviews	Benefit-risk trade-offs for breast screening: a qualitative study
Six	Eye-tracking trials	Benefit-risk trade-offs for breast screening: an eye- tracking study
Seven	Assimilation of findings from reviews and empirical studies	Discussion

Table 1.1:	Chapter	titles,	topics	covered	and	methods	emple	oyed
		,					· •	

The thesis starts with a general literature review presented in Chapter Two, which introduces the underlying theories, concepts and methods in economic valuation. Whilst focussed on health, this chapter draws upon examples and research from environmental, transport and food economics. The chapter describes why DCEs are an appropriate valuation method for eliciting preferences for benefit-risk trade-offs, but the challenges of communicating risk in a choice setting.

Chapter Three explains why, and how, breast screening was selected as a case study for this PhD. The chapter reviews existing screening DCEs identified by the systematic review. Chapter Five also describes the development and design of the DCE survey, including the identification of the two risk communication formats via a structured review. The iterative stages of piloting the survey with patients, experts, and the public are also explained. The chapter finishes by describing the final DCE used in the subsequent empirical chapters (see Chapters Four, Five and Six) of this thesis. Two versions of the DCE were created: version 1) risk attributes communicated as percentages only; and version 2) risk attributes communicated as percentages and as an icon array. These two versions were used as the basis for comparing the impact of different risk communication formats in a DCE.

Chapter Four details a study of a large sample online DCE which was administered to female members of the public recruited via an internet panel. The DCE designed in Chapter Three was completed by a sample of women randomised to one of the two risk communication formats (risk communicated with percentages only; or with percentages and an icon array). The chapter explains the econometric specification of discrete choice models and the final analytical approach taken. The quantitative results of the large DCE study are presented in terms of the women's benefit-risk trade-offs and willingness-to-pay (WTP).

Chapter Five presents a qualitative study using the think-aloud method. The 'think-aloud' method was identified from a systematic review of qualitative research methods alongside DCEs and a survey to the authors of these DCEs to understand the usefulness of such methods. The DCE designed in Chapter Three was fielded in a sub-sample of women randomised to one of the two risk communication formats (risk communicated with percentages only; or with percentages and an icon array). The results of the think-aloud interviews are presented with verbatim quotes from the interviewees.

Chapter Six describes eye-tracking as a method and explains how it could be useful in understanding DCE respondents' decision strategies. The chapter reports the methods and
results of an eye-tracking study which used the DCE designed in Chapter Three with a further sub-sample of women randomised to one of the two risk communication formats (risk communicated with percentages only; or with percentages and an icon array).

An overview of the findings of the empirical research (conducted in Chapters Four, Five and Six) are summarised and discussed in Chapter Seven. This chapter also considers the implications of the findings of this research for other DCE practitioners before suggesting future research topics and potential methods to use.

This chapter has briefly outlined and explained the rationale behind the research conducted in the thesis. The next chapter, Chapter Two, builds upon the key ideas introduced here and describes the foundations of DCEs within health economics decision-making and economic theory.

Chapter Two

Why and how to elicit preferences for benefit-risk trade-offs in healthcare?

2.1 Introduction

The aims of Chapter Two are to describe: 1) why it is important to elicit preferences for healthcare interventions; 2) the methods available to elicit preferences; and 3) the challenges of using the stated preference method, DCEs, for valuing benefit-risk trade-offs. The Chapter presents an overview of the relevant literature to address these aims. Formal systematic review methods were not used but the identified literature provides an extensive overview of the current evidence on applied research and methodological challenges when designing DCEs.

2.2 The role of health economics

Health resources are finite because decisions, such as those made by health service commissioners or policy makers, are subject to a strict budget constraint set by the available annual healthcare budget (Appleby, 2013). Within a finite budget, the funding of any healthcare intervention or service creates an opportunity cost (Gold et al., 1996). This opportunity cost arises as another area of expenditure must be reduced or sacrificed completely. Efficiency is achieved through allocating resources to produce the most benefit, but also through ensuring healthcare production is achieving an optimal output given the resources consumed (Retzlaff-Roberts et al., 2004).

In the UK, decision-making bodies such as the National Institute for Health and Care Excellence (NICE) use information from economic evaluations to inform the efficient allocation of healthcare resources (Drummond & McGuire, 2001). Table 2.1 defines the available methods of economic evaluation.

Type of economic evaluation	Definition (from Drummond & McGuire (2001))	Theory under- pinning the method ^a	Applied example
Cost- effectiveness analysis (CEA)	An evaluation where the costs of a health technology or service are weighed against outcomes measured in units that are readily counted, such a life years gained.	Extra- welfarist	Puri et al. (2012) compared radiotherapy to surgery in the treatment of lung cancer through the primary outcome, life-years gained.
Cost-utility analysis (CUA)	An evaluation where benefit is measured using a combination of health state and length of life, for example, the Quality Adjusted Life Year (QALY).	Extra- welfarist	Athanasakis et al. (2012) evaluated a hypertension intervention using QALYs to compare treatment as opposed to no intervention at all.
Cost-benefit analysis (CBA)	An evaluative framework that elicits people's valuation of benefit for a good or service using 'money' as the unit of measurement and compares this to the associated costs.	Welfarist	Muangchana & Warinsatian (2011) evaluated an influenza vaccination programme using WTP values elicited through an online survey with the specific aim of capturing the wide range of intangible benefits from the vaccination.

Table 2.1: Definitions of methods of economic evaluation

^{*a*} Welfarist is defined as maximising utility to individuals; extra-welfarist is defined as maximising something other than, or in addition to, utility (for example, individuals' health).

The aim of using evidence generated from economic evaluations is to minimise opportunity costs or maximise benefits (Gafni & Birch, 2003). Therefore decision-makers allocating finite healthcare resources need to consider which intervention or service creates the most benefit. Methods of economic evaluations compare the total benefits with the total costs of a new intervention or technology to current practice thus indicating where resources are best allocated (NICE, 2013).

NICE currently takes an extra-welfarist view and uses model-based cost-effectiveness analysis to inform decisions. Model-based cost-effectiveness analysis involves comparing the relative costs of a new intervention with existing interventions with gains in health status, quantified in terms of extension to, and the quality of life, by using the QALY (NICE, 2013). Cost-benefit analyses, underpinned by the welfarist view, have yet to be used to inform national resource allocation decisions in the UK and this is largely down to problems with the methodologies used to elicit WTP values (Grutters et al., 2008; Cookson, 2003).

2.2.1 The role of public and patient preferences

There have been increased calls for more patient and public involvement in healthcare decision-making at all levels (Hogg, 2007; Mockford et al., 2011; Stewart et al., 2011). Patient involvement is important at an individual level (shared decision-making), policy level (NICE patient experts) and at a commissioning level (incorporating patient preferences in technology evaluations). As well as occurring at different levels, patient

involvement can occur to varying degrees. For example, patients may play a crucial role in developing the direction of a research project or they may provide vital data on their utility for health states (Boote et al., 2010; Brazier et al., 1999). However, in order to incorporate public and patients' views into healthcare decision-making, the methods used must be robust, underpinned by theory, and able to accurately elicit preferences (Cookson, 2003).

2.3 Valuation methods

Methods used to value preferences for healthcare services and interventions are generally split into revealed and stated preference methods.

2.3.1 Revealed preferences

Revealed preference data are acquired through direct observation of consumer behaviour and analysis of real demand curves through observable changes in existing markets (Ryan & Gerard, 2003a). Therefore to use revealed preference data, an observable market must exist.

2.3.1.1 Travel cost method

One revealed preference method available for valuing a good or service is the travel cost method (Hanley et al., 2001). Under this approach, the costs incurred by an individual when they choose to visit an amenity or service, such as their cost of travel or loss of work, are said to be indicative of their valuation (Hanley et al., 2001). The approach has been used extensively within environmental economics where data are regularly published (Willis & Garrod, 1991; Chen et al., 2004; Font, 2000). As an example, Blakemore & Williams (2008) used the travel cost method to value a beach in Turkey taking into account British tourists' expenditure on flights, accommodation and meals as well as the opportunity cost of spending their time there as opposed to enjoying other activities. Although the travel cost method is less popular in health economics, published health-related studies have looked at access costs to clinics to estimate patients' WTP for more local service provision (Clarke, 2002). However, the method has been criticised because of its superficiality and reliance on an accurate valuation of the cost of time, the exclusion of equipment costs, and the difficulty in distinguishing what is valued by consumers (Randall, 1994).

2.3.1.2 Hedonic pricing

Another approach used to measure revealed preferences is the hedonic pricing method, which involves assessing the market-value of a particular good or service through price premiums or discounts (Perman et al., 2003). Studies in health economics, Jensen &

Morrisey (2011) and Robst (2006), have considered the demand for various health insurance policies and the trade-offs consumers were willing to make between premiums and cover. Market failure in healthcare means prices are rarely present and it can be difficult to find data to support this method particularly in a publically-provided healthcare system such as the UK (McIntosh et al., 2010).

2.3.1.3 Averting behaviour

Behaviour revealed in wage premiums accepted by workers for riskier jobs, or individuals' averting behaviour through their demand for risk-reducing safety measures in vehicles such as airbags or seat-belts can also be used to measure revealed preferences (Rohlfs et al., 2015; Hakes & Viscusi, 2007; Perman et al., 2003). For example, Bresnahan et al. (1997) looked at the costs involved in either averting or mitigating dangers to value clean air in Los Angeles with measures of the steps taken to prevent smog inhalation and behavioural reactions to 'bad-air days' considered to indicate value. However, difficulties in distinguishing between people who do not change their behaviour in response to risks because they have permanently abstained from the risky activity and those who do not value the activity at all were acknowledged by Bresnahan et al. (1997) as creating bias in the results.

Researchers can, in general, be confident that revealed preference data are reflecting actual behaviour but the data are not always available or useful for valuations of healthcare interventions or services (Ryan et al., 2008). In healthcare, the goods or services being valued may not presently be offered and, therefore, revealed preference methods can only be used for auditing current policy, rather than informing new policy or proposed changes to the service (Ryan et al., 2008). In addition, the revealed preference data are often difficult to identify in the UK, where patients rarely face market conditions when making decisions about healthcare consumption (McIntosh et al., 2010).

There are two key challenges with using revealed preference methods in the context of healthcare: 1) decision-makers often require information on a new service or a service for which people may not have prior experience; and 2) the need to apply the principles of opportunity cost means we need to understand what, if anything, was 'given up', partially forgone or not chosen (Ryan et al. 2008). Another common criticism, particularly within the environmental literature, is that revealed preference data do not allow for non-use values (Cicchetti & Wilde, 1992). For example, despite no observational data on some people accessing a clinic, these individuals may still derive value from its availability for others or for their own future consumption.

Limited and absent data from revealed preferences has resulted in the development of stated preference methods which may be preferable in the context of public goods such as healthcare (Train, 2009). Section 2.3.2 describes available stated preference methods and how they overcome some of the criticisms of the revealed preference methods described in section 2.3.1.

2.3.2 Stated preference methods

Stated preference methods can be used to elicit preferences to generate values of monetary benefit, for use in a CBA, or values in terms of preference weights (to generate utility values) associated with a health state, for use in a CUA. Stated preference methods involve asking people to state how they would behave, or what they would prefer, based on a hypothetical scenario (Louviere et al. 2000). Humans make choices on a daily basis. Given that humans live in a choice-based world, one suggested approach to classifying the available stated preference methods is to group them into trade-off based and non-choice approaches with a strong recommendation to use trade-off based methods which reflect day-to-day human decision-making behaviour (Carson et al., 2009). Categorisation using system is not based on theory *per se* but is a practical approach that has been adopted by a number of authors (see: Carson et al. 2009; Louviere & Islam 2008). Figure 2.1 was developed to explain different preference methods after a comprehensive review of the stated preference literature. Figure 2.1 is used to inform sections 2.3.2.1 to 2.3.2.5 which describe the available types of stated preference methods.





2.3.2.1 Non-choice methods

The visual analogue scale (VAS) is a non-choice based approach and could be used to generate a utility value for use in CUA (Drummond & McGuire, 2001). The VAS method has been used to elicit individuals' preferences for different health states often by providing a scale between 0, death, and 1, perfect health (Whitehead & Ali, 2010). Respondents then rate various health states which can be used to estimate the weightings they place on each state and transform length of life into QALYs (Parkin & Devlin 2006). A fundamental criticism of the VAS is that it does not consider opportunity costs as the respondent is not required to trade or give-up anything when rating the health states (Brazier et al., 1999), and therefore VAS is generally not recommended as a method for eliciting utility values.

Other methods, which align with the concept of opportunity cost, include the trade-off based stated preference methods. In trade-off based methods respondents are asked to trade and make decisions about value by exchanging health for time, money, risk or other goodrelated attributes. Whole service trade-off based stated preference methods tend to be used to value 'whole services' but, in principle, there is no reason why different attributes of a service could be valued in separate surveys. In practice, this step-wise approach to valuing service attributes is not done due to potential problems with cognitive overload for survey respondents, resulting in strategic bias (such as task learning) (Holmes & Boyle, 2004).

2.3.2.2 Standard gamble

Standard gamble is used to elicit utility values for use in CUA and involves asking people to disclose how much risk, quantified using a probability, that they are prepared to accept to undergo a treatment that will either cure them from a health state or cause death (Jones, 2007). The less desirable the health state, the more risk is taken and therefore it is possible to estimate the relationship between different levels of quality of life and provide weightings for QALYs (Drummond & McGuire, 2001). It is well established that people find risk a difficult concept to understand and, although trade-offs are made in everyday life, explicitly stated probabilities and changes in magnitudes have been difficult to interpret for standard gamble respondents (Dolan & Iadarola 2008). The concept relies heavily on the assumption that these differences in magnitude are appreciated by the respondent and therefore the method is highly susceptible to framing effects (Abellan-Perpinan et al., 2005).

2.3.2.3 Time trade-off

The time trade-off is a method also used to elicit utility values for use in CUA (Drummond & McGuire, 2001). The time trade-off method was developed specifically to value health states and it is not based on the economic utility theory that supports the standard gamble method (Bleichrodt & Johannesson, 1997). In the time trade-off method, respondents are required to trade length of life for quality of life (Torrance, 1986). For example, respondents may be willing to forgo a long life in poor health in exchange for a shorter life of better quality. The length of life they are willing to exchange indicates the relative preferences for different health states and this in turn can be used to weight QALYs (Clarke et al., 2010). If a health state is so bad, then it may be that people are willing to trade their whole life because it is seen as worse than death, however, the time trade-off method in this case cannot distinguish between different 'worse than death' health states (Devlin et al. 2011).

2.3.2.4 Contingent valuation

Contingent valuation is a trade-off based method used to provide monetary values for use in a CBA (Boxall et al., 1996). In a contingent valuation study, individuals are directly asked their WTP or willingness-to-accept (WTA) compensation for a particular event or scenario. WTP is an approximation of the compensating variation required to maintain each individual at the same 'utility' level, before and after the prospective change (Diener et al., 1998). WTA is an estimate of the equivalent variation, the amount an individual requires to achieve the utility level if a change had occurred (Diener et al., 1998). The WTP or WTA valuations can then be averaged from a sample of people and compared with total costs to see whether a good, service or change to provision is worth pursuing (O'Brien & Gafni, 1996).

Contingent valuation has been used extensively within many areas of economics (Venkatachalam, 2004; Klose, 1999). For example, in environmental economics, this method was used in a valuation of the compensation to be paid by Exxon after an oil leak in Alaska (Carson, 1994). Within health economics, contingent valuation has been used for estimating WTP for a range of goods and services. For example, the method was used to estimate values for mortality and morbidity risk reduction through WTP for preventative medication (Nielsen et al., 2012). Although there have been many studies published, contingent valuation has yet to be used in UK healthcare decision-making (Smith & Sach, 2010). There are many criticisms of the method. Contingent valuation is often used to value unfamiliar goods (when revealed preference methods could not be used) but this means respondents may find it difficult to price something, particularly if it is health-related (Ryan et al. 2008). The contingent valuation survey approach is open to many known sources of bias including zero-bids (where respondents state 'zero' when they actually have a positive WTP), protest-bids (where respondents give artificially high values to influence decisions), anchoring effects (when respondents are influenced by their previous responses or suggested responses), ordering effects (where valuations differ depending on the order of questions) and payment vehicle effects (where certain terminology, for example using 'tax' or 'subsidy' influences WTP values) (Hanley et al., 2001; Holmes & Adamowicz, 2003).

Given these problems with using the contingent valuation method, methods such as conjoint analysis, DCEs and best-worst scaling, have been suggested as viable alternatives and are increasingly used in the healthcare context.

2.3.2.5.1 Attribute trade-off-based stated preference methods

Conjoint analysis, DCEs and best-worst scaling are three types of attribute trade-off-based stated preference methods. The methods all present a series of scenarios and ask respondents to state (through choosing, rating or ranking) the preferred scenario that describes a good or service using pre-defined attributes and levels. These methods are used to value individual attributes that make up goods or services and allow estimation of the trade-offs between different attributes. Trade-offs are defined as a balance of opportunity costs which maximise utility. Attribute based methods are grounded in Lancaster's Theory (1966) (see also section 2.4.1), which suggests consumers value particular features of goods or services rather than the product as a whole.

2.3.2.5.2 Conjoint analysis

Conjoint analysis (CJA) is a term used inconsistently in the published literature. Both the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and Sawtooth survey software (Sawtooth, 2012) refer to DCEs as CJA (Bridges et al. 2011; Louviere et al. 2010). However, this thesis distinguishes between the approaches in line with the recommendations of Louviere et al. (2010), and defines CJA specifically as a method which requires respondents to rate or rank hypothetical alternatives which are described by attributes of varying levels (Ryan et al., 2008).

Under ranking CJA, respondents are asked to order the alternatives to represent their preferences for each alternative (Boyle et al., 2001). In rating CJA, respondents are asked to consider not just the order of their preferences but also the strength of their preference (Boyle et al., 2001). Inclusion of a cost attribute allows analysis of the exchange of money for different levels of other attributes and thus we can derive their value and WTP (Johnson et al., 2011). CJA has been used for identifying preference weights to generate QALYs. It can therefore be thought of as both welfarist and extra-welfarist in its ability to inform both CBA and CUA (Flynn, 2010).

The development of specialised Sawtooth software (Sawtooth, 2012) has stimulated the use of adaptive conjoint analysis (ACA) (Fraenkel, 2010). In ACA, the choice sets are presented to respondents as either a rating or ranking task that are then updated based on their previous choices. The ACA approach is used as an attempt to improve the efficiency of the experimental design (Fraenkel et al., 2001). However, ACA surveys have been criticised for their time-burden (with surveys taking longer to complete than traditional designs) which could result in lower completion rates (Cunningham et al., 2010).

CJA is neither recommended, nor used, in the UK decision-making processes. The CJA ranking or rating task is not necessarily representative of a typical decision-making process. When choosing a good in a supermarket, individuals rarely rank or rate the goods. Louviere et al. (2010) further criticise CJA because it does not align with neoclassical economic theory that relies on a discrete choice rather than ranking or rating. This means that the analysis is not underpinned by economic theory and so the models used to explain uncertainty in typical economic theory are not applicable. In addition, because the data elicited are not representative of actual decision-making behaviour, it is not appropriate to compare the stated preference data with revealed preference data (Louviere et al., 2010). Louviere et al are quite clear when they state: *"Conjoint analysis should be seen for what it really is, namely a purely descriptive way to fit a statistical model to a set of observed ranking or rating data with no ability to inform questions about how consumer behaviour is likely to change in response to changes in the choice context"* (p. 69).

2.3.2.5.3 Discrete choice experiments (DCEs)

DCEs ask respondents to make a choice, usually pair-wise, to indicate their preference between a discrete set of two or more alternatives describing goods or services using attributes and levels (Ryan et al., 2008). Representing a choice task as used in DCEs is arguably more natural because they more closely represent decisions made by humans in everyday life. As DCEs involve respondents completing a relatively simple task, choosing only one option, but the elicited data will be 'weakly ordered' (Louviere et al., 2010). 'Weakly ordered data' means information about the strength of preference for each individual attribute is not completely attained. However, there is a balance between using methods which provide stronger ordering data, such as CJA, and task complexity.

If a cost attribute (or price proxy) can be included in a DCE then it is possible to estimate a monetary valuation for the product being valued and therefore provide benefit measures for use in a CBA (Ryan & Gerard, 2003b). Furthermore, the relative importance of the different attributes included in the DCE can also be quantified. If health states are used as the attributes in the DCE, the method can also be used to generate preference weightings for the different components of a multi-attribute health status measure and the preference weights used to create utility values subsequently used in a CUA (Brazier et al., 2012; Stolk et al., 2010; Flynn, 2010).

Within the discipline of health economics, applications of DCEs have gone beyond providing data for economic evaluations and now answer research questions in a variety of contexts and settings (Clark et al., 2014; de Bekker-Grob et al., 2012). Two systematic reviews of published DCEs identified that the method had been used to address a variety of questions such as: estimating preference weights (utility); valuing health outcomes; measuring patient experiences; investigating the trade-off between health outcomes and patient experiences; investigating job-choices; and developing priority setting frameworks (de Bekker-Grob et al., 2012; Clark et al., 2014).

2.3.2.5.4 Best-worst scaling

A relatively new valuation method, often seen as an extension to DCEs, is best-worst scaling (BWS). There are three distinct cases of BWS methods (case one, case two, and case three). In case one, also known as object case, respondents choose the attribute that is the best and the attribute that is the worst (Louviere & Flynn, 2010). In case two, profile case, respondents choose the best attribute and the worst attributes in an alternative (Ratcliffe et al., 2012). The third case, multi-profile case, is the most similar BWS task to a traditional DCE as respondents choose the alternative that is the worst that is the worst from a choice set (Marti, 2012).

BWS is argued to be advantageous compared with DCEs because the choices made reveal more information about the strength of people's preferences through fewer choice sets which could, in turn, reduce the response error (Xie et al., 2014). However, studies comparing BWS and other methods have found it produces similar but slightly less reliable estimates (Ratcliffe et al., 2011). A recent study by Whitty et al. (2014) used qualitative interviews to compare the validity and acceptability of DCEs and BWS, and found more support (in terms of consistency and trading behaviour) for the traditional DCE task.

2.4 Theoretical foundation of DCEs

DCEs and BWS share a solid theoretical foundation that underpins the implementation of the methods and interpretation of the results. In contrast, CJA does not share the same theoretical basis. Rating or ranking experiments are not considered a useful approach for eliciting preferences, and therefore are not discussed further in this thesis. BWS was an emerging method and was not perceived to be a relevant method for eliciting benefit-risk trade-offs when this thesis was conceptualised. Traditional DCEs have been used more often than BWS and are a more established method (Clark et al., 2014).

Furthermore, given evidence suggesting that DCEs are less burdensome and possibly produce more reliable results, a traditional 'choose the most preferred' choice experiment was selected as the focus for this PhD. For this reason the remainder of the thesis focusses on the design, analysis and application of DCEs. DCEs are based on two important theories: Lancaster's Theory and RUT described in section 2.4.1 and 2.4.2, respectively.

2.4.1 Lancaster's theory

Prior to the 1960s, it was traditionally thought that consumers valued goods and services as whole 'offerings' and decisions were made on psychological taste for the good or service rather than on their distinguishing features (Debreu, 1960; Johnson, 1958). In the 1960s, this traditional economic theory was challenged and new theories developed to try and explain choice behaviour; suggesting goods or services are made up of attributes or characteristics (Lancaster, 1966; Alcaly & Klevorick, 1970).

All attribute-based valuation methods, including DCEs, are based on theories developed by Lancaster in his 1966 paper: A New Approach to Consumer Theory. This paper queried the traditional theories of consumerism that "goods are goods" (p.132, Lancaster 1966) and suggested three hypotheses: 1) consumers value the attributes that a good possesses, rather than the good itself; 2) goods are made up of many attributes which are not necessarily unique to that good; and 3) two goods together may possess different attributes from those when they are separate. DCEs incorporate these three hypotheses through the development of alternative profiles, choice sets and interactions between attributes, respectively.

A key inference to be made from Lancaster's paper is that goods are distinguishable because of their attributes, and possess inherent differences on which a consumer chooses. This theory allows an objective explanation of *why* some things are good substitutes and why others are complements, rather than relying on the assumption of 'intrinsic properties' that traditional consumer theory depended upon (Johnson, 1958). Another key inference that can be drawn from Lancaster's theory is that the values placed on each characteristic can be summed to estimate the value of the good or service as a whole (Ryan, 2004).

2.4.2 Random utility theory

The way individuals choose between the alternatives presented in a DCE can be explained using probabilistic choice theory. Two hypotheses can explain choice behaviour: 1) decisions are random and utility is deterministic; and 2) decisions are deterministic and utility is random (and 'actual' behaviour cannot be modelled). In the first hypothesis, the individual can be assumed to choose on impulse as influenced by psychological factors (Tversky, 1972). In contrast, the second hypothesis regards individuals as utility maximisers, but there is a random component to this maximisation (Thurstone, 1927). The second hypothesis, that suggests decisions are not made randomly but rather utility has a random component, has been explored most in the context of DCEs through random utility models to explain behaviour and provides the model for this thesis (Ryan et al., 2008).

The random utility models used to explain the uncertainty around predicting consumer and respondent choices are underpinned by RUT. RUT was originally investigated by Thurstone (1927) who looked at the derivation of satisfaction, or utility, through a 'Law of Comparative Judgement' with a psychological perspective. The theory was developed substantially in the 1970s with econometric input from the 2000 Nobel Prize Winner, Daniel McFadden (McFadden, 1974, 1986).

RUT provides a deterministic-decision framework which is not trying to explain irrational behaviour, but model the researcher's lack of information. The lack of information results in an error which could be due to measurement errors or latent attributes that influence choice or heterogeneity in preferences (unobserved differences in taste). Therefore, the psychological factors which influence choice are incorporated into this random component of utility (Louviere et al., 2010).

RUT is based on the simple axiom that it is not possible to observe the 'actual' utility function, however, it is possible to infer what is affecting utility from deterministic decisions being made (Louviere et al., 2010). Utility (U_{in}) is said to be a latent construct that people hold in their mind (hence it is unobservable), and this construct is made of both systematic (V_{in}) and random (ε_{in}) components (Louviere et al., 2000). In equation 2.1, U is the unobservable utility of alternative i for individual n, V is the deterministic utility and ε is a random unobservable component:

$$U_{in} = V_{in} + \varepsilon_{in} \qquad [2.1]$$

In a DCE, *V* is constructed of: the attributes and levels in the alternative scenarios presented; and covariates to explain individuals (such as socioeconomic differences). The random component reflects all unobservable factors influencing decision-making behaviour. This random element to behaviour and decision-making means it is not possible to accurately predict the alternative the respondent will choose and therefore choices are modelled as probabilities.

RUT assumes individuals seek to maximise utility, U, but as their actual utility function is unobservable, a probabilistic utility function should be used to estimate their choices. This probabilistic utility function estimates the likelihood of the individual choosing an alternative out of a set of feasible alternatives. Equation 2.2 illustrates this concept and defines the probability of choosing alternative *i* over another alternative *j* as:

$$P_{i} = Prob(U_{i} > U_{j}) = Prob(V_{i} + \varepsilon_{i} > V_{j} + \varepsilon_{j}) = Prob(V_{i} - V_{j} > \varepsilon_{i} - \varepsilon_{j})$$

$$[2.2]$$

The first part of equation 2.2 explains that the probability that the utility from alternative i is greater than alternative j is the same as the probability of the deterministic utility of alternative i, given random errors, is greater than the deterministic utility of alternative j. This implies that if there is a high probability of choosing alternative i there is considerably more utility offered upon its consumption than alternative j.

If, in a particular set of alternatives, *i* has more desirable levels in its attributes, then the probability of choosing alternative *i*, P_i , tends towards one because the difference in deterministic utility between alternatives *i* and *j* increases. If the utility difference was only small, then the probability would tend towards 0.5 as it is more difficult for the individual to choose between the alternatives, and therefore the decision becomes more random; like tossing a coin. Estimates of these probabilities can be calculated from data elicited from a DCE by looking at the choices made by the respondents in each set as the proportion of respondents choosing alternative *i* represents the probability of an individual choosing *i*.

Popular discrete choice models include the conditional logit, multinomial logit (MNL) and mixed logit (MXL) (de Bekker-Grob et al., 2012; Clark et al., 2014). There are no agreed gold standards for the design, analysis or interpretation of a DCE. However, there are recommendations published by organisations and experts in the field (Bridges et al., 2011; Lancsar & Louviere, 2008; Ryan & Gerard, 2003a). These guidelines were used to inform the DCE conducted as part of this thesis (see Chapter Three).

2.5 Risk in healthcare valuation

Within healthcare, the number of studies using DCEs has been gradually increasing since 1990, and the number containing risk as an attribute has also increased (de Bekker-Grob et al., 2012; Ryan & Gerard, 2003b; Clark et al., 2014; Harrison et al., 2014).

2.5.1 Definition of risk

The British Medical Association (BMA) defines risk as "the probability that an adverse event will occur, analogous to the likelihood, the odds or chance" (p.6, Mansfield et al. 2012). In the simplest definition, risk can be thought of as concept linking the probability of an outcome, usually negative, occurring and the severity of that outcome (Mansfield et al., 2012). A risk may be categorised as 'high' because of the magnitude of its likeliness or because it has a very unpleasant outcome (Lipkus, 2007). Similarly, a negative risk (probability of something bad happening) can be re-framed to be a positive risk (the probability of something good). For example, the risk of mortality (a negative frame) can be translated into a positively-framed chance of survival (Edwards et al., 2001). Although often used interchangeably, risk is distinctly different from uncertainty. For the purpose of this thesis, a definition by Mishel (1990) is taken which describes uncertainty as an unknown, or unquantifiable, probability associated with an event occurring.

Risk is made up of multiple components. When considering risk, individuals may take into account: the severity of the possible outcome and its irreversibility; the baseline level of risk that is typical for everyone; the duration of risk exposure and whether they will return to a baseline level; event time; the certainty surrounding the risk; the objectivity of the risk and their preformed perceptions about their likelihood; and risk latency with regards to whether the risk will occur now or in the future (Hammitt & Graham 1999).

Another key component is whether the risk is viewed as being voluntary or imposed on an individual (Edwards et al., 2002). Risks can be voluntary when, for example, there is an established link that being overweight can create an increased risk of diabetes and yet individuals do not lose weight (Abdullah et al., 2010). Risks can also be imposed on an individual over which they have no control, for example, exposure to polluted water and the risk of parasitic disease (Wu et al., 1999). Other risks are combinations of voluntary and imposed risks, such as a genetic predisposition to a condition which may be increased by certain lifestyle factors (Criswell et al., 2006). The different components of risk are defined in Table 2.2.

Table 2.2 Components of risk

Component	Definition
Severity of outcome	The significance of the hazard or event.
Irreversibility of outcome	The opportunity to undo the event and return to a previous health state.
Baseline level of risk to everyone	The risk elevation for an individual.
Duration of exposure to risk	The length of time an individual is exposed to a risk which could be temporary or permanent.
Event time horizon	The time until the hazard will occur; this could be immediate or long-term.
Certainty of risk estimates	The risk could be calculable as a point estimate or defined with confidence intervals.
Objectivity of risk	The risk statistic may be applicable to a whole population or specific to an individual.
Perceived risk and experience	The risk may be influenced by an individual's experience of the hazard.
Risk latency	The risk may not be immediate but something that becomes elevated in the long-term.
Imposed and voluntary risks	An individual may or may not have control over their risk for a particular hazard.

Source: Hammitt & Graham (1999)

2.5.2 Risk in healthcare

Risk is a ubiquitous component of health and healthcare delivery in the form of iatrogenic effects. In addition, lifestyle choices such as overeating or smoking can cause a risk resulting in a negative effect on health. A balance between benefits and risks when making decisions about health can occur at various stages of care. Table 2.3 provides examples of different risks an individual may face relating to their health or healthcare choices although this list is not exhaustive.

Table 2.3: Examples of risk in types of healthcare

Type of healthcare					
	Preventative	Prognosis	Diagnosis	Treatment	Screening
Example	Risks	Risk of	Risk of	Risk of an	Risk of a false
risk	associated	mortality	misdiagnosis	adverse	negative/positive
	with lifestyle			event	

The potential for interactions between Tables 2.2 and 2.3 show that risk is a complicated and multifaceted concept and constitutes more than a statistical statement of a probability.

2.5.3 Risk in a DCE

As a ubiquitous part of healthcare delivery, there are many situations when individuals can be faced with different types of risk in the process of their treatment and for this reason risk is a commonly occurring attribute in DCEs (Ryan & Gerard, 2003b; de Bekker-Grob et al., 2012). In line with previous reviews, a recent systematic review by Harrison et al. (2014) showed an exponential year-on-year increase in the number of DCEs containing risk attributes since the 1990s. One explanation for this increased use is that incorporating risk in a DCE will allow estimation of the maximum acceptable risk for a benefit and analysis of preference proportionality (Johnson et al., 2009). If a price proxy is also included, it allows information on the WTP for a risk reduction and an alternative valuation of a statistical life (Telser & Zweifel, 2007).

Recently, there have also been suggestions that quantitative estimates of benefit-risk preferences could be used to support regulatory decision-making as a complement to the clinical evidence of interventions (Hauber et al., 2013). With this new potential to inform drug approval decisions it is important that the methods used to elicit people's benefit-risk thresholds are accurate and robust.

Whilst valuation of benefit-risk trade-offs are useful, some studies that presented risk in stated preference surveys have found it has an insignificant effect on preferences or it produces values which do not align with the value of statistical life estimates (Corso et al. 2001; Watson et al. 2009). Corso et al. (2001) proposed that benefit-risk valuations from stated preference methods may appear to be inaccurate because of the following hypotheses:

- Risk estimates have already been formed by respondents who have their own perceived risk;
- 2) Economic theories are not compatible with real human behaviour and assumptions, such as linear responses to risk, falsely indicate 'incorrect' answers;
- 3) Differences in risk magnitude are not being accurately communicated.

The underlying economic theories (described in section 2.4) supporting the use of DCEs as a valuation method require attributes to be interpretable by respondents. The presence of any of the hypotheses from Corso et al. (2001) could introduce bias and invalidate any valuations derived from a DCE.

Hypotheses 1) and 2) could also be due to an increase in the cognitive burden of a DCE which may force respondents to employ simplifying heuristics ignoring the risk attributes. If this occurs, then no information about the trade-offs of risk can be made (Bridges et al., 2011a). Additional support for hypothesis 3) has been confirmed by other authors, for example, Watson et al. (2009) included risk of pelvic inflammatory disease from Chlamydia in their DCE and hypothesise that: *"the insignificance of the risk of pelvic inflammatory disease may reflect the difficulties that respondents had in understanding this attribute"* (p. 623). A systematic review by de Bekker-Grob et al. (2012) stated:

"Studies have continued to include risk as an attribute. [Previous reviews] noted the difficulties individuals have understanding risk, and they commented on health economists giving little consideration to explaining the risk attribute to respondents. There appears to be little progress here... This is clearly still an important area for future research." (p.150).

In a more recent systematic review of risk in DCEs, Harrison et al. (2014) found a lack of training, information and descriptions with fewer than 20% of the studies reviewed containing risk communication aids. The authors also noted that those designs that did contain risk communication aids were published only recently. Also, evidence from other valuation methods such as the standard gamble and contingent valuation, has shown that risks are often misinterpreted by respondents when no aids are provided (Sharma et al., 2002; Corso et al., 2001).

It may be appropriate to provide risk communication aids to respondents to enhance the respondents' understanding of the DCE and their ability to make comparisons and trade-offs. Possible risk communication formats which may aid respondents include comparative graphs, risk grids or risk ladders (Schapira et al., 2001). However, there are many different risk communication formats and, as yet, there is no evidence of their relative effectiveness for trade-off decisions. Exploration of the different risk communication formats which could be used in to aid respondents understanding of attributes in a DCE is expanded on in a structured literature review presented in Appendix 3.1 and described in section 3.2 of Chapter Three.

2.5.4 Decisions with risk

Every day, people are confronted with decisions to make about risky behaviour. Many decisions are made based on gut feeling rather than on the interpretation of a mathematical probability, and this if often because no explicit statements about the risk exist (Slovic et al., 2005). For example, choices about what to eat or a mode of transport to travel on are often based on 'feeling' or intuition rather than 'analysis' or logical reasoning (Slovic et al., 2005).

Non-perfect reason based on feelings of fear helped humans evolve a fight or flight reaction to a hazard (Renn, 2006). In risky situations, humans have had to make quick decisions and judgements based on instinct or intuition rather than hesitating to assess the situation and take into account probabilities. Whilst this has worked well for human survival, it has resulted in a brain inherently poorly equipped to reason with factual risks (Kahneman, 2012). This cognitive shortcut is described by Slovic et al. (2005) as the

'affect heuristic'. As a consequence of the 'affect heuristic', when risk information is presented, humans' evolutionary traits prove an obstacle to effective communication.

There are, however, occasions where individuals need to weigh-up situations with numbers and facts of risk likelihood. In healthcare, there have been increasing calls for shared decision-making with individuals (patients, clinicians) deciding treatment based on an assessment of the potential consequence and their likelihood (Charles et al., 1997; Stiggelbout et al., 2012). Choices about the benefits and risks of healthcare treatment should go beyond that of gut feeling and minimising the affect heuristic to communicate risk effectively is an essential part of accurately eliciting people's preferences.

2.5.4.1 Expected utility theory

When choosing between alternatives that are risky or uncertain, an individual must evaluate the presented prospects under risk or ambiguity, respectively. In behavioural economics, decision-making under risk or ambiguity is explained through a number of theories. One well explored theory is Expected Utility Theory (EUT). EUT suggests individuals will weight the possible outcomes by their likelihood (Von Neumann & Morgenstern, 1947). Whilst it is assumed that an individual will still seek to maximise their utility, it is also assumed that they will consistently choose the alternative which is in line with their risk preferences and reject options which involve too much risk for too little reward (Rabin, 2000).

In addition to EUT, there are a number of generalisations which could also explain respondents' behaviour. Hey & Orme (1994) looked at experimentally generated preference data to see how the different generalisations (such as prospective reference theory, weighted utility theory, regret theory and disappointment aversion theory) and different restrictions breaking neutrality (such as absolute risk aversion) fitted with their collected data. Hey & Orme (1994) showed that, in the stated preference data, expected utility held and, with additional structure of the error term, it could predict individual choice. However, other studies have found EUT was frequently violated, with preferences more in line with prospect theory and biases induced from loss aversion (Rabin, 2000; Avineri & Prashker, 2004). RUT overcomes some of the shortcomings of EUT and is the standard method for the analysis of choice data as it allows for unobserved, random, influences on choice and will therefore be used to analyse choices in the empirical chapters of the thesis (Lloyd, 2003).

2.5.4.2 Risk literacy

There is a substantial evidence base (see Lipkus et al. 2001; Gigerenzer et al. 2007) which shows a general lack of understanding about risk across all demographics, possibly related to the psychological reasons described in section 2.5.4. Effective risk communication is also susceptible to the challenges of any numerical or scientific information. When asked which risk was biggest out of: 1 in 10; 1 in 100; or 1 in 1,000, 28% (n=282) of Germans and 25% (n=249) of Americans provided an incorrect answer (Galesic & Garcia-Retamero, 2010). The authors of this study suggested that because risks with the largest magnitude were described with the smallest denominator frequencies, the subjects became confused.

Similarly, Denes-Raj & Epstein (1994) asked individuals to choose between two bowls: 1) a bowl with nine white beans and one red; and 2) a bowl with 100 beans with nine or fewer being red. The individuals who took part won a prize of \$7 if they picked one red bean. Despite knowing that probability of bowl two was against them, the individuals were observed to indicate that they felt they had a better chance when there were more red beans available. This 'ratio bias' has been replicated in other studies and emphasises the challenges of choosing framing affects using different numerators and denominators. There are multiple examples of biases and framing effects in the risk communication literature; the term 'collective statistical illiteracy' has been used to describe these challenges (Gaissmaier & Gigerenzer, 2008; Gigerenzer et al., 2007).

2.5.5 Communicating risk information

Communicating information on risk introduces the concept of framing affects and how the choice of risk information and its presentation may affect the interpretation of the probability statistic.

2.5.5.1 Absolute and relative risk

A popular example of miscommunicated risk is the 1995 'pill scare' where a new generation of the contraceptive pill was found to increase women's risk of thrombosis by 100% (Committee on Safety of Medicines, 1995). The subsequent newspaper headlines failed to explain an increased relative risk of 100% was an increase in absolute risk from 0.014% to 0.028%. The news coverage and misleading risk information resulted in an estimated 13,000 additional abortions at a cost of around £46 million to the NHS (Furedi, 1999). The example of the 'pill scare' illustrated the importance of using absolute risk information to communicate risk.

A key step in transparency is providing individuals with enough information to make their own choices (not necessarily the 'right' answer). A lack of information means people make their own inferences of the situation using subjective value judgements. For this reason, it is generally accepted that risk should be communicated as an absolute value rather than relative risk.

2.5.5.2 Sufficient information

In addition to presenting the appropriate statistic, the risk information should be contextualised in a transparent way. A study by Gigerenzer et al. (2005) surveyed a typical weather forecast to understand how the information was interpreted by members of the public. When asked what a *"30% chance of rain tomorrow"* meant, most respondents interpreted this to mean that there would be rain 30% of the time or in 30% of the area. The correct response, that it will rain 30% of the days like tomorrow, was the least chosen alternative in the multiple choice question. Gigerenzer et al. (2005) hypothesise that this could be due to ambiguity of such statements and stress the importance of re-iterating risk information using clear and simple terminology and, where possible, examples.

2.5.5.3 Positive and negative frames

Another challenge is the framing of the risk. A positive 'survival rate' can be interpreted differently to a negative 'mortality rate'. For example, Tversky & Kahneman (1981) asked individuals:

Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

- *i) Choose between:*
 - A. A sure gain of \$240
 - B. 25% chance to gain \$1000 and 75% chance to gain nothing
- *ii) Choose between:*
 - C. A sure loss of \$750
 - D. 75% chance to lose \$1000 and 25% chance to lose nothing

Most of the study respondents chose the risk averse option of A in decision i), and the risk seeking option of D in decision ii). When risk was framed positively, it was seen that there was a potential gain to be made and the behaviour was more risk averse. However, when risk was framed negatively, the potential for a loss induced more risk-seeking behaviour. These behaviours are systematic and consistently present in choices about financial or health risks (Fishburn & Kochenberger, 1979; McNeil et al., 1982; Kahneman & Tversky, 1979).

In healthcare, the appropriate framing depends on the objectives of the communication as framing can influence patients' decisions for healthcare (Haward et al., 2008). In many areas of public health, risk communication is intended to change actions or behaviour (Rothman & Salovey, 1997). However, risk communication in a DCE context is not about changing behaviour but instead about effectively informing respondents so they can make a decision that is right for them and reflects their preferences.

2.5.5.4 Risk communication formats

One of the key challenges of communicating risk is to encourage risk literacy without controlling or manipulating people's feelings, preferences or motives. To help the objective communication of risk, a number of risk communication aids have been created. In addition to different numerical expressions of the information (percentages, frequencies, decimals, odds ratios), visual stimulants such as graphs or pictures have been developed to help people's interpretation (Peters et al., 2011; Lipkus & Hollands, 1999). As there are many different ways of presenting the probabilistic information to people, the overwhelming literature needed to be approached using a structured and specific focus to identify possible risk communication formats for use in a DCE.

2.6 Conclusion

This chapter has described why valuations using stated preference methods of current and new interventions are useful in the UK healthcare system. The chapter also identified DCEs as an appropriate method for the valuation of healthcare goods and services by comparing its strengths and limitations with alternatives such as contingent valuation and CJA. Evidence of the challenges of using value attributes, particularly risk, in a DCE were also identified by the literature review.

The chapter concluded that if DCEs are to be used to elicit people's preferences for benefit-risk trade-offs, it is imperative that the information is communicated effectively so respondents can make informed choices which accurately reflect their preferences. The following chapter, Chapter Three, describes the development of the DCE used in the empirical chapters of this thesis.

Chapter Three

Valuing risks and benefits: a case study of population-based screening for breast cancer

3.1 Introduction

The aim of this chapter is to report the development of a DCE to value respondents' preferences for benefit-risk trade-offs. The chapter addresses two main objectives: 1) to identify a suitable case study; and 2) to design a DCE. The chapter is presented in three sections. The first section (starting at section 3.2) describes the identification of risk communication formats to be investigated in the DCE. Section 3.3 starts by describing key criteria for selecting a case study to value benefit-risk trade-offs, and how population-based screening for breast cancer was selected. The second section (starting at section 3.4) explains the clinical background of breast cancer, screening programmes, the type of benefit-risk trade-offs involved in participation and the implications of generating quantified preference data for decision-makers seeking to allocate resources efficiently. The third section (starting at section 3.5) describes the process for generating the DCE design including possible attributes and levels, experimental properties and pilot studies. The chapter finishes with a description of the final DCE to be used in the empirical studies described in Chapters Four, Five and Six.

3.2 Risk communication formats

Prior to the selection of a case study, two risk communication formats were identified. These formats were found through a structured review which aimed to systematically identify different formats which can be used to communicate risk information to DCE respondents. If risk is communicated effectively, then it will assist the respondents in making informed choices based on their preferences and thus improve the quality of the data collected. A complete description of the review can be found in Appendix 3.1. The review was also published as a subsection in a paper by Harrison et al. (2014) which can be found in Appendix 3.3

A systematic search of five electronic databases (Medline, Embase, Web of Science, PsychINFO and EconLit) was conducted in April 2013. The search checked the titles of articles in each database for the terms risk and either communication or format, and the whole abstract for topic terms (aid, presentation, display). The exact strategy included the following terms: Title=(risk*) AND Title=(communicat* or format*) AND Topic=(aid* or present* or display*). The search of all databases revealed 1,207 possible hits. After the removal of non-peerreviewed material and duplicates, 390 titles and abstracts were reviewed. In total, 215 full articles were retrieved for further review and 99 papers were included in the final review. Of the included studies, 65 were empirical tests of risk communication formats, 21 were overviews of the risk literature with no reported search strategy, nine were systematic reviews with a structured search strategy, and four looked at presenting uncertainty (for example, the use of confidence intervals as opposed to point estimates). A list of all included studies can be found in Appendix 3.2.

The results of this review uncovered a fine line between providing sufficient information for respondents to understand the risk and overwhelming them with too much. Training materials for DCE respondents should therefore be presented at a level that is accessible to all but should not be so elementary that respondents will ignore it. A prescriptive approach to the DCE, with extensive piloting and revisions, should therefore be conducted to ensure materials are appropriate for respondents.

The identified empirical studies also highlighted the importance of assessing numeracy skills on a respondent's ability to interpret probabilistic information. Preformed ideas about perceived risk based on experience were also identified as a challenge to effective communication of risk information. In understanding how respondents interpret risk in a DCE, numeracy, experience and perceptions could be useful covariates to collect in supplementary questions in a DCE study.

The review identified icon arrays as the risk communication format with the most empirical support which would make it an appropriate comparator to current practice in DCE literature which are percentages only (Harrison et al., 2014). However, no empirical study investigated the use of icon arrays in a trade-off task so it is unclear whether they would be superior in a DCE context. The next sections of this chapter describe the design of a DCE to incorporate the selected risk communication formats (icon arrays and percentages; and percentages only) identified by the review.

3.3 Selecting a case study

This PhD used a case study to quantify benefit-risk trade-offs and investigate the impact of how risk is communicated in the DCE. A list of case study selection criteria were developed based on practical considerations and literature identified in Chapter Two. The criteria provided a pragmatic approach to selection and aimed to ensure the chosen case study would be feasible and a suitable context in which to investigate the effect of different risk communication formats in a DCE.

3.3.1 Case study selection criteria

Eight criteria were used to select a case study:

Criterion 1: The good or service described in the DCE must represent at least one clear example of the different types of risk. As risk manifests itself in healthcare through multiple means (see Tables 2.2 and 2.3), a case study for this PhD thesis should incorporate at least one of these risks in order to investigate the effect of different communication formats.

Criterion 2: The disease area should have significant burden, incidence and prevalence rates as proxies for interest to decision-makers in primary care. Primary care is the first point of access and often the only point of care for many patients. This PhD received funding from the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR). Therefore, the outputs of the research must produce results applicable to primary care and be applicable to a large number of current and potential patients.

Criterion 3: Consumption of the good or services described by the DCE should result in a quantifiable risk. As discussed in Chapter Two, risk and uncertainty are distinctly different concepts with different underlying theories, and with the latter representing an unknown probability of the hazard occurring. The focus of this PhD thesis is the best way of communicating risk attributes which implies that the probabilities must be identifiable and quantifiable.

Criterion 4: Opting-out of the choice set should be a realistic and plausible alternative. If respondents are particularly risk averse it would be advantageous to have a 'no treatment' option for them to choose. Without the ability to opt-out, the estimated valuation may be upwardly biased, implying the DCE respondents are more tolerant of risk than they truly are. In order to add a realistic opt-out, the risk probabilities must be available for current care, the status-quo or no intervention.

Criterion 5: The case study should contribute to the DCE literature with an original application. As there are a growing number of applied DCEs using a variety of healthcare goods and services, it would be preferable to conduct original research rather than duplicate an existing study.

Criterion 6: Access to experts to verify the DCE is appropriate in its use and description of the attributes and levels. There are practical considerations in conducting a high quality DCE in the time span of a PhD project. If the case study could tie in with other areas of

research where patient and expert involvement is already established, this could improve the rate of study progression.

Criterion 7: Stated preference methods should be appropriate and the outcomes of the DCE should be of interest to decision-makers. Preferably there would not be revealed preference data available so the results of a stated preference study are of use. This could be because the intervention is not currently provided by the NHS or because there is no market data due to the reasons discussed in Chapter Two, section 2.3.1.

Criterion 8: The case study should ideally be familiar, in order to minimise respondents' confusion about the clinical context. This criterion aims to minimise the other sources of cognitive burden in a DCE task by ensuring the selected case study is not too complex. This means that respondents are likely to have better formed preferences and as a consequence the training materials prior to the choice sets can be briefer. A familiar topic may also keep the DCE respondents engaged in the task, generating higher quality data.

3.3.2 Identification and appraisal of case studies.

A number of suggested topics were discussed as potential case studies using the criteria described in section 3.3.1. These suggestions were largely driven by originality of the research and links to research on-going in the department which would allow for rapid access to experts, healthcare professionals and patients. Table 3.1 compares the possible case studies based on these criteria.

Table 3.1: Potential case studies for a benefit-risk DCE

Service	Case study	Prevalence	Available	Primary care	Risk attribute	Life-style	Life-time risk
			treatment			risk	
Genetic testing	Age-related macular	Approximately 1.64%	Yes (not for all	Potentially results	Risk of AMD with current lifestyle;	Yes	6% for
	degeneration (AMD)	of UK population. ¹	types and not all on NHS)	from a general practitioner (GP).	risk of false negatives/positives.		Women; 3% for males ⁶
	Retinitis pigmentosa	Approximately	No	No current link.	Risk of developing retinis	No	Unknown
	10	0.025% of UK			pigmentosa; risk of false		
		population. ²			negatives/positives.		
Pharmacotherapy	Hypertension	3% <40, 28% 40-79,	Yes – anti-	Yes	Risk of disease with/without drug;	Yes	80-90% ⁷
		50% 80+ years old of	hypertensive		risk of adverse events (risk of stroke,		
		UK population. ⁴	drugs		heart failure, aneurysm); risk of side		
			-		effects.		
	Hyper-	Approximately 3% of	Yes – statins	Yes	Risk of disease with/without drug;	Yes	Risk of heart
	cholesterolemia	UK population. ⁵			risk of adverse events; risk of side		disease
					effects (risk of heart disease).		increased ⁸
Screening	Breast cancer	Most prevalent cancer	Yes - chemo-	GP prescribing,	Risk of radiation induced cancer;	Complex	12.5% for
-		in women. ³	preventative	referral and aftercare.	risk of false negative; risk of false	-	women ³
			tamoxifen	Some screening in	positive (over diagnosis); risk of		
				primary care vans.	mortality from breast cancer.		

¹*O*'Shea & Harvey (2003)

²Bundey & Crews (1984)

³Cancer Research UK (2010)

⁴*Lane & Lip* (2001)

⁵*Midlands Therapeutics Review & Advisory Committee (2008)*

⁶Women's Eye Health Task Force (2003)

⁷Department of Health and Human Services (2004)

⁸Austin et al. (2004)

Three broad categories were identified as risk-prevalent areas of healthcare: genetic testing; pharmacotherapy; and screening. Two new genetic tests were identified for the inherited eye conditions: AMD and retinitis pigmentosa. The two conditions slowly lead to loss of sight and there are very limited treatments available (O'Shea & Harvey, 2003; Bundey & Crews, 1984). As a consequence of lack of treatment, understanding people's preferences for these tests would introduce the need to value the role of information. However, the genetic test for AMD was so new that the risk of false negative or positives from the test was uncertain (with large confidence intervals) and therefore the case study did not satisfy criterion 3. The genetic test for retinitis pigmentosa was more developed, with known sensitivity and specificity, however, the prevalence of the condition was considerably lower and therefore the applicability of genetic testing would be for a small proportion of the population. A case study of a genetic test for retinitis pigmentosa would therefore fail criteria 2 and 8.

A case study of drugs for either hypertension (high blood pressure) or hypercholesterolemia (high cholesterol) would overcome the problems of genetic testing as these are prevalent disease areas with significant burdens. The disease area is also well researched with the risks of different treatments (hypertensive drugs or statins) well established through multiple long-term randomised controlled trials (RCTs) (LaRosa et al., 1999; ALLHAT, 2002; Lindholm et al., 2005). However, the potential risks are heavily influenced by lifestyle factors such as an individual's smoking status, body mass index (BMI), engagement in physical activity, diet (potassium and salt intake), and alcohol consumption style (Boden-Albala & Sacco, 2000). Therefore calculating reasonable risk attributes for a generic DCE would be complicated and particularly vulnerable to subjective interpretation by respondents who may believe they are more or less at risk than average because of their behaviour. As a result, these two possibilities of pharmacotherapies for hypertension or hypercholesterolemia were dismissed.

Preventative screening is an intervention to detect illness at an early stage and aims to reduce the risk of the disease (Jepson et al., 2000). However, screening is rarely perfect and therefore consumption of a screening programme generates its own risks. For example, tests often have some degree of inaccuracy so there is often a chance of a false-positive or false-negative screening result. Understanding how these risks of participation in a programme must be balanced with the chance of detecting a disease is important for decision-makers trying to assess which services provide the most benefit. The UK's Office for National Statistics (ONS) reports that one of the most prevalent illnesses is cancer, with different forms frequenting the top causes of mortality for both men and women (ONS, 2012). Breast cancer is the most common cancer for women (Ferlay et al., 2010), and the leading cause of death for middle aged women (ONS, 2012).

As a consequence of its prevalence, studies surrounding breast screening are plentiful and there is a substantial evidence base comprising multiple randomised controlled trials and systematic reviews from across the world (Gøtzsche & Olsen, 2000; Kerlikowske et al., 1995; Ohuchi et al., 1995). Despite the UK offering screening, it is not mandatory and therefore an option of 'No Screening' is a plausible alternative in a breast screening DCE. The University of Manchester is also home to the largest independent cancer research organisation in the UK, with world experts in breast screening and breast cancer research working in the institution. Breast cancer and breast screening have also been the topics of long lasting public health campaign with targeted leaflets and media attention (Kamenova et al., 2014; Slaytor & Ward, 1998). The campaigns, combined with high prevalence, means that many people are familiar with the disease (McMenamin et al., 2005).

Breast screening was selected as a service to evaluate in a DCE for this thesis because it satisfied key criteria for a case study.

3.4 Screening for breast cancer

A national breast screening programme was identified as the most appropriate case study from the available options described in Table 3.1. The following sections describe breast cancer as a disease area, the need for a national breast screening programme and the debates surrounding the screening programme in the UK.

3.4.1 Background to breast screening

Breast cancer is a major health issue not just because of the prevalence and mortality risks, but also for financial reasons. Breast cancer costs the UK economy an estimated £1.5 billion pounds annually and the cost to the NHS of a single case of breast cancer is over $\pounds7,000$ (Leal et al., 2012; Dolan et al., 1999). The high expense of general cancer care (an estimated £15 billion annually in the UK) means many Western countries now encourage participation in screening for common malignant diseases (von Karsa et al., 2008).

In England, the NHS currently invites all women between the ages of 50 and 70 years for screening using mammography (an X-ray of the breast) every three years as part of NHSBSP (Cancer Research UK, 2013). This programme is based on the premise that regular screening can identify tumours and ensure therapy is begun as soon as possible. In 2011, about 1.73 million eligible women were screened for breast cancer (excluding other referrals and those outside the current age range), an uptake rate of about 75% (The NHS Information Centre Screening and Immunisations, 2012). The NHSBSP is estimated to have an annual expense of £96 million (Public Health England, 2013).

3.4.2 Facts, figures and debates

Screening for breast cancer using mammography has been proven to detect cases of breast cancer earlier, and in 2010-2011 an estimated 14,725 undiagnosed cancers were detected by screening (The NHS Information Centre Screening and Immunisations, 2012). Women who participate in screening programmes have been shown to have improved mortality rates because of earlier interventions (Independent UK Panel on Breast Cancer Screening, 2012). In addition, women who participate in screening and receive a true-negative have the reassurance of knowing that they are cancer-free (Johnston et al., 1998).

In any X-ray, there is exposure to small amounts of radiation which result in a very small risk of radiation-induced cancers (Yaffe & Mainprize, 2011). In addition, as the mammogram produces an image which is interpreted by a radiographer, there is a chance of a false-negative and a cancer being missed. There is also a risk that the image will locate a true cancer but one which is so slow growing it would never have been harmful in the woman's life-time (referred to as 'over-diagnosis') (Welch & Black, 2010).

In addition to the risk of over-diagnosis, women can be recalled back to the screening centre for biopsies and further unnecessary tests as a result of a false-positive (Hofvind et al., 2004). These unnecessary follow-ups are not only painful, but an incorrect test result can cause undue worry for the woman and her family (Johnston et al., 1998). Additionally, the screening procedure, the mammogram, is not pleasant and many women find it uncomfortable and invasive, and some find it extremely painful (Sapir et al., 2003).

Screening for breast cancer does not change the incidence of cancer and no fewer women will develop breast cancer as a result of screening. Even without a screening programme, the same number of women will get cancer, although they may be diagnosed later by their GP after discovery of a lump or other symptoms. The NHS Information Centre (2012) produced a statistical report on the results of the breast screening programme in England using routinely collected data. These data are summarised in Table 3.2 for the age range 50-70 years. In terms of the cancers detected, other estimates have suggested that as many as half of these cases would have been picked up by other means regardless of screening participation (Independent UK Panel on Breast Cancer Screening, 2012).

Outcome of screening	Risk per 1,000 women screened
Recall	$\approx 108/1,000$ screened
Benign biopsy	$\approx 80/1,000$ screened
Cancer detected	$\approx 8/1,000$ screened
Invasive cancer <1.5cm detected	$\approx 4/1,000$ screened

Table 3.2: Risks to women of breast cancer

Source: The NHS Information Centre Screening and Immunisations (2012)

Some of the possible outcomes of cancer screening are described in Figure 3.1, which shows the progression of women who choose to participate in screening and the relative benefits and harms. This diagram aimed to represent a sufficient simplification of reality and is therefore not exhaustive. Other benefits and risks are not incorporated in this diagram, nor are longer term outcomes which would overcomplicate the figure and lead to a Markov-style model of the probabilities of different states and potential routes (Taylor, 2010).

The benefits and risks of participating in a national breast screening programme has sparked a debate in the medical literature, which caused a complete review of the UK's screening policy by the Independent UK Panel on Breast Cancer Screening in 2010, who published their report in 2012. Whether a NHSBSP causes more harm than good continues to be debated by academics in leading medical journals (Baum, 2013; Kirwan, 2013). To contribute to the debate, there have been efforts to qualitatively discover what women consider when they weigh up the benefits and risks, research into better communication, and pushes for further involvement of women in decision-making (Hersch et al., 2013). By analysing a DCE of benefit-risk trade-offs in breast screening, and how these are affected by risk formats, this thesis adds to these important and on-going debates about the merits of a NHSBSP.



Figure 3.1: Progression of women through screening

3.4.3 Benefit-risks trade-offs in breast screening

The decision to participate in screening is based on the individual's perception of the advantages and disadvantages of screening and the risks of foregoing the mammogram. In the UK, women are invited for screening from the age of 50 years but as of 2011, there are still half a million women in the eligible age range who have been invited for screening but have not participated (The NHS Information Centre Screening and Immunisations, 2012); possibly because they do not believe the benefits outweigh the risks.

Screening women more frequently and for a larger age range, would increase the detection rate at the expense of increased risk of over-diagnosis and radiation-induced cancer (Tabár et al., 1987; Moss et al., 2006). Understanding women's preferences for breast screening and how they balance the associated benefits and risks would help decision-makers provide an effective screening programme which maximises value.

3.4.4 Healthcare setting

The NHS currently screens for breast cancer in the community, with women invited by a letter sent their home address to attend their local screening centre, which may be a clinic, a mobile unit or a local hospital. Few women are referred to the breast screening service (<5% in 2010-2011) and most screening takes place in a primary care setting (The NHS Information Centre Screening and Immunisations, 2012).

3.4.5 Potential outputs of the DCE

As risk is an inherent component of any healthcare good or service, the decision to both approve the funding and use of services on the NHS involves service commissioners weighing up the risks and benefits. As discussed in section 3.4.2, there has been a well-publicised debate about whether the harms outweigh the benefits of breast screening. Quantification of women's preferences through DCEs could contribute to this debate. The elicitation of women's benefit-risk trade-offs in breast screening could provide a number of possible outputs including a measurement of their tolerance of risk in the programme and their WTP for the current screening programme.

Recent UK policy recommendations for breast screening have suggested expanding the screening age so it is available to both younger and older women. The DCE data could identify the drivers of demand and how subgroups of women could be targeted to improve uptake.

3.4.6 Existing DCEs relevant to cancer screening: a rapid review

The next section summarises the published DCE literature which has investigated preferences for cancer and screening generally, and breast screening programmes where available. The DCEs were identified from a systematic review conducted as part of this thesis (see Appendix 5.1 and section 5.2 of Chapter Five).

The results of the systematic review presented in Appendix 5.1 did not identify any studies that have elicited women's preferences for the UK's breast screening programme. However, studies of preferences for similar screening programmes have successfully used DCEs to elicit values for the associated benefit-risk trade-offs. This rapid review involved hand-screening the studies identified by the systematic review presented in Appendix 5.6 to find DCEs looking at preferences for breast cancer-related healthcare or diagnostic screening. For the purpose of this rapid review, Public Health England's definition of screening was used:

"A process of identifying apparently healthy people who may be at increased risk of a disease or condition. They can then be offered information, further tests and appropriate treatment to reduce their risk and/or any complications arising from the disease or condition."

(Public Health England 2013b)

The rapid review identified 17 relevant studies related to breast cancer or screening programmes. Two additional studies were identified (X1 and X11) which were not found in the systematic review. X1 was a study in Japanese so did not meet the English-language inclusion criterion, however, an English language abstract was provided. X11 is a working paper which was not peer-reviewed. The reviewed studies are summarised in Table 3.3. The complete list of references for the included studies is listed in Appendix 3.4. The review aimed to identify studies relevant to large-scale screening programmes in cancer rather than selective or multiphasic screening where the attributes were vastly different to breast screening and therefore uninformative for this case study.

Two DCEs were identified relating to breast screening (X1 and X11). X1 was in Japanese and could not be translated and X11 looked at clinicians' preferences for referring women for screening outside of the current programme's coverage. Three other studies were identified relating to breast cancer that considered women's preferences for surgery (X2 and X4) and their preferences for genetic counselling (X3). The generic breast cancer studies were only tenuously linked to the chosen case study and had little transferable information relevant to the design of a DCE for this PhD.
In the identified screening DCEs, there were commonly occurring attributes such as frequency of screening, cost of the programme, test accuracy, discomfort or pain from screening, risk reduction from participation, and location of test provider. The most commonly occurring risk attributes referred to the programmes' sensitivity or specificity, featuring in nine of the DCEs (X6, X7, X8, X10, X12, X14, X16, X17 and X18). In addition to specificity, the risk was communicated as: chance of a false-positive; chance of being recalled; chance of an unnecessary treatment; chance of an unnecessary follow-up; or risk of unnecessary colonoscopies. The sensitivity of the screening intervention was also described as chance of a false-negative; missed polyps; and possibility of missing a cancer. In line with the findings of Harrison et al. (2014), risk was most commonly communicated as a percentage.

The most common choice question for the screening DCEs was a ternary choice (used by ten studies: X5, X7, X8, X9, X10, X11, X13, X14, X15 and X19) with respondents choosing from two programmes or neither. The neither options were defined as either 'No Screening', indifferent or status-quo. The screening DCEs mainly elicited preferences from members of the public although three studies (X11, X12 and X14) administered their DCE to clinicians.

The review highlighted the potential originality of using breast screening in the UK as a case study for the DCE (only one other study (X11) had a UK setting). The review also identified some common designs in existing cancer screening-related DCEs such as the types of attributes, choice set designs and samples, which are useful starting points for consideration in the design of a breast screening DCE.

Table 3.3:	Breast	cancer	and	screening	DCEs
-------------------	--------	--------	-----	-----------	-------------

ID	Condition	Attributes		Risk format	Choice Set	Sample	Quantitative Model
X1 (2010)	DCE for mammography	Total amount of time taken for the screening Degree of breast pain Possibility of breast cancer being missed during the screening	Effectiveness of reducing deaths caused by breast cancer Total cost required for the screening	Not applicable	Binary choice (ambiguous)	301 members of the public (Japan)	Conditional logit
X2 (2010)	Follow-up breast cancer surgery	Attendance at an educational group programme Frequency of visits	Waiting time Contact mode Type of healthcare provider	Not applicable	Binary forced-choice	331 breast cancer patients (Netherlands)	Random parameter logit
X3 (2006)	Genetic counselling	Information Preparation	Surveillance Direction	Not applicable	Binary forced-choice	210 female breast cancer study participants (Australia)	Random effects probit
X4 (2011)	Breast reconstruction	Materials for reconstruction Number and duration of operations Short-term complications	Long-term complications Aesthetic result Waiting time	Percentages	Ternary choice (including opt- out)	386 breast cancer patients (Netherlands)	MNL
X5 (2009)	Barrett esophagus surveillance	Number of times tested over 10 years	Reduction in risk of dying from esophageal carcinoma	Percentages	Ternary choice (including opt- out)	247 study patients (Netherlands)	Conditional logit
X6 (2009)	Colorectal cancer screening	Test process Test frequency Requirement for a follow-up test if the initial screening test is positive Test-related pain or discomfort	Preparation for the test Risk of complications Test accuracy as measured by sensitivity Test accuracy as measured by specificity	Mix: Percentages for sensitivity and specificity, and frequencies (1/X) for risk of complications.	Binary response (test/ no test)	1588 members of the public (Canada and the United States of America (USA))	Bivariate probit
X7 (2003)	Colorectal cancer screening	Deaths prevented Unnecessary colonoscopies	Result notification	Frequencies (Y/10,000 deaths prevented and X colonoscopies /death prevented)	Ternary choice (including opt- out)	301 members of the public (Australia)	Random effects probit
X8 (2006)	Cervical cancer screening	Time between smears Waiting time for results Chance of being recalled	Chance of abnormality Chance of dying from cervical cancer Cost of the smear	Percentages	Ternary choice (including opt- out)	641 screened and unscreened mixed members of the public (Scotland).	Nested logit

ID	Condition	Attributes		Risk format	Choice Set	Sample	Quantitative Model
X9 (2011)	Cervical cancer screening and vaccination	Lifetime risk of cervical cancer Lifetime risk of genital warts Frequency of smear testing Need for vaccine booster	Target group to vaccinate Risk of side effect Cost of the vaccine	Percentages	Ternary choice (including opt- out)	1157 member of the public (Canada)	MXL
X10 (2013)	Prostate cancer screening	Risk reduction of prostate cancer- related death Screening interval Risk of unnecessary biopsies	Risk of unnecessary treatments Out-of-pocket costs	Simple graphics and frequencies (X/1000)	Ternary choice (including opt- out)	427 member of the public (Netherlands)	MXL
X11 (2011)	Breast screening	Breast cancer risk Patient age Comorbidities	Physical functionability Cognitive functionability	Qualitative descriptor (normal/raised)	Ternary choice (screen/ no screen/ can't decide)	139 clinicians (UK)	MNL
X12 (2006)	Colorectal cancer screening	10-year mortality reduction after screening Risk of false negative Risk of false positive	Annual remuneration Number of avoided deaths-per 100,000 Information	Percentages and frequencies (X/10,000)	Binary choice with constant alternative	294 clinicians (France)	Standard probit Random effects probit
X13 (2010)	Colorectal cancer screening	Reduction in mortality, Frequency of screening Complication risk Location	Duration of screening Patient preparation Pain from screening	Percentages and qualitative descriptors ('small risk of complication')	Ternary choice (including opt- out)	276 screened and unscreened mixed members of the public (Netherlands)	MNL
X14 (2009)	Cervical cancer screening	Screening interval GP descriptors Time since pervious screen Doctor's recommendation	Incentive payment Chance of false negative Chance of false positive	Frequencies (1/X)	Ternary choice (2 different pap tests, standard and liquid, and an opt-out)	382 members of the public and clinicians mix (Australia)	MNL
X15 (2010)	Colorectal cancer screening	Test type Screening interval	Mortality risk reduction	Percentages and icon arrays	Ternary choice (including opt- out)	1034 screened and unscreened mixed members of the public (Netherlands)	MNL
X16 (2009)	Colorectal cancer screening	Test accuracy – find/miss cancers Test accuracy – find/miss polyps Unnecessary colonoscopies	Cost Preparation Collection	Frequencies (X/100)	Binary choice	1157 members of the public (Australia)	MXL
X17 (2007)	Colorectal cancer screening	Test process Preparation Pain	Specificity Sensitivity Cost	Percentages	Binary choice (could state after if would opt-out)	547 members of the public (Canada)	Bivariate probit, standard logit.

ID	Condition	Attributes		Risk format	Choice Set	Sample	Quantitative Model
X18 (2010)	Colorectal cancer screening	Provider Sample type Sensitivity	Risk of unnecessary colonoscopy Test result provider Cost to you	Percentages and fractions	Binary choice with constant alternative	656 members of the public (France)	Random effects probit
X19 (2010)	Colorectal cancer screening	Preparation Location Pain	Risk of complication Frequency Duration	Percentages and icon arrays	Ternary choice (including opt- out)	280 screened and unscreened mixed members of the public (Netherlands)	MNL

3.5 Breast screening: designing the DCE

The next section of this chapter will describe the key stages of creating a DCE building on the step-wise approach following ISPOR guidelines (see Bridges et al. 2011b) for good research practice.

3.5.1 Research questions

This PhD sought to understand if, and how, the communication of risk attributes in a DCE affects respondents' choices and the valuations derived. The case study for breast screening led to a specific set of defined research questions. The DCE aimed to elicit women's preferences for breast screening and a quantification of their balance of the benefits and risks associated with screening programmes. The results of the DCE will contribute to the debate discussed in section 3.3 and the comparisons of different risk communication formats will investigate whether preferences are susceptible to change with the format used.

3.5.2 Sampling frame

There is extensive debate in the literature about whose preferences should be used in order measure the values people place on different aspects of healthcare. The sample population who will answer the DCE survey must be appropriate to answer the research question and large enough to allow for subgroup analysis if necessary (Ryan & Watson 2011). A key criticism of published DCEs has centred around the characteristics of the respondents whose preferences were elicited (Bryan & Dolan 2004). It has been acknowledged that patients', the public's and health professionals' preferences will differ and this should be carefully considered when designing a DCE (Montgomery & Fahey 2001). The sample chosen will also relate back to the perspective taken and whether an extra-welfarist or welfarist view point is pursued.

In health economics more generally, it has been argued that the general public's (taxpayers') preferences should be taken in to account for a publically-funded health service (Bryan & Dolan 2004). Other contrasting views suggest that the patient experience plays an important role in determining preferences and that clinicians also have views to be incorporated (de Wit et al., 2000). DCEs have elicited preferences from healthcare professionals, hospital managers or policy makers (Bech, 2003; Gidman et al., 2007; Payne et al., 2011). There is some consensus (based on advice from the Panel on Cost-Effectiveness in Health and Medicine) that preferences taken from the public contain a broad range of views thus making them most informative for policy makers in healthcare (Weinstein et al., 1996). Although breast cancer affects both men and women, the screening programme in the UK is only targeted at women. This is because of the considerably higher prevalence rate in women (Anderson et al., 2010). It is therefore, in this context, more appropriate to elicit preferences from women. A DCE eliciting preferences of men for a non-existent and unplanned screening programme would be neither useful, nor informative, for decision-makers.

In Table 3.3, one DCE (X5) was described which elicited preferences from people who were at an increased risk of developing cancer. However, this DCE was in the context of a service providing surveillance of a rare disease. As the NHSBSP routinely invites all women, it is difficult to rationalise why higher-risk women's preferences would be more important.

Female members of the public are both tax-payers, and potential patients, because the NHSBSP as it currently stands will invite all women in England once they reach the age of 50 years. For the purpose of collecting preferences about breast screening, female members of the public were defined as the relevant study sample.

The key premise for the selected study sample is underpinned by guidelines published by NICE for the implementation of more commonly-used stated preference methods (time trade-off and EuroQol (EQ-5D)) that advocate preferences from a sample of the general public (NICE, 2013).

3.5.3 Identification of attributes and levels

The next stage in designing a DCE is to identify the characteristics of the good or service which drive demand. The identification of attributes and levels began as an iterative process of interviews with clinical experts (n=4), a patient representative (n=1) and female members of the public (n=4), and literature reviews based on the attributes identified in the rapid review of section 3.4.6.

3.5.3.1 Attributes

Breast screening is made up of positive and negative attributes which women must tradeoff in their decision to attend the programme (see section 3.3). A DCE comprising few attributes was desirable so the effect of risk communication on choice could be investigated thoroughly while being cognisant of the potential cognitive burden of the task. A DCE needs to provide sufficient information about an intervention for respondents to make choices. The selection of attributes for the breast screening DCE balanced the requirement of a simple choice set with the need to generate informative preference data. Commonly occurring attributes in existing cancer screening DCEs were identified (see section 3.4.6): test accuracy; discomfort or pain from screen; risk reduction from participation; location; test provider; cost; and frequency of the programme. Each of these identified attributes was discussed extensively with the supervisory team, breast screening experts and patients (n=5), and female members of the public (n=4).

The accuracy of the screening test was the most commonly occurring risk attribute included in the identified screening-related DCEs. Accuracy was frequently described in terms of the test's specificity (risk of a false-negative) and sensitivity (risk of a falsepositive). The specificity of a breast screening programme is the risk that the mammogram misses a true cancer. In the case of breast screening, the risk of missing a true cancer is vanishingly small and the problem of over-diagnosis is much greater. For this reason, test specificity was not included as an attribute in the DCE.

In breast screening, the sensitivity of the programme is most often described as a risk of either over-treatment or over-diagnosis. The definition of over-diagnosis varies as experts debate the inclusion of slow-growing cancers which are not always distinguishable from dangerous tumours and that of biopsies with negative results (Carter et al., 2015). In this DCE, the appropriate attribute selected was the risk of unnecessary treatment where women are called back for something which would not have harmed them and this could include repeat scans, biopsies which identify no cancer, and treatment for slow-growing tumours.

As mammograms are known to be uncomfortable (Sapir et al., 2003), pain would be an obvious attribute to include. However, pain is a difficult concept to communicate and suffers from subjective interpretation and although scales exist to communicate pain such as the Brief Pain Inventory (Cleeland & Ryan, 1992), it was felt that discomfort from screening would have induced another framing issue in addition to the risk attribute. In addition, the pain a woman might experience from having a mammography will be dependent on the equipment used, the skills of the radiographer and the woman's physiology (Sapir et al., 2003). It was decided, therefore, that explaining what a mammogram involves, including the possible pain involved, in the training materials at the start of the DCE would be more beneficial than including pain as an attribute of a programme.

There is no potential reduction in the risk of developing breast cancer from participating in a breast screening programme. Screening women does not prevent the onset of breast cancer and the incidence rates of developing breast cancer are the same whether a women is screened or not. Therefore an attribute for reduced risk of cancer is not appropriate. However, the long term health-related outcomes of breast cancer have been shown to change as a result of screening, with some studies concluding that women who participate in screening programmes are less likely to die of breast cancer, and they are also less likely to require a mastectomy (Berry et al., 2005).

The potential inclusion of an attribute to capture the risk of mortality from breast cancer was discussed with expert clinicians (n=4) at the Nightingale Centre, Wythenshawe Hospital, but they strongly believed it was not a relevant motivator of screening attendance. The experts stated that from their experience, women attended screening simply because they wanted to detect cancer but mortality was too long-term for them to seriously consider. For this reason, the probability of detecting cancer by screening was identified as an appropriate attribute to explain the benefits of breast screening. This attribute allowed investigation of a positive risk; the probability of a favourable event (a benefit) occurring.

Three of the screening DCEs identified in rapid review of section 3.4.6 included the process attributes relating to the provider of the programme or location of the programme (studies X14, X18 and X19). As the research questions relating to this DCE were to investigate the communication of risk attributes, the outcomes of screening were seen to be more important and therefore attributes relating to the provider were not included. It was feared that additional attributes may add to the cognitive load of respondents and potentially detract from the main attributes of interest (risk).

A cost attribute is useful to include in a choice experiment as it allows monetary valuations and welfare calculations. To understand if, and how, elicited valuations differ between risk communication formats; a cost attribute was included. Cost is a notoriously difficult attribute to frame when respondents are used to consuming healthcare free at the point of use (Johnson et al., 2011). The cost attribute was described as out-of-pocket expense of attending the screening programme. The cost attribute was framed with the aim of reducing the incidence of protest bids or zero-response bias.

Frequency of screens occurred as an attribute in nine of the screening DCEs identified in the rapid review of 3.4.6. The frequency of screens is potentially important to women as the procedure is uncomfortable and inconvenient. Preferences for screening intervals are also of interest to policy makers with trials examining the effect of biennial screening. DCE responses could help understand how this might affect demand (Moss et al., 2006). Including screening frequency was problematic in this DCE as it would be directly correlated with the cost of screening attribute. Pre-pilot exploratory interviews with four female colleagues revealed that women were confused by the inclusion of screening frequency. The pilot respondents would try to identify the 'best value' programme by dividing or multiplying the cost by the frequency.

To reduce strategic behaviour, the programme was described as lasting twenty years (in line with current policy) with screening occurring five times over that period. This screening frequency allowed the cost to be presented over a lifetime (starting from the age of 50 years), and the cost per screen. Using this approach prevented women from making frequency-based calculations in the stated preference exercise.

The three attributes chosen for inclusion in the pilot study were:

- 1) Probability of detecting a cancer by screening over a lifetime
- 2) Risk of unnecessary treatment over a lifetime
- 3) Out-of-pocket cost of the programme (per screen and over a lifetime)

3.5.3.2 Levels

Assigning levels to these attributes describing the risks and benefits of screening, was challenging. The national breast screening programme has a high uptake rate which covers a large proportion of the population (The NHS Information Centre Screening and Immunisations, 2012). It is therefore difficult to calculate the baseline rates of benefits and risks for an unscreened population given the challenges of accounting for selection bias from women who opt-out. For example, the women who choose not to attend may be at particularly low risk or attend for mammograms privately. The range of levels used reflected estimates identified from the literature. The exact figures used in the levels depended on the perceived definition, population sample and whether the data were collected from randomised controlled trials or observational studies.

The attribute 'probability of detecting a cancer' was assigned levels based on a study by Welch & Frankel (2011). The actual value depends on age, race and family history. It was possible, however, to identify estimates for the average woman entering screening at aged 50 years with a 20 year probability of detecting a cancer set at about 3.5%. This estimate was calculated as the product of the risk of developing breast cancer (approximately 6%) and proportion of breast cancers found by mammography (approximately 65%).

The levels chosen for the attribute reflecting the probability of detecting a cancer by screening were: 3%, 7%, 10% and 14%. If women who perceived themselves to be at a

low risk of developing breast cancer, opted out of screening, then the proportion of breast cancers found by mammography could be underestimated and the 3.5% probability calculated by Welch & Frankel (2011) would be a conservative estimate hence representing the lower bound of the level range. A higher range for the level was set by reflecting detection rates achievable through stratified or more frequent screening (Hall et al., 2014).

The attribute 'risk of unnecessary treatment' was also assigned levels from a literature review. The levels selected represent the range of views that centre on the definition of what constitutes an over-treatment or over-diagnosis (see section 3.4.2). The Independent Review of Breast Cancer Screening (2012) estimated that *"just over 1% of women invited for screening over a twenty year period would have an over-diagnosed cancer"* (p.1778). This estimate only incorporated over-treatment for cancer and did not include women who underwent unnecessary biopsies. A review of the Norwegian screening programme estimated false recalls after mammography to be approximately 20% (Hofvind et al., 2004). The final levels used for the attribute reflecting the risk of unnecessary treatment were: 1%; 5%; 10%; and 20%.

Four levels of out-of-pocket costs were used: £20 per screen (£100 over a lifetime); £50 per screen (£250 over a lifetime); £150 per screen (£750 over a lifetime); and £200 per screen (£1,000 over a lifetime). The lower bound represents the lowest amount of out-of-pocket expenses, associated with taking time off work and travelling to a screening centre. The upper-bound was set as a realistic maximum based on the price of private mammograms in the UK (BreastHealthUK, 2014). The full list of attributes and levels selected for piloting in the DCE are shown in Table 3.4.

Attribute	Definition	Levels
Probability of detecting a cancer	The chance of detecting a cancer from screening over a 20 year period	3%; 7%; 10%; 14%
Risk of unnecessary treatment	The probability of being recalled for a procedure or procedures when no harm existed over a 20 year period	1%; 5%; 10%; 20%
Out-of-pocket cost of screening over a lifetime	The costs of attending the programme including original screens and recalls. These could include transport, time-off work, carer costs.	£100; £250; £750; £1,000

Table 3.4: Attributes and levels used in pilot DCE

3.5.4 Choice format

There are advantages and disadvantages of labelling the alternatives in a DCE compared with using a generic (un-labelled) question format. Kruijshaar et al. (2009) established labelling alternatives in a DCE added realism to the choices but de Bekker-Grob et al. (2010) found that labelling a DCE detracted respondents' attention from the attributes. In this thesis, the key research questions relate to the communication of risk attributes and anything that detracted from these would have been a detriment to the study. In addition, for this example, screening is conducted via a mammogram and there are no immediately apparent labels for different screening programmes. Therefore an unlabelled DCE was considered appropriate for this case study, and given the research question, thus the programmes took a generic (un-labelled) name of programme A or programme B.

Recent estimates of breast screening in the UK suggest uptake to the national programme is approximately 75% (The NHS Information Centre Screening and Immunisations, 2012). Whilst this is relatively high compared with other screening programmes such as bowel cancer (Weller et al., 2006), there is still a large number of women who are invited but do not attend. These uptake figures imply that some women have revealed 'no demand' for the programme, and consequently, a forced choice experiment would be inappropriate. Therefore an opt-out alternative of 'No Screening' was included in addition to the two programmes, A and B, in the pilot study.

3.5.5 Experimental design

The experimental design refers to the generation of alternatives and choices sets from a selection of attributes and levels. A full factorial design for a DCE of three attributes, each containing four levels would result in 64 possible profiles. This would be too many alternatives for any one person to consider, and would likely result in respondent fatigue. Therefore, the DCE aimed to use a fractional factorial design but still be capable of creating precise parameter estimates.

Using all the possible combinations of levels can create many profiles, and too many profiles can be expensive to distribute, demanding for respondents and time-consuming to analyse (Ryan & Watson 2011). To reduce this full factorial to a more manageable size, the profiles can be blocked or fractional factorials can be designed (Street et al., 2005). Blocking involves splitting the profiles into different chunks and creating multiple questionnaires (Ryan et al., 2008). However, if one questionnaire block suffers from a particularly poor response rate, it can lead to an inconclusive analysis (Viney et al., 2002; Witt et al., 2009).

Reducing a full factorial design to a fractional factorial design may involve assuming that all interactions are negligible or zero, and therefore only the main effects are estimated. It could also allow for a select few interactions to be present rather than estimating them all (Viney et al., 2005). A fractional factorial could be created by selecting profiles at random from the full factorial, but more systematic methods that ensure efficiency are usually employed to ensure an orthogonal or D-efficient design (de Bekker-Grob et al., 2012).

In orthogonal design plans, there is level balance as all attribute levels occur with equal frequency and there is no correlation between the attributes (Street & Burgess, 2007). This level balance allows identification of the effect of variables in levels of all effects independently as the attributes of the design are statistically independent, and minimises the variance in the parameters (Ryan & Watson 2011). These inherent characteristics are present in the fractional factorial design plans available in the Hahn & Shaprio (1966) catalogue.

If there are unequal numbers of levels for each attribute or the combination does not appear in the Hahn & Shaprio (1966) catalogue, then it is possible to overestimate and 'collapse' at the expense of level balance (Ryan & Watson 2011). If three attributes have four levels and one attribute has two levels, then it is possible to follow the design for a four attribute, four level plan with the two level attribute appearing twice as often, at the expense of level balance.

The other option is to employ a D-efficient design. D-efficiency involves minimising the inverse of the variance-covariance matrix in a maximum likelihood estimation (Johnson et al., 2013), and often involves the use of a software algorithm (using programmes such as Ngene®) to search for a design which maximises this efficiency (de Bekker-Grob et al., 2012). Previously, researchers have generated D-efficient designs assuming that parameters are equal to zero, however, using a Bayesian approach and incorporating prior assumptions (expected values) to improve statistical efficiency is a recent advance that is increasingly being used (de Bekker-Grob et al., 2012).

Ultimately, D-optimal designs (based on minimising the D-error) are a balance between statistical and response efficiency (Johnson et al., 2013). However, when designing a DCE there is a trade-off in allowing for estimation of effects thought to be important, and a design efficient in terms of allowing estimates from a smaller sample size. The method chosen will depend on the research question and the resources available for the study.

From the profiles suggested by the mathematical design, the choice sets of alternatives for the DCE need to be created. A recent systematic review by Clark et al. (2014) found most

authors now use software to design their DCEs. One popular package, Ngene®, is a mathematical software used to design choice experiments for given design criteria (such as orthogonality, level balance, blocking or validity tests). The software is then able to calculate the most 'efficient' design depending on the criteria set (for given sample sizes and model specification). Ngene® identifies the design from algorithms that use nested Monte Carlo simulation to select the most efficient design after running an extensive number (~2 million) of iterations where each iteration produces a design with the predefined number of choice situations. The software automatically compares each design produced from each iteration using the D-error, computed on the basis of expected values (priors) of the predicted model parameters.

The pilot DCE was constructed using a Bayesian efficient design. This involved defining priors for each of the main and interaction effects predicted to be in the final estimation. As neither the main effects nor interactions were perfectly known, point estimates were specified with a normal distribution in the design algorithm explained in the software syntax (see Figure A1 Appendix 3.5). The software package Ngene® (Choice Metrics, 2012) was used to generate a design minimising the D-error for a MNL. The initial pilot design in Ngene® used 2,000 Halton (random) draws to sample from the distribution of priors. This approach incorporated the uncertainty around the priors, and allowed deviation from the expected values whilst still ensuring efficiency of the design was maximised.

For the pilot study, four blocks of ten choices were chosen as ten choices seemed an appropriate starting amount for the number of questions to use in the pilot study. de Bekker-Grob et al. (2012) found over three-quarters of DCEs used fewer than 16 choice sets to reduce the cognitive burden of the exercise. Ngene® randomly selected from the full factorial design (64) which was defined by the number of attributes (three) and levels (four). The pilot design can be found in Table A2, Appendix 3.5.

3.5.5.1 Validity tests

The reliability of DCE data may be demonstrated by proving responses are consistent and respondents have made informed choices (Telser & Zweifel, 2007). However, validity of data can be tested through tests within the choice set and by comparisons with revealed preference data in internal and external validity checks, respectively.

One way of ensuring the results of a DCE are internally valid is through the inclusion of tests within the survey. Internal validity tests involve the development of specific choice sets to check that the respondents' choices exhibit the properties of: transitivity where if A is better than B, and B is better than C, then A is preferred to C; monotonicity where if A is

preferred to B and B is at least as good as C, then A is also preferred to C; stability where if A is preferred to B in one set, then A is preferred to B in another set too; and dominance when the respondent chooses an option that is superior in all ways (Johnson et al., 2009; Varian, 1992).

Tests for behaviour in line with these axioms can be incorporated in the design of a DCE to check for its internal validity. One of the fundamental and most common tests is for monotonicity (de Bekker-Grob et al., 2012). Each of the blocks in the DCE contained a check for internal validity to verify whether the respondent was answering in line with expectations and theory underpinning monotonicity. The test included a choice-set where one alternative had a higher probability of detection a cancer; a lower risk of unnecessary treatment; and a lower cost of screening. Failure of the test would not necessarily mean the respondent was irrational, but could instead indicate further explanation about the health intervention or choice experiment was required.

3.5.6 Survey materials

The DCE choice questions were completed as part of an online survey which included information (training materials) about the screening programme and an explanation of the forthcoming task. After the choice task, the respondents were asked for background information which might be useful to understand if, and how, different women gave different responses. The development of the preceding training materials is described in section 3.5.6.1 and the background questions are explained in section 3.5.6.2.

3.5.6.1 Training materials: pilot study

The information preceding the choice questions in a healthcare DCE survey is an important stage of informing the respondents about the disease area and proposed good or service. Whilst the information should be accessible to all potential respondents, it should also be interesting enough to keep their attention. As a starting point, breast screening pamphlets distributed as part of the NHSBSP were examined to extract appropriate terminology and phrasing (Slaytor & Ward, 1998; Cancer Research UK, 2013). Publically-available information resources on Public Health England's website (Public Health England, 2013a) were also consulted to create a comprehensive yet simple introduction to breast screening and the choice task. The information was supplemented with Figure 3.1 to show women the potential progression through screening and the related risks and benefits in a visual way.

3.5.6.2 Background questions: pilot study

Factors which could influence women's preference for breast screening or their interpretation of risk, were collected for possible use in a subgroup analysis or as covariates. Socio-demographic questions were taken from a review of established surveys (such as the Census and Understanding Society) which informed the context, question frame and answer categories (see Appendix 3.6).

In addition to socio-demographic questions, the survey also assessed respondents' numeracy by using the standardised questions identified by the structured review described in section 3.2 and Appendix 3.1 (see Figure A3). In addition to questions on numeracy, their experience of probabilities and their perception of risk in general were captured through Likert scale questions. Examples of the additional questions can be found in the pilot DCE shown in Appendix 3.8. The order of the questions in the survey reflects their importance. As respondents could drop out, after an initial screen to confirm the participants wished to continue, and a verification of their age for sampling, the first questions respondents answered were the choice sets.

The background questions were then ordered by their perceived importance for the analysis to capture as much data as possible if a respondent dropped out. Socio-demographic questions came first and questions on their feelings about probability came last.

3.5.7 Ethical approval

On the 18th September 2013, ethical approval was granted by The University of Manchester Research Ethics Committee (see Appendix 3.7).

3.5.8 The piloting process

After receiving ethical approval, a preliminary survey was created using online Sawtooth software (Sawtooth, 2012). The software provides an alternative to website development by providing a platform to enable survey webpages and choice set-style questions. The survey was hosted online using The University of Manchester's secure servers and databases to store survey responses. The software's inbuilt functions for design and analysis were not used in this PhD due to the availability of specialised, and more sophisticated, packages such as Ngene®. Starting with the pilot DCE (see Appendix 3.8), an iterative piloting process was used. The stages of piloting described in sections 3.5.8.1 to 3.5.8.2 were concurrent processes of changing and re-testing. After approximately ten drafts and re-drafts, a satisfactory final DCE was developed.

3.5.8.1 Experts and healthcare professionals

The first stage of piloting involved confirming that the attributes and levels selected were appropriate for a breast screening DCE. Whilst the design stages involved expert consultation, the complete pilot was also tested with experts at the Nightingale Centre, Wythenshawe Hospital and genetic counsellors at St Mary's Hospital, Manchester Royal Infirmary (n=4). In addition, a link to an online version of the survey was distributed to attendees (n=11) of a Genesis Breast Cancer meeting held in December 2013 in Manchester.

The discussion with four experts identified a need for a change in terminology from 'unnecessary treatment' to 'unnecessary follow-up', as they thought this was a more accurate description of the recall process. The experts also suggested including a set of validated questions on risk perception and concerns about developing breast cancer (see Evans et al. 1994; Evans et al. 1993; Hopwood et al. 2001). The importance of religious questions and feelings of fate, already incorporated in the questionnaire, were emphasised as important by the experts.

3.5.8.2 Qualitative interviews

In-depth face-to-face interviews using the think-aloud method (see Chapter Five) were conducted with female members of the public (n=4) and a patient representative (n=1). Women were recruited with an online advertisement placed on The University of Manchester's Research Ethics Website (see Appendix 3.9). Women who expressed an interest in the study were given an information sheet (see Appendix 3.11) to read over before the interview was arranged. As a thank you, the interviewees were given a £10 Amazon voucher after the interview was completed. The interviews took place on The University's Main Campus once the women consented (see Appendix 3.12 for the consent form). Contact with the patient representative was made via a clinical expert involved in the piloting.

The aims of these interviews were two-fold: 1) to pilot the DCE survey (see Appendix 3.8); and 2) to test the think-aloud interview schedule (see Appendix 3.10). Although interviews were recorded, the qualitative data collected was not formally transcribed and instead the field notes were used to make changes to the DCE and the preliminary interview schedule.

It became immediately apparent that the younger members of the public were not familiar with the screening programme and the procedure of a mammogram. Therefore a more interactive resource was sought to maintain the respondents' attention. The NHS Choices' website video describing the screening programme was selected to add into the training material for the main DCE survey. The video was edited to remove any persuasive language which may have swayed women towards screening, and instead provided a factual account of the process. Use of the video was approved by NHS Choices (see Appendix 3.13).

The 'No Screening' option appeared to be ignored by the interviewees in piloting. A change to the DCE was made. The selected levels for the 'No Screening' option were explicitly stated in the main DCE survey: a zero probability of detecting a cancer through screening; a zero probability of having an unnecessary follow-up from screening; and no cost, if the respondent opted out.

It was also apparent some interviewees found thinking-aloud difficult. After contacting an author of a think-aloud study identified in the systematic review presented in Appendix 5.1 (personal communication with Dr Sudeh Cheraghi-Sohi in July 2013), a warm-up exercise was included in the main think-aloud study. The warm-up exercise asked respondents how many windows they had in their house, which helped them to verbalise their thoughts, even if it felt unnatural.

The pilot interviews also provided an insight into the practicalities of recruitment and to an approximate interview length, which was around 45 minutes.

3.5.8.3 Internet panel: pilot study

To rapidly acquire responses to test the priors used in the experimental design, the DCE was piloted with an internet panel. The panel involved an initial sample of fifty-six women who were recruited via the internet panel provider, ResearchNow®. The sample was restricted to women aged 50 years and over to acquire information about the perspectives of older women not captured by the pilot interviews. The pilot internet panel also allowed for potential technical complications associated with hosting an online survey and collecting data to be identified and eliminated prior to the main survey. For example, the generation of different pass-in/pass-out IDs; survey redirect links for participant payments; and an idea of the ResearchNow® referral and response rate.

The pilot data were analysed using conditional logit and heteroscedastic conditional logit models. Details of the results of the internet panel pilot can be found in Appendix 3.14. All coefficients were significant and had the expected signs. However, a feature of the design

meant linearity of preferences in risk and detection could not be tested as a 0% level only appeared in the 'No Screening' option. This was changed in the main survey design.

The alternative specific constant (ASC) on 'No Screening', represent the mean of the distribution of the unobserved effects in the random component, ε , was insignificant in the pilot study. Free-text comments from the survey indicated that some respondents were giving protest responses that occurred in women who objected to the idea of paying for screening. Other women were sceptical of the purpose of the survey, possibly because they were used to completing surveys for private companies. Therefore, the main survey reiterated that screening would remain free on the NHS and that the survey was being used for the purpose of a PhD research study. A few women also commented that they were unsure what the word 'attribute' was referring to. The numeracy questions suffered from a particularly low response rate with many women skipping these questions (n=16, 26%).

A 0% level was introduced into the attribute risk of unnecessary follow-up. It was not introduced into the detection rate as this would generate inferior alternatives with no benefit from screening and yet potential costs (both financial and to health). The updated priors from the initial sample were then used to generate a new experimental design (with the additional level) following the steps described in section 3.5.5 which only required two blocks of ten choice sets. The final design included eleven choice sets including an additional dominant choice as a test of internal validity.

The new design was tested in a second pilot study using a sample of 58 women collected via internet panel provider ResearchNow® and also included defined levels for the 'No Screening' option. The training materials were changed to include the NHS video described in section 3.5.8.2, and it was explained explicitly that answers would be used only for university research. An additional question about women who had experienced screening and the cost incurred to them was added to provide information about real-life out-of-pocket costs incurred. The use of the word 'attribute' was removed and replaced with 'characteristic', and a survey logic was added which required women to give a reason for skipping the numeracy questions. The second internet panel experimental design, survey and results can be found in Appendix 3.15, 3.16 and 3.17, respectively. The second internet panel survey resulted in no additional changes to the survey or design.

3.5.9 DCE design: main study

The final attributes, attribute definitions, levels and specification of the opt-out alternative are shown in Table 3.5.

Table 3.5: Attributes and levels used in final DCE

Attribute	Definition	Levels for programmes	Levels for opt-out of 'No Screening'
Probability of detecting a cancer	The chance of detecting a cancer from screening over a 20-year period	3%; 7%; 10%; 14%	None: no cancers detected (0%)
Risk of unnecessary follow-up	The probability of being recalled for a procedure or procedures when no harm existed over a 20- year period	0%; 1%; 5%; 10%; 20%	None: no unnecessary follow- ups (0%)
Out-of-pocket cost of screening over a lifetime	The costs of attending the programme including original screens and recalls. These could include transport, time- off work, carer costs.	£100 (£20 per screen); £250 (£50 per screen); £750 (£150 per screen); £1,000 (£200 per screen)	None: no cost to you (£0)

An example of the final choice sets in each of the survey versions can be found in Figure

3.2, where risk was communicated as either a percentage or as a percentage and icon array.

Appendix 3.15 and 3.16 shows the experimental design and the DCE survey subsequently used in the empirical studies described in Chapters Four, Five and Six.

Figure 3.2: Example choice questions

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

(1 of 11 choice questions)

	Programme A	Programme B	No Screening
Women who will have cancers detected by screening	3%	14%	None: no cancers detected
Women who will have an unnecessary follow-up	1%	0%	None: no unnecessary follow-up
Out-of-pocket cost to you of screening programme per screen	£20 per screen (£100 over your lifetime)	£50 per screen (£250 over your lifetime)	None: no cost to you
	0	0	O

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

(1 of 11 choice questions)

	Programme A	Programme B	No Screening
Women who will have cancers detected by screening	3%	14%	None: no cancers detected
Women who will have an unecessary follow-up	1%	0%	None: no unnecessary follow- ups
Out-of-pocket cost to you of screening programme per screen	E20 per screen (£100 over your lifetime)	E50 per screen (£250 over your lifetime)	None: no cost to you
	\bigcirc	\bigcirc	0

3.6 Conclusion

Eliciting women's preferences for the benefits and risks of a NHSBSP was identified to be a useful case study for this thesis. Breast screening was found to be topical and welldebated, suggesting results from the DCE could be useful for decision-makers seeking to make changes to a breast screening programme. The topical nature also meant that it was a familiar disease area, which diminished the challenge of communicating an obtuse clinical topic to DCE respondents.

The DCE was extensively piloted with experts and female members of the public. The final pilot indicated that the attributes and levels selected were appropriate and the priors acquired informed an efficient design for the final study. Pilot work also ensured that the DCE was accessible to the public, confirming they understood the subject matter and the choice task. In the pilot, all attributes were estimated to be significant in women's decision to participate in screening and also had coefficient signs in line with expectations. The DCE design allowed for investigation of how a positively-framed risk (benefit) compared with a negatively-framed risk. The inclusion of the cost attribute allowed for valuation of these risk attributes using marginal WTP (MWTP). The following three chapters (Chapters Four, Five and Six) describe the empirical studies designed to investigate the effect of risk communication on individual's preferences and decision-making heuristics.

Chapter Four

Benefit-risk trade-offs for breast screening: results of the DCE

4.1 Introduction

This chapter uses econometric regression methods to quantify women's preferences for the benefit-risk trade-offs associated with breast screening and assess the effect of risk communication format on these preferences. The chapter addresses three main objectives: 1) to identify the trade-offs women are prepared to make in a NHSBSP using the DCE designed in Chapter Three; 2) to investigate whether the framing of risk affects their benefit-risk trade-offs made; and 3) identify the presence and determinants of preference heterogeneity. The chapter starts with a description of the key research questions in section 4.2. In section 4.3, the methods for: 1) data collection; and 2) the analysis are described. In section 4.4, the results of the DCE study are presented, and in section 4.5 the key findings, implications and study limitations are detailed. The chapter concludes in section 4.4.

4.2 Aims and objectives

The aim of the large scale DCE study was to identify if, and how, the communication of risk attributes affect women's preferences for breast screening. The sample used the final DCE designed in Chapter Three which can be found in Appendix 3.15. The design included 11 choice sets, with a test for internal validity, and a two-block design generated using Ngene® (Choice Metrics, 2012). To investigate whether the risk communication format had an effect, the design included two experimental conditions: percentages only (PO); or icon arrays and percentages (IAP). The study sought to address a number of key research questions:

- 1) What is the structure of women's utility for breast screening?
- 2) Is the randomness of choice affected by the framing of the risk attributes?
- 3) After accounting for randomness, do individual's preferences differ with the framing of risk?
- 4) Do groups of women have similar preferences in the form of latent-classes?
- 5) What are the predictors of class membership?
- 6) What are the marginal rates of substitution (MRS) between the attributes?
- 7) How do changes in risk affect the probability of a woman attending screening?

The research questions were answered using a variety of methods discussed in section 4.3.

4.3 Methods

The following sections describe the methods used to elicit and analyse preference data. The recruitment of the study sample is described in section 4.3.1. The models chosen to answer the research questions stated in section 4.2 are described in section 4.3.2. In section 4.3.3, other analyses are presented including calculations to determine predicted uptake and the MRS between attributes.

4.3.1 Data collection

An online survey was designed using Sawtooth software version 8.3.8 (Sawtooth, 2012) and uploaded to the internet as described in Chapter Three. Women were recruited to complete the survey via an internet panel provider, ResearchNow®. Internet panels were identified as a suitable source of achieving high quality yet low cost responses, with studies showing that this approach yields reliable response data efficiently (Ščasný & Alberini, 2012; Mulhern et al., 2015). Women who completed the survey received 'points' from the internet panel provider which they could accumulate and exchange for gift vouchers.

The first page of the survey (as shown in Appendix 3.16) explained the task and provided the opportunity for the respondents to opt-out of the study. The women were then randomly allocated to receive the DCE with either the PO or IAP risk format. After the desired sample of women was acquired, the survey was closed to prevent new entrants.

4.3.1.1 Recruitment

As a guide to DCE sample size, Orme (1998) suggested the power calculation presented in equation 4.1. The required sample size depends on the number of choice tasks (T), the number of alternatives in a choice set (A), and largest number of levels in any attribute (l):

Sample size
$$> \frac{500l}{TA}$$
 [4.1]

In this DCE, the Orme (1998) power calculation would suggest a minimum sample size of 152 (allowing for tests of the two risk communication formats). Despite the calculation by Orme (1998) suggesting a sample size of 152, this study aimed for a much larger sample. The large sample size was in response to literature suggesting preferences for healthcare are highly heterogeneous; even more so than for other areas of DCE research (such as transport and environment) (Fiebig et al., 2010). As a result, the sample size was constrained only by financial resources and instead the aim was to achieve as many views and choices as possible.

Female members of the public were identified as an appropriate sample in section 3.5.2 of Chapter Three. Quotas for each of the age bands selected for the survey (see p.351 of Appendix 3.16) were used to select a sample. These age bands were identified after the discussions with experts (n=4) described in section 3.5.3 of Chapter Three. The views of women currently in screening (who may have better formed preferences) and women about to enter screening (with regards to upcoming changes in policy) were felt to be of particular interest to the screening experts interviewed. Therefore the last two age bands (45 to 49 years and over 50 years) were over-sampled.

4.3.1.2 Data storage and transformation

The choice data generated were anonymous and stored as the survey progressed using The University of Manchester's secured server. After the sample was collected, the data were downloaded and transformed from comma-separated values (CSV) files into STATA® (StataCorp, 2011) .dta files and LatentGold® (Statistical Innovations, 2013) .lgf files.

4.3.1.3 Coding

The data were coded to rescale some attributes, such as cost, to make the coefficients more easily interpretable by reducing the number of decimal places. In addition, both effects coding (EC) and dummy coding (DC) were used, depending on the model specification.

Dummy coding involves dropping a level of an attribute and transforming the other levels into individual (dummy) variables which are equal to one when the level is present or zero if it is absent. The effects of the dummies are then comparable to the dropped level, although the effect of this level is consequently captured by the constant term and so the estimated coefficients are consequently correlated with it (Bech & Gyrd-Hansen, 2005). Effects coding allows for more information to be gained about the dropped level and eliminates the identification problem that occurs when the dropped level is incorporated within the intercept (Bech & Gyrd-Hansen, 2005). Effects coding allows the attribute interaction effects to be independently estimated whereas dummy variables would result in artificially high interactions and main effects because the utility associated with the dropped level is not distinguishable from the other elements of utility also incorporated in the intercept estimate (Ryan et al. 2008; Bech & Gyrd-Hansen 2005). The coding of the attribute levels are shown in Table 4.1.

Table 4.1: Attributes, levels and coding

Attribute	Label	Levels	Coding	EC1	EC2	EC3	EC4	DC1	DC2	DC3	DC4
Probability of	Detect	3%	3	-1	-1	-1		1	0	0	
detecting a cancer		7%	7	+1	0	0		0	1	0	
		10%	10	0	+1	0		0	0	1	
		14%	14	0	0	+1		0	0	0	
Risk of unnecessary	Risk	0%	0	-1	-1	-1	-1	0	0	0	0
follow-up		1%	1	+1	0	0	0	1	0	0	0
		5%	5	0	+1	0	0	0	1	0	0
		10%	10	0	0	+1	0	0	0	1	0
		20%	20	0	0	0	+1	0	0	0	1
Out-of-pocket cost	Cost	£100	1	-1	-1	-1		0	0	0	
of screening		£250	2.5	+1	0	0		1	0	0	
		£750	7.5	0	+1	0		0	1	0	
		£1000	10	0	0	+1		0	0	1	

EC: effects coding

DC: dummy coding

Other dummy variables were generated for key items of interest and to simplify some complex response options. For example, '*To what extent are you concerned about your own risk of breast cancer*?' gave women five options (not at all, a little, quite a lot, very much, no idea) so a dummy variable was generated to identify women who stated that they were not at all concerned. A dummy was also created to identify respondents who had failed the internal validity test, and an additional dummy variable was generated to identify the risk format the respondent had received. The dummy variables generated and their definitions and coding and presented in Table 4.2.

Table 4.2: Dummy variable definitions

Variable	Definition	Coding
IAP	Identifier for risk format respondent received	1 = IAP
		0 = PO
Wrong	Identifier for respondents who failed the internal validity	1= Failed test
	test by choosing the inferior alternative.	0 = Passed test
Unconcerned	Dummy variable collapsed from question: 'To what extent	1 = Not concerned about
	are you concerned about your own risk of breast cancer?'	risk
	to identify respondents who stated 'Not at all' about their	0 = Concerned/No idea
	risk of cancer.	
Employed	Dummy variable collapsed from question: 'What is your	1 = Employed
	occupational status?' to identify respondents in work	0 = Not employed
	(either employed full-time, part-time or self-employed).	
White	Dummy variable collapsed from question: 'What is your	1 = White
	ethnic group?' to show White respondents (either 'White	0 = Non-White
	British/Irish' or 'White Other').	

4.3.2 Models

Descriptive statistics were used as a preliminary exploration of the DCE choice data. Respondent characteristics were initially explored to determine whether any differences had occurred, by chance or error, in the randomisation process to the different risk formats or the two design blocks. After ensuring randomisation was successful, investigation of the research questions presented in section 4.2 began. The research questions were answered using multiple discrete choice models which are now explained.

The choice data were modelled using a random utility maximisation framework (McFadden, 1974). Using equation 4.2 as a starting point, an individual's (n) indirect utility (U) for an alternative (j) can be written as:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \qquad [4.2]$$

In equation 4.2, utility comprises a deterministic component (V_{nj}) and an unobservable, random component incorporating unobservable attributes (ε_{nj}) . The probability of an individual *n* choosing alternative *i* over alternative *j* is given by equation 4.3:

$$P_{ni} = Prob(U_{ni} > U_{nj} \text{ for all } j \text{ where } i \neq j) \qquad [4.3]$$

Assuming that ε_{nj} is iid extreme value with a variance σ^2 then a closed-form choice probability can be given by equation 4.4:

$$P_{ni} = \frac{exp^{V_{ni}}}{\sum_j exp^{v_{nj}}} \qquad [4.4]$$

Where V_{nj} , the deterministic component of utility is explained by:

$$V_{nj} = \beta X_{nj} \qquad [4.5]$$

In equation 4.5, β is a vector of parameters to be estimated and is the relative level of satisfaction for individual *n*, and X_{nj} is a vector of observed attributes for alternative *j*.

4.3.2.1 Discrete choice models

The first research question sought to understand the structure of women's preferences for breast screening. This involved: 1) assessing whether the effects of attributes on utility had the expected sign; 2) investigating whether an ASC was needed for any alternative; and 3) investigating whether there existed non-linearity in women's preferences for the risks associated with breast screening. To do so, a simple conditional logit model was used.

4.3.2.1.1 Conditional logit model

Model 1 was a conditional logit model. The conditional logit model is a simple discrete choice model which makes a number of assumptions. Combining equations 4.4 and 4.5 means the conditional logit model can be given by:

$$P_{ni} = \frac{exp^{(\beta X_{ni})}}{\sum_{j} exp^{(\beta X_{nj})}} \qquad [4.6]$$

The conditional logit model presented in equation 4.6 was used to a conduct a preliminary investigation into the structure of women's utility. It was hypothesised that the attributes 'risk of unnecessary follow-up' and 'cost' would have negative coefficients as more risk and more cost are intuitively seen as less desirable, and the attribute 'probability of detection' would have a positive coefficient. These *a priori* expectations did not influence the analysis, but provided a first step of checking that the DCE was theoretically valid through a simple model (Model 1A).

Conditional logit models were also used to make simple specification changes to the model. Equation 4.7 simply investigated the effect of each attribute on utility where $Risk_{nj}$, $Detect_{nj}$ and $Cost_{nj}$ were the exogenous attributes associated with alternative *j*, and ε_{nj} the unobserved random component which has a variance and a set of covariances.

$$U_{ni} = \beta_1 Detect_{ni} + \beta_2 Risk_{ni} + \beta_3 Cost_{ni} + \varepsilon_{ni}$$
 [4.7]

In this DCE, it could be expected that there may be a difference in ε_{nj} between the 'No Screening' option and Programmes A and B, as individuals may have gained utility from just participating in screening, beyond what was explained by the attributes. In equation 4.8, α_{n0} , an ASC was introduced. This ASC represents the mean of the distribution of the unobserved effects in the random component, ε_{nj} , associated with the opt-out for individual *n*. This ASC can therefore be interpreted as the average effect of the unobserved screening attributes on the utility of not participating in a programme relative to any form of screening. Equation 4.8 was estimated using a conditional logit (Model 1B)

$$U_{nj} = \alpha_{n0} + \beta_1 Detect_{nj} + \beta_2 Risk_{nj} + \beta_3 Cost_{nj} + \varepsilon_{nj}$$
 [4.8]

In equation 4.9, ASCs were introduced for Programme A and Programme B to investigate any left-right bias in respondents' choices.

$$U_{nj} = \alpha_{nA} + \alpha_{nB} + \beta_1 Detect_{nj} + \beta_2 Risk_{nj} + \beta_3 Cost_{nj} + \varepsilon_{nj}$$
 [4.9]

Equation 4.9 was also estimated using a conditional logit (Model 1C). A left (right) bias was present if α_{nA} was significantly higher (lower) than α_{nB} .

The functional form of women's risk preferences was investigated using Model 1D (a conditional logit estimation of equation 4.10). Assuming a linear utility function when preferences were non-linear may result in biased estimates (Torres et al., 2011). Therefore

a dummy for each level of risk was incorporated in the model. As a result, the utility from alternative j for individual n is:

$$U_{nj} = \alpha_{n0} + \beta_1 Detect_{nj} + \beta_2 Risk_1 n_j + \beta_3 Risk_5 n_j + \beta_4 Risk_1 0_{nj} + \beta_5 Risk_2 0_{nj} + \beta_6 Cost_{nj} + \varepsilon_{nj}$$

$$[4.10]$$

Comparing Model 1D with Model 1B allowed an investigation into the linearity of women's risk preferences. To do so, a likelihood ratio (LR) test was conducted, which tested whether Model 1B or Model 1D was best-fitting. If Model 1D was best-fitting, this would indicate that women's risk preferences are non-linear. In addition, non-linearity was tested by plotting the coefficient of the dummy variables representing each level of risk in Model 1D. Investigation into non-linearities in preferences for the other attributes could not be conducted as the 0 levels (0% of detection and £0) only occurred in the 'No Screening' option and were therefore perfectly collinear.

4.3.2.2 Testing icon array effects

Research questions two and three were concerned with the effect of risk format on the randomness of respondents' choices and on utility parameters (individuals' preferences), respectively. Under the conditional logit model described in section 4.3.2.1, the error variance across individuals was assumed constant. That is to say, choice consistency was the same across choices (and the scale parameter was normalised to one). However, this DCE may have unequal error variance as respondents received different risk communication formats, depending on the DCE they were randomised to. This would mean that differences in estimated parameters in the conditional logit models across risk formats would not reflect true differences in preferences, as these differences would also capture differences in the scale parameter (choice consistency) across groups. To allow for this, investigations into scale heterogeneity were made in order to answer research question two: is the randomness of choice affected by the formatting of the risk attributes?

4.3.2.2.1 Impact of risk format on choices: preferences and scale.

The choice sets the respondents completed were identical in all attributes and experimental design resulting in no other context-specific components to the utility functions described in section 4.2.1.1. However, how the risk communication formats could have affected respondents' decision rules. The structured review of risk communication formats presented in Appendix 3.10 found icon arrays as a preferable risk communication format which could translate into different stated preferences (as respondents' can make choices

which better reflect their preferences as they are more informed) or differences in scale (as task complexity is reduced and, as a consequence, choice consistency improves).

To test if the risk communication format affected the utility parameters (preferences), the null hypothesis that preferences were the same between the two communication formats was H₀: $\beta_{PO} = \beta_{IAP}$, where β_{PO} represents the estimate utility parameters for those receiving the percentages only risk format, and β_{IAP} represents the utility parameters for those receiving icon arrays and percentages. However, estimation of the conditional logit models described in section 4.3.2.1.1 actually involved estimates of $\lambda_{PO}\beta_{PO}$, $\lambda_{IAP}\beta_{IAP}$ where λ was a scale parameter confounded with the β estimates. Therefore, with a standard conditional logit, it was not possible to simply compare model coefficients from the different risk formats as differences in parameter estimates could be due to: 1) differences in utility parameters (preferences) (the β s); 2) differences in consistency or scale factors (the λ s); or 3) both differences in utility and scale. Crucially, if the scale parameters differences in preferences.

Plotting β_{PO} against β_{IAP} results in a gradient $\frac{\lambda_{IAP}}{\lambda_{PO}}$ (the ratio of the error variance for PO respondents to the error variance for IAP respondents). If the points lie on a straight line, this would indicate a ratio of one and suggest that the scale parameters are approximately equal i.e $\lambda_{po} \approx \lambda_{IAP}$. However, the DCE used in this study had only three attributes (detect, risk, cost) which made parameter plots limited in their ability to confirm or disconfirm choice consistency (as there are only three data points to determine linearity). Therefore to understand whether choice consistency was affected by the risk communication format, heteroscedastic conditional logit models were used to estimate the size, direction and significance of the scale parameter. Introducing the scale parameter directly into the estimated utility function allowed for individual-level heteroscedasticity (variance in response error) as shown by equation 4.11:

$$P_{ni} = \frac{exp^{(\lambda_n \beta X_{ni})}}{\sum_j exp^{(\lambda_n \beta X_{nj})}} \qquad [4.11]$$

In equation 4.12, λ_n is a scale term which is allowed to vary by IAP_n , a dummy variable which equals unity if individual *n* received the IAP version of the DCE:

$$\lambda_n = \exp(\delta IAP_n) \quad [4.12]$$

 δ reflects the influence of the risk communication format on the error variance, and therefore the scale parameter. As shown in equation 4.12, for identification, the baseline

scale parameter is set equal to one for respondents completing the PO version of the DCE. Testing if $\delta = 0$ is a test of constant error variance across respondents from all versions. Accounting for scale heterogeneity in the heteroscedastic conditional logit means that any differences in estimated utility parameters across risk formats reflects true differences in preferences.

Therefore, to understand how the risk communication format affected any of the marginal utilities associated with an attribute, the data was pooled and equations 4.13 and 4.12 were estimated using a heteroscedastic conditional logit model (Model 2A):

$$U_{nj} = \alpha_{n0} + \lambda_n \beta_1 Detect_{nj} + \lambda_n \beta_2 Risk_{nj} + \lambda_n \beta_3 Cost_{nj} + \lambda_n \beta_4 Detect_{nj} IAP_n + \lambda_n \beta_5 Risk_{nj} IAP_n + \varepsilon_{nj}$$
[4.13]

In equation 4.13, $Detect_{nj}IAP_n$ and $Risk_{nj}IAP_n$ represent interactions terms between the probability of detecting a cancer attribute and the probability of unnecessary follow-up, with the dummy variable indicating the risk format received, respectively. No such interaction term was included for the cost attribute, as cost was framed in an identical manner in both versions of the DCE. As a result, to investigate differences in preferences due to the risk format (and therefore to answer research question three), the following hypothesis was tested: H_0 : $\beta_4 = \beta_5 = 0$ and H_1 : β_4 , $\beta_5 \neq 0$. To investigate whether risk format affected choice consistency (and therefore to answer research question two), the following hypothesis was tested: H_0 : $\delta = 0$ and H_1 : $\delta \neq 0$.

To investigate whether respondents who failed the internal validity test had different preferences and/or choice consistency, equations 4.14 and 4.15 were estimated using a heteroscedastic conditional logit model (Model 2B):

 $U_{nj} = \alpha_{n0} + \lambda_n \beta_1 Detect_{nj} + \lambda_n \beta_2 Risk_{nj} + \lambda_n \beta_3 Cost_{nj} + \lambda_n \beta_4 Detect_{nj} IAP_n + \lambda_n \beta_5 Risk_{nj} IAP_n + \lambda_n \beta_6 Detect_{nj} Fail_n + \lambda_n \beta_7 Risk_{nj} Fail_n + \lambda_n \beta_8 Cost_{nj} Fail_n + \varepsilon_{nj}$ [4.14]

$$\lambda_n = \exp(\delta IAP_n + \varphi Fail_n) \quad [4.15]$$

Equation 4.14 is identical to equation 4.13 but with additional interaction terms between all attributes and a dummy for whether an individual failed the internal validity test, $Fail_n$, and equation 4.15 is identical to equation 4.12, but with $Fail_n$ included as an additional predictor of scale. Hypotheses $H_0: \beta_6 = 0, H_1: \beta_6 \neq 0; H_0: \beta_7 = 0, H_1: \beta_7 \neq 0$; and $H_0: \beta_8 = 0, H_1: \beta_8 \neq 0$, were then tested to test whether failing the internal validity test affected women's preferences for the detection, risk and cost attributes, respectively, and

 $H_0: \varphi = 0$ and $H_1: \varphi \neq 0$ tested to test whether those who failed had less consistent choices.

All estimations of Models 1 and 2 were conducted in STATA® using maximum likelihood estimation (MLE) to acquire parameter estimates.

4.3.2.3 Investigating further preference heterogeneity through preference groups

Research questions four and five were concerned with establishing whether groups of women have similar preferences, and whether observables describing women different preference groups. The conditional and heteroscedastic conditional logit models presented previously were aggregated models which may hide underlying variation in preferences. The estimation of latent-class models allowed identification of subgroups with similar preferences by investigating preference heterogeneity amongst responses to the DCE.

The conditional and heteroscedastic conditional logit models are also based on the independent of irrelevant alternatives (IIA) assumption. Under IIA, it is assumed that the ratio of choice probabilities for any two alternatives is unaffected by the presence or absence of another alternative (Louviere et al., 2000). For example, assume an individual is faced between choosing a doctor and a hospital nurse, and that this respondent chooses between these options with equal probability, 0.5. If another option, say district nurse, was introduced the IIA assumption requires that this equal probability must be maintained, so the probability of choosing a doctor, district nurse and hospital nurse must all be 0.33. However, if the introduction of this new alternative resulted in probabilities of 0.5, 0.25 and 0.25 respectively, the IIA assumption would be invalid. In addition to allowing the investigation of further preference heterogeneity, latent-class models also relax this IIA assumption.

4.3.2.3.1 Latent-class analysis

The third estimated model (Model 3) was a latent-class model. In latent-class models, respondents are allowed to be probabilistically assigned to classes which differ in their utility functions. The choice probabilities and the marginal utilities are dependent on the class, *c*. The latent-class model is therefore specified as:

$$P_{ni|c} = \frac{exp(\beta_c X_{ni})}{\sum_{j=1}^{J} exp(\beta_c X_{nj})} \qquad [4.16]$$

In equation 4.16, $P_{ni|c}$ is the probability that an individual, *n*, chooses alternative *i*, given that this individual is in class, *c*. Class membership is modelled in MNL form:

$$P_{nc} = \frac{exp(Z_n\Phi)}{\sum_{c=1}^{C} exp(Z_n\Phi)} \qquad [4.17]$$

In equation 4.17, Z_n is a vector of individual respondent characteristics included as explanatory variables. In this study, details about the respondent characteristics were collected in the final questions of the survey (see Appendix 3.16, p. 368-385). The questions were chosen because they were believed to potentially influence a respondent's choices and were therefore incorporated as preference-class covariates. Estimates Φ indicate which of these covariates characterise difference preference groups. In addition, the *IAP_n* dummy, which represents the risk communication format received by the respondent, was also included as a preference-class covariate. Identification was achieved by imposing $\sum \Phi = 0$ so:

$$P_n = \sum_{c=1}^{C} P_{nc} P_{n|c} \qquad [4.18]$$

Equation 4.18 was estimated using MLE. The number of classes must be imposed exogenously, thus the total number of preference-classes, C, was varied and each model compared using information criteria to select the best fitting model via a calibration process. Information criteria are measures of the goodness of fit for maximum likelihood estimation models. The most widely used information criteria are the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and Consistent Akaike Information Criterion (CAIC). These criteria vary in their penalties (such as for sample size, number of parameters) so different information critera can support different models depending on their emphasis for sensitivity (having enough parameters to model relationships) or specificity (over-fitting) (Dziak et al., 2012). The results of the different information criteria can all be interpreted in the same way: the model with the lowest comparative value, the better (absolute values are meaningless). There is debate around which information criterion is preferable (Dziak et al., 2012) and models were also compared using the predictive error (how many choices were correctly predicted by the model).

4.3.2.3.2 Scale-adjusted latent-class analysis

The latent-class model was then extended to allow for each latent-class to comprise subsets of respondents with the same preferences but different error variances (choice consistency).

As in the conditional logit model, if the scale parameter differed between risk communication formats, then a standard latent-class model may not reveal true preference heterogeneity, as differences in estimated parameter across groups will reflect both differences in preferences and differences in the scale parameter (choice consistency). Scale-adjusted latent-class models were used to overcome this. Scale-adjusted latent-class models capture differences in the scale parameter by explicitly grouping respondents with similar choice consistencies. The scale-adjusted latent-class model was used to identify latent preference-classes (Pc) within which different scale-classes (Sc) with different scale parameters, λ , were present. Each of these Sc are associated with a scale-membership probability. Incorporating the scale parameter back into equation 4.16, yields equation 4.19:

$$P_{ni|c,s} = \sum_{s=1}^{S} \pi_s \frac{exp(\lambda_s \beta_c X_{ni})}{\sum_{j=1}^{J} exp(\lambda_s \beta_c X_{nj})}$$
[4.19]

Equation 4.19 was estimated using a scale-adjusted latent-class model (Model 4). π_s is the scale membership probability with $0 \le \pi_s \le 1$, $\sum_{s=1}^{S} \pi_s = 1$. Variables which could predict choice consistency could be included to predict scale-class membership. In this study, self-reported task difficulty and risk format were included as Sc covariates. The scale-adjusted latent-classes were estimated using the software LatentGold® (Statistical Innovations, 2013) and the number of scale and preference-classes were selected through comparisons of the BIC, AIC and CAIC. These information criteria were also used to compare these models against standard latent-class models.

4.3.3 Marginal rates of substitution

Research question six was concerned with MRS between attributes. The results of Models 1 to 4 provided information which could be used to investigate how respondents were prepared to trade between attributes through their MRS. The parameter estimates can be interpreted in terms of willingness to exchange money (WTP) or WTA risk. Using the cost attribute, the MWTP for particular changes in non-cost attributes was estimated by dividing the coefficient of the attribute of interest by the coefficient of the cost attribute. The MWTP for the attributes probability of detecting a cancer and risk of unnecessary follow-up were calculated using equations 4.20 and 4.21, respectively.

MWTP for a 1% *point increase in detection* =
$$\frac{\beta_{Detect}}{-\beta_{Cost}}$$
 [4.20]

MWTP for a 1% *point decrease in risk of unecessary follow*
$$-up = \frac{-\beta_{Risk}}{-\beta_{Cost}}$$
 [4.21]

The generated WTP values were then used to estimate an overall value of the screening programme. As identified in Chapter Three, there are challenges in identifying the current levels of each attribute in the NHSBSP as it currently stands due to extensive debates (see

section 3.4.2 of Chapter Three). Therefore, predicted WTP for a screening programme was calculated for different scenarios as detailed in Table 4.3.

Table 4.3: WTP for hypothetical	screening pr	ogrammes
---------------------------------	--------------	----------

	Screening worst-	Screening	Screening best-
	case	intermediate	case
Probability of detecting a	3%	10%	14%
cancer			
Risk of unnecessary follow-up	20%	10%	0%

The MRS between the attributes 'probability of detecting a cancer' and 'risk of unnecessary follow-up' show the benefit-risk trade-offs women were prepared to make for an NHSBSP. The ratio of coefficients of the probability of detection and risk of unnecessary follow-up ($\frac{\beta_{Detect}}{-\beta_{Risk}}$) represents the number of unnecessary follow-ups women are willing to accept for an additional cancer detected, per 100 women screened.

Research question seven sought to understand the external validity of the DCE. As there was no revealed preference data, uptake was estimated for the different screening scenarios presented in Table 4.3 and compared to current levels of screening uptake.

4.4 Results

This next section presents the results of the large scale internet panel DCE and utilises the Models 1 to 4 to answer the research questions stated in section 4.2.

4.4.1 Descriptive statistics

The final study comprised 1,000 women who completed the DCE survey entirely. 1,018 women completed the DCE choice sets, but 18 women did not finish all the background questions. Out of the 1,018 women who completed the choice questions, 507 people received the PO version and 511 people received the IAP version. Out of the 1,000 women who completed the DCE entirely, 498 people received the PO version and 502 people received IAP version.

It was difficult to attribute an exact response rate to an internet panel survey. However, 38,251 emails were sent from ResearchNow® and 23,082 of these were delivered. 2,205 women clicked on the link to the survey. 511 women clicked on this when their age quota was full and another 60 clicked on the link after the survey had closed. The response rate in terms of survey loads as a proportion of emails delivered was therefore around 9%. A full breakdown of the exact response rate can be found in Appendix 4.1.

Table 4.4 shows the sample characteristics for a number of key variables, for all respondents and the IAP and PO subsamples. A more detailed breakdown of the sample characteristics can be found in Appendix 4.2. No significant differences occurred in the characteristics between the two risk communication formats. A logistic regression was also used to check if any of the covariates predicted risk format membership but none were statistically significant (see Appendix 4.3). The computerised randomisation to each survey version of the DCE survey appeared to be successful.

	Overall		РО		IAP		
Age group	n	%	n	%	n	%	
18-24	104	10.2%	44	8.7%	60	11.7%	
25-34	103	10.1%	51	10.0%	52	10.2%	
35-44	253	24.9%	127	25.0%	126	24.7%	
45-49	253	24.9%	131	25.8%	122	23.9%	
50+	305	30.0%	154	30.4%	151	29.6%	
Occupational Status							
Employed full-time	343	34.0%	167	33.3%	176	34.7%	
Employed part-time	211	20.9%	114	22.7%	97	19.1%	
Self-employed	58	5.8%	31	6.2%	27	5.3%	
Unemployed/Long-term	68	6.7%	30	6.0%	38	7.5%	
Other: e.g Retired/ House wife	329	32.6%	160	31.9%	169	33.3%	
Education							
Less than 4 O-levels	150	14.9%	74	14.7%	76	15.0%	
5+ O-levels	112	11.1%	54	10.8%	58	11.4%	
NVQs/ A-levels	308	30.5%	165	32.9%	143	28.2%	
Undergraduate degree	270	26.8%	129	25.7%	141	27.8%	
Master's/ PhD	99	9.8%	52	10.4%	47	9.3%	
Other qualification	70	6.9%	28	5.6%	42	8.3%	

Table 4.4: Sample characteristics

4.4.2 Internal validity

As described in section 3.5.5.1 of Chapter Three, an internal validity test was included with a dominant choice option. In this study, 9% of the total sample failed this test. As shown in Table 4.5, out of the 90 respondents who answered incorrectly, 42 received the PO version and 48 received the IAP; this difference was not statistically significant (p=0.533). In

addition to the validity test, there was no statistically significant difference between the two formats with regards to self-reported task difficulty (p=0.640).

	Overall		РО		IAP	
Validity test	n	%	n	%	n	%
Failed	90	8.8%	42	8.3%	48	9.4%
Task difficulty						
Very easy	346	34.2%	172	34.2%	174	34.2%
Easy	299	29.6%	155	30.8%	144	28.3%
Neither easy nor difficult	260	25.7%	125	24.9%	135	26.5%
Difficult	97	9.6%	45	9.0%	52	10.2%
Very difficult	10	1.0%	6	1.2%	4	0.8%

 Table 4.5: Validity test failure and self-reported task difficulty.

With confidence that the randomisation procedure was successful, discrete choice modelling of the choice data continued, starting with the conditional logit models.

4.4.3 Investigating the functional form of utility

The conditional logit model results are shown in Table 4.6. The results of Model 1A show the attribute coefficients from the conditional logit were in line with *a priori* expectations with cost and risk having negative and statistically significant coefficients, whereas detection was positive and statistically significant. The negative coefficient on risk indicates that the higher the level of this attribute in an alternative, the less likely an individual was to choose this alternative. The positive coefficient for detection indicates that the women preferred a programme with higher levels of this attribute.

Table 4.6: Results of the conditional logit Models 1A, 1B and 1C

Model 1A			Model 1B			Model 1C		
ASC(on A)					1.588***	(0.05)		
ASC(on B)					1.534***	(0.05)		
ASC(on none)			-1.569***	(0.05)				
Detect	0.141***	(0.00)	0.080***	(0.00)	0.079***	(0.00)		
Risk	-0.022***	(0.00)	-0.047***	(0.00)	-0.046***	(0.00)		
Cost	-0.036***	(0.00)	-0.108***	(0.00)	-0.107***	(0.00)		
	Ν	Obs						
	1018	33594						

p*<0.005; *p*<0.01; ****p*<0.001; standard errors in parentheses
In Model 1B, the ASC for the option 'No Screening' was found to be statistically significant and negative. This suggests respondents were more likely to choose either A or B over no screening at all, indicating that women gained utility from attending screening beyond that explained by the attribute levels. In Model 1C the ASCs on Programme A and B indicate respondents were likely to opt-in but were no more likely to choose A over B, or *vice versa*, and indicates no left-right bias was present. A likelihood ratio test showed Model 1B offered better fit (p<0.01) than Model 1A, and thus further models incorporated the ASC for 'No Screening'.

In Model 1D the continuous variable of risk was recoded into levels (see Table 4.1). The results of Model 1D presented in Table 4.7 showed the coefficient on risk increased by approximately equal intervals with overlapping confidence intervals implying preferences for risk were monotonically increasing and could consequently be considered linear. A likelihood ratio test showed that allowing for non-linearity in the risk attribute in Model 1D did not increase fit (p>0.05), and thus further models assumed that a linear representation of the attributes was compatible with the data.

	Model 1D	
ASC(on none)	-1.624***	(0.06)
Detect	0.074***	(0.00)
Risk		
1%	-0.030	(0.04)
5%	-0.289***	(0.06)
10%	-0.355***	(0.05)
20%	-0.962***	(0.05)
Cost	-0.111***	(0.00)
	Ν	Obs
	1018	33594

Table 4.7: Results of the conditional logit Model 1D

* *p*<0.005; ** *p*<0.01; ****p*<0.001; standard errors in parentheses

In addition to the LR test, the parameter estimates were plotted which visually confirmed an approximately linear relationship (see Figures 4.1 to 4.3). A linear relationship was assumed for subsequent investigations.

Figure 4.1: Functional form of the risk attribute for DCE respondents



Figure 4.2: Functional form of the detection attribute for DCE respondents



Figure 4 3: Functional form of the cost attribute for DCE respondents



With confidence that the functional form was appropriately specified, more sophisticated models were estimated to investigate the presence of scale and preference heterogeneity to answer subsequent research questions.

4.4.4 Testing risk format effects

Estimation results from the heteroscedastic conditional logit models (Model 2A and 2B) are presented in Table 4.8. In Model 2A, all attribute coefficients were statistically significant and had the expected signs, again consistent with *a priori* expectations. All coefficients relating to the interactions between attributes and risk format, and the coefficient of risk format in the scale term were insignificant, suggesting that the risk format did not lead to any direct differences in either marginal utility or choice consistency, and that the format of risk had no effect in this study sample.

		Model 2A		Model 2B		
	ASC(on none)	-1.497***	(0.07)	-1.583***	(0.07)	
	Detect	0.081***	(0.00)	0.092***	(0.00)	
Utility	Risk	-0.047***	(0.00)	-0.060***	(0.00)	
-	Cost	-0.103***	(0.01)	-0.132***	(0.01)	
	IAP*detect	-0.010	(0.01)	-0.006	(0.01)	
	IAP*risk	0.005	(0.00)	0.004	(0.00)	
	Fail*detect			-0.086***	(0.01)	
	Fail*risk			0.094***	(0.01)	
	Fail*cost			0.222***	(0.02)	
Scale	IAP	0.094	(0.06)	0.063	(0.05)	
	Fail			-0.029	(0.13)	
		Ν	Obs		. ,	
		1018	33594			

Table 4.8: Heteroscedastic conditional logit results

* *p*<0.005; ** *p*<0.01; ****p*<0.001; standard errors in parentheses

However, in Model 2B the coefficients relating to interactions between attributes and failure of the validity test suggested respondents who failed the internal validity test made a significant difference to the parameter estimates. A likelihood ratio test comparing Models 2A and 2B showed Model 2B had statistically significant (p<0.001) better fit. However, responses of those who failed were included in the subsequent stages of analysis in line with recommendations from the literature (see Hougaard et al. 2012; Lancsar & Louviere 2006) and because their responses were no more random (with a statistically insignificant effect of 'Fail' on the scale parameter).

4.4.5 Investigating preference heterogeneity

With confidence that the data from each risk communication format could be pooled because of the insignificant scale and interaction terms for IAP in Model 2, latent-class models were estimated to explore preference heterogeneity further and answer research question four and five. In Table 4.9, a summary of latent-class models with different numbers of preference-classes and their associated information criteria and log likelihood (LL) values are presented. On the BIC and CAIC, the 6-Pc model was preferred, however, the 7-class model was preferred based on the fact it minimised the AIC. The AIC is more sensitive (lower false-negative rate) but BIC is more specific (lower false-positive rate). In other studies, BIC has performed better than AIC (which tended to over-fit the data) in tests (Nylund, 2007). Therefore, on the basis of the BIC result, a 6-Pc model provided the best fit.

Number of preference classes	LL	BIC(LL)	AIC(LL)	CAIC(LL)
1-Pc	-9925.1295	19877.922	19858.2591	19881.922
2-Pc	-7954.8521	16186.3331	15989.7042	16226.3331
3-Pc	-7485.4462	15496.4875	15122.8925	15572.4875
4-Pc	-7042.2923	14859.1456	14308.5846	14971.1456
5-Pc	-6804.6897	14632.9065	13905.3795	14780.9065
6-Pc*	-6661.3985	14595.29	13690.7969	14779.29
7-Pc	-6541.0412	14603.5415	13522.0823	14823.5415

Table 4.9: Information criteria for varying preference-class numbers in Model 3

*Chosen model

In the 1-Pc model, the prediction error was 0.4227, which meant that just over half (57.7%) of choices were predicted correctly. In the 6-Pc model, the prediction error reduced to 0.1906 with over 80% of choices accurately predicted. Details of 6-Pc Model 3 can be found in Appendix 4.4.

4.4.6 Accounting for scale heterogeneity

The latent-class models presented in section 4.4.5 did not account for differences in scale, which could produce biased class parameter estimates or even suggest an incorrect number of latent-classes, C. Fitting a scale-adjusted latent-class model, allowing for groups of respondents to have different error variances, allowed for further analysis of preference heterogeneity without the confounding scale factor.

A second calibration procedure to identify the correct number of Sc and Pc was conducted. In Table 4.10, the different class models and associated information criteria are presented. When the latent-class analysis was run with scale-classes, the 3-Sc-6-Pc model had both the lowest BIC and lowest CAIC. Table 4.10: Information criteria for varying class numbers in latent-class and scale-adjusted latent-

Number of classes	LL	BIC(LL)	AIC(LL)	CAIC(LL)
1-Pc Model	-9925.13	19877.92	19858.26	19881.92
2-Pc Model	-7954.85	16186.33	15989.7	16226.33
3-Pc Model	-7485.45	15496.49	15122.89	15572.49
4-Pc Model	-7042.29	14859.15	14308.58	14971.15
5-Pc Model	-6804.69	14632.91	13905.38	14780.91
6-Pc Model	-6661.4	14595.29	13690.8	14779.29
7-Pc Model	-6541.04	14603.54	13522.08	14823.54
2-Sc 5-Pc Model	-6647.69	14332.74	13595.39	14482.74
2- Sc 6-Pc Model	-6488.3	14262.91	13348.59	14448.91
2- Sc 7-Pc Model	-6415.02	14365.33	13274.04	14587.33
3- Sc 5-Pc Model	-6613.01	14277.2	13530.01	14429.2
3-Sc 6-Pc Model	-6479.56	14259.28	13335.13	14447.28
3-Sc 6-Pc Model (with covariates)*	-6328.24	14039.62	13056.47	14239.62
3-Sc 7-Pc Model	-6407.54	14364.2	13263.08	14588.2

class models.

*Chosen model

Regardless of the number of classes, the scale-adjusted latent-class models were always preferred to standard preference-class models based on the information criteria. Adding scale covariates (observed factors which might influence choice consistency) improved the model fit further. The final model selected was a 3-Sc-6-Pc model with both preference and scale covariates. Table 4.11 describes parameters of the 3-Sc-6-Pc model, with only the preference and scale covariates which were statistically significant in at least one preference and scale class included. Preference parameters for Sc-2 and Sc-3 can be obtained by multiplying the preference parameters from Sc-1 by their scale factor.

Preference-class	Pc-1	Pc-2	Pc-3	Pc-4	Pc-5	Pc-6
ASC (on none)	-171.997***	-153.514***	-9.971***	-18.993***	1.189	-6.159***
	(65.932)	(53.775)	(2.253)	(3.869)	(1.366)	(1.437)
Detect	2.785***	0.846***	0.279**	0.152***	-0.104	0.956***
	(0.518)	(0.184)	(0.110)	(0.055)	(0.268)	(0.288)
Risk	-0.331***	-0.999***	-0.504***	-0.155***	-0.723***	-1.704***
	(0.083)	(0.234)	(0.140)	(0.041)	(0.234)	(0.421)
Cost	-1.003***	-0.303***	-5.354***	-2.268***	-2.417***	-0.638***
	(0.229)	(0.086)	(1.208)	0.478)	(0.770)	(0.173)
Pc proportions	32.35%	29.15%	7.48%	18.70%	7.73%	4.60%
Preference covariates						
Unconcerned dummy**	-0.337***	-0.202*	0.231	0.117	0.120	0.071
	(0.128)	(0.115)	(0.150)	(0.126)	(0.148)	(0.199)
Employed dummy*	0.190**	0.029	-0.183	0.127	0.095	-0.257*
	(0.080)	(0.079)	(0.120)	(0.096)	(0.123)	(0.155)
White dummy**	0.227	0.289	0.048	-0.283	-0.508***	0.228
	(0.189)	(0.192)	(0.241)	(0.174)	(0.188)	(0.353)
IAP dummy	-0.064	0.111	0.054	0.025	-0.082	-0.045
	(0.074)	(0.075)	(0.115)	(0.091)	(0.116)	(0.152)
Scale-class	Sc-1	Sc-2	Sc-3			
Sc proportions	23.84%	64.87%	11.29%			
Scale factor	1	0.139	0.014			
Scale covariates						
Failed internal validity dummy**		-0.088	6.630**			
		(4.133)	(3.101)			
Task difficulty***						
Very easy (1)		-0.815***	-3.358			
		(0.292)	(2.225)			
Easy (2)		-0.129	-3.072			
		(0.312)	(2.349)			
Neither easy/ hard (3)		0.139	2.197**			
		(0.340)	(0.954)			
Hard (4)		0.759	1.606			
		(0.490)	(1.175)			
Very hard (5)		0.046	2.627*			
		(0.958)	(1.496)			
IAP dummy	•	0.042	0.220			
		(0.106)	(0.305)			

*p<0.10; **p<0.05; ***p<0.01; standard errors in parentheses

Most women (80%) fell into either Pc-1, Pc-2 or Pc-4, which were the largest classes. Pc-3 accounted for 7.5% of the sample, Pc-5 for 7.7%, and Pc-6 for 4.6% Across all classes, coefficients for each attribute had the expected signs and were statistically significant (apart from detect in class five which was insignificant). Although a number of covariates were hypothesised to influence preference-class membership (education, religion, age,

children, experience of mammography, and experience of breast cancer with either friends and/or family), the only significant covariates were 'unconcerned about risk of breast cancer', 'employment status' and 'ethnicity'. Figure 4.4 shows the relative importance of the attributes and the ASC for the opt-out for each preference-class.





The largest preference-class was Pc-1 which accounted for almost a third of the sample (32.3%). Women in this class treated the detection attribute as the most important, and had a large negative coefficient on the ASC for 'No Screening', suggesting a large amount of disutility in the absence of a breast screening programme. The women in this group were significantly less likely to be unconcerned about their risk of breast cancer (they had no idea, were concerned or were very concerned). The women in this class were also statistically significantly more likely to be employed.

Pc-2 was similar to class one. However, in this class, risk of unnecessary follow-up was more important than the attribute probability of detecting a cancer. Again, the coefficient on 'No Screening' was negative, large and highly significant suggesting women acquired utility from participating in the programme beyond that explained by the levels. Women in this class were also significantly likely to be unconcerned about their risk of breast cancer. This was a large class, accounting for 29% of the sample.

Another large class (18.7%) was Pc-4. In this class, the ASC on 'No Screening' was significant, negative and large (in line with the other preference-classes). However, women in this class were highly sensitive to cost, with this attribute being the most important.

In Pc-5, the ASC on 'No Screening' was positive (implying women in this group would rather not attend a screening programme) although statistically insignificant. This implies, in contrast to the other preference-classes, there was no baseline utility derived from participating in the screening programme. The most important attribute in this class was cost, followed by risk of unnecessary follow-up. The coefficient on the detection attribute was negative and the smallest of all preference classes. The coefficient of the detection attribute was also not statistically significant suggesting it was not an important factor of screening choice for women in this preference class. To test if women in this preference class ignored the attribute (exhibited attribute non-attendance (ANA)), the model was restricted; constraining the coefficient on the detection parameter to zero. This definition and assessment of ANA was based on published quantitative examples (see Gibson et al., 2015; Lagarde, 2012; Scarpa et al., 2009) which rely on an ex-post measurement rather than inferring attendance directly from survey questions. Restricting the model to allow for non-attendance to the detection attribute in Pc-5 improved model fit (BIC reduced to 14023.53), suggesting women in this class completely ignored this attribute. These women were significantly more likely to be from ethnic minorities (non-White). There is existing literature which could explain why this group of women who choose not to attend screening, and this is drawn upon in the discussion presented in section 4.7.

The two smallest classes were Pc-3 and Pc-6. Pc-3 accounted for only 7.5% of the women sampled. In the smallest class, Pc-6, the coefficient on the ASC for 'No Screening' was relatively small (compared with other classes) although still statistically significant. Women in this class were statistically significantly less likely to be employed.

Relative to this base class (Sc-1), the scale factor in Sc-2 was 0.14 (high error variance) and in Sc-3 was 0.01 (very high error variance). Respondents were split between the three scale-classes with 23.8% in Sc-1; 64.9% in Sc-2; and 11.3% in Sc-3. Three-quarters of the respondents had a relatively high error variance (low choice consistency).

The largest scale-class was Sc-2 which accounted for almost 65% of the sample. The women in this class were significantly less likely to have reported the task as being 'very easy', which is in line with the high error variance and low choice consistency associated with a scale factor of 0.14. In Sc-3, the women were significantly more likely to have failed the internal validity test and report the task as hard, again in line with the very high error variance (low choice consistency) in this class.

Risk format was not a statistically significant covariate in either preference-class or scaleclass membership suggesting the formating of risk did not affect either respondents' preferences or their choice consistency.

4.4.7 Marginal rates of substitution

To generate the MRS between the attributes, ratio calculations were conducted. The MRS explain the rate at which women were willing to give up the benefits of screening

(probability of detecting a cancer) in exchange for increased financial cost or increased risk of unnecessary follow-up, whilst maintain the same level of utility.

4.4.7.1 Willingness-to-Pay

The MWTP estimates from the heteroscedastic conditional logit for each class are shown in Table 4.12. The variation across the different preference-classes illustrated the heterogeneity in women's preferences. The MWTP for a one percentage increase in the probability of detecting a cancer ranged from £5.20 in Pc-3 to £279.48 in Pc-2. Pc-1 and Pc-2 have large MWTP for detection relative to the other classes, and considerably larger than the heteroscedastic conditional logit results which were £79.17. The MWTP for a one percentage decrease in the risk of unnecessary follow-up ranged from £6.83 in Pc-4 to £330.15 in Pc-2. Pc-2 and Pc-6 have very large MWTP for the attribute 'risk of unnecessary follow-up', suggesting these classes were very intolerant of this risk.

In Pc-5, MWTP was calculated using the insignificant parameter estimates hence MWTP to detect a cancer is negative (indicating that women in this class would need to be compensated for better detection).

Table 4.12: MWTP for breast screening attributes

	Heteroscedastic conditional logit	Pc-1	Pc-2	Pc-3	Pc-4	Pc-5*	Pc-6
MWTP detect	£79.17 (£66.98-£91.35)	£277.81	£279.48	£5.20	£6.71	-£4.28	£149.83
MWTP risk	-£46.01 (-£51.19£40.84)	-£33.04	-£330.15	-£9.42	-£6.83	-£29.90	-£267.01

*Insignificant coefficient on detection; confidence intervals in parentheses

Using current estimates for the levels of the screening attributes, a total WTP for a breast screening programme was estimated in Table 4.13. These values were calculated using the estimated parameters from the heteroscedastic conditional logit model in column one of Table 4.13 and the estimated parameters from each preference class in the scale-adjusted latent-class model. The levels for the screening programmes were taken from Table 4.3. The results of the heteroscedastic conditional logit show that WTP for best-case screening scenario (with no risk of unnecessary follow-up and a 14% detection rate) was over three times higher than for the worst-case scenario (with 3% probability of detecting a cancer but a 20% risk of unnecessary follow-up).

The WTP for the screening programmes in Table 4.13 again highlights the preference heterogeneity within the sample. The large values in Pc-1 and Pc-2 are generated by the

high values of the ASC on the 'No Screening' option, essentially reflecting that women in this group have a very strong preference to participate in a NHSBSP. In contrast, individuals in Pc-5 would pay nothing for screening in the best-case scenario, and they would have to be compensated to attend screening in the intermediate or worst-case scenarios. Similarly, whilst individuals in Pc-6 were willing to pay for the screening programme in the best-case scenario, they would have to be compensated for participation in the intermediate and worst-case scenarios.

Table 4.13: WTP for a lifetime	e breast screening programme
--------------------------------	------------------------------

	Heteroscedastic conditional logit	Pc-1	Pc-2	Pc-3	Pc-4	Pc-5*	Рс-6
Screening worst-case	£772.86 (£734.69-£812.32)	£17,329.25	£44,967.88	£13.42	£721.13	-£597.97	-£3,925.49
Screening intermediate	£1,787.13 (£1714.75-£1860.17)	£19,604.36	£50,225.05	£144.05	£836.46	-£298.99	-£206.36
Screening best-case	£2,563.90 (£2493.87-£2633.97)	£21,046.05	£54,643.89	£259.07	£931.64	£0.00	£3,063.23

* Insignificant coefficient on detection; confidence intervals in parentheses

4.4.7.2 Risk tolerability

The WTA an increased risk of unnecessary follow-up for a one percent increase in the probability of detecting a cancer is shown in Table 4.14. The results showed that women were willing to accept nearly two additional unnecessary follow-ups (1.72) for an additional cancer detected, per 100 women screened. Again, results from the scale-adjusted latent class analysis indicate substantial preference heterogeneity with WTA ranging from 0.55 unnecessary follow-ups for an additional cancer detected in Pc-3, to 8.41 unnecessary follow-ups in Pc-1. In Pc-5, women were completely intolerant of the risk as they derived no utility from the detection attribute.

Table 4.14: WTA risk to detect an additional cancer

	Heteroscedastic conditional logit	Pc-1	Pc-2	Pc-3	Pc-4	Pc-5*	Pc-6
Willingness-to-accept risk	1.72 (1.47-1.97)	8.41	0.85	0.55	0.98	-0.14	0.56

* Insignificant coefficient on detection; confidence intervals in parentheses

Standard errors, or confidence intervals, were not included for the preference classes in Tables 4.12, 4.13 or 4.14 due to the challenges of exporting variance-covariance matrices from a scale-adjusted latent-class models with scale-covariates in LatentGold®. The variance-covariance matrix was required for simulation procedures, which produce

confidence intervals and standard errors, such as the Krinsky-Robb method (Krinsky & Robb, 1986).

4.4.7.3 Predicting uptake

Using responses from the heteroscedastic conditional logit, the probability of a woman participating in screening was estimated using equation 4.22. Assuming a probability of detecting a cancer rate of 10%; risk of unnecessary follow-up of 10%; and a lifetime screening cost of £100, the probability of a woman participating in screening is given by:

$$P_i = \frac{e^{V_i}}{e^{V_i + e^{V_j}}} = \frac{e^{(\beta_1 10 + \beta_2 10 + \beta_3 1)}}{e^{(\beta_1 10 + \beta_2 10 + \beta_3 1)} + e^{\alpha}} = \frac{1.267}{1.491} = 0.850$$
 [4.22]

In this case, the probability of a woman choosing screening over 'No Screening' was 85%. Table 4.15 shows how this varied for the different screening scenarios described in Table 4.3. In the best-case scenario, uptake was predicted to be over 90%. In the worst, two-thirds of women would still attend. Only when risk of unnecessary follow-up was 47% (given detection at 10%) did the screening uptake drop to 50%. In a catastrophic scenario of no cancers being detected and all women in the programme receiving some unnecessary follow-up, 4% of women would still participate in screening.

Table 4.15: Probability of participating in a screening programme over 'No Screening'

	Screening worst-case	Screening best-case	Screening intermediate
Probability of	66.8%	92.6%	85.0%
participating in screening	(66.47-66.80)	(92.57-92.72)	(84.87-85.13)
	.1		

Confidence intervals in parentheses

Figure 4.5 shows graphically the effect increasing the risk of unnecessary follow-up on the probability of a woman participating in a screening programme for the different preference classes and the heteroscedastic conditional logit (hetclog) holding detection rates constant at 3% and cost constant at £20 per screen. For Pc-1 and Pc-2, women will always attend regardless of the risk of unnecessary follow-up. In Pc-4, women were very tolerant but uptake diminished as the risk of unnecessary follow-up exceeded 75%. In Pc-3 and Pc-6, women were tolerant of low levels of risk. In Pc-5, there was only a positive probability of a woman participating in screening at very low levels (<2%) of risk.

Figure 4.5: The effect of increasing risk on uptake for breast screening by preference-class



(Assuming 3% detection, £20/screen)

Figure 4.6 shows graphically the effect of increasing the probability of detecting a cancer on the probability of a woman participating in a screening programme holding risk of unnecessary follow-up constant at 5% and cost constant at £20 per screen. Women in Pc-1, Pc-2 and Pc-4 would always attend for screening, even at very low (and zero) levels of detection. In Pc-3, there were still high levels of uptake when probability of detecting a cancer is low and all would attend at detection rates higher than around 10%. Some women in Pc-5 would not attend for screening, regardless of the detection rate. In Pc-6, was low levels of uptake when the probability of detecting a cancer was low, but, similar to Pc-3, would always attend screening at detection rates higher than approximately 10%.





(Assuming 5% risk, £20/screen)

In Figure 4.7, the uptake rates given varying levels of screening programme cost is presented for the different preference classes and the heteroscedastic conditional logit

holding detection constant at 3% and risk constant at 5%. At zero cost, there was some uptake in all classes; however, this was very low in Pc-5 (where women were highly sensitive to the price of the screening programme). In Pc-3, Pc-4 and Pc-6, the probability in participating in screening dropped substantially as the lifetime cost of screening approached £1,000. Uptake rates in Pc-1 and Pc-2 were largely insensitive to cost even at values greater than £4,000.





4.5 Discussion

This study investigated whether the communication formats of risk attributes affected respondents' choices (or utilities) in a DCE. The results provided confirmatory evidence that the DCE is theoretically valid, with many women trading the attributes presented and results showing significant coefficients with signs in line with *a priori* expectations. In addition, sample characteristics indicated successful randomisation of DCE respondents to each risk format.

The key findings will be discussed in section 4.7.1, with reference to the original research questions stated in section 4.2. The implications of these findings for this thesis and research more generally is presented in section 4.7.2, and in section 4.7.3, the results of this study are compared to the existing literature. The strengths and limitations of the study are discussed in section 4.7.4.

4.5.1 Key findings

The overarching aim of this research was to understand if, and how, the communication of risk attributes in a DCE influenced respondents' choices, the consistency of these choices,

⁽Assuming 3% detection, 5% risk)

and the valuations derived. In this study, results from multiple models found no statistically significant difference in scale or preferences between the two risk formats.

The results of the heteroscedastic conditional logit model (Model 2A) found that the interaction terms between IAP and the risk and detection attributes were insignificant and that IAP had an insignificant effect on the scale term. This suggests that the risk format did not lead to any direct differences in either marginal utility or choice consistency, implying that the communication of risk had no effect in this study sample. These results were confirmed in the scale-adjusted-latent-class analysis (Model 4) which found that the risk format was not a significant predictor of either preference-class or scale-class membership. These results answer research questions two and three: randomness of choice and estimated utility parameters are not affected by the communication of risk.

This result is, in some respects, surprising because the structured review presented in Appendix 3.1 suggested risks were better understood when communicated with icon arrays. However, the review also identified that risk communication comprises multiple components, including risk perception and experience. In the latent-class models, concern about breast cancer risk was a statistically significant predictor of preference-class membership, indicating that risk perception, rather than risk communication, plays a significant role in women's preferences and their valuations.

Research question four was investigated through latent-class analysis which found women's preferences for breast screening were highly heterogeneous, with six distinct preference-classes. The largest class, class one, comprised women who tended to be more concerned about their risk of breast cancer and were significantly more likely to be employed. There was a high utility associated with participating in a programme with detection being the most important attribute in this class.

In Pc-5 five, women appeared to be uninterested in the screening programme with the estimate of the ASC and detection attribute both positive and insignificant. These women tended to be from ethnic minorities which is consistent with a large literature acknowledging the low up-take rates from these populations (Hoare, 1996; Renshaw et al., 2010), and that White women are most likely to attend screening, with little difference between age groups (Renshaw et al., 2010; Bansal et al., 2012). To target increased uptake for women from ethnic minorities, the benefits of screening could to be communicated better. One possible option is to offer GP endorsement letters and multilingual leaflets, for which there is existing evidence that this improves uptake for ethnic minorities (Bell & Branston, 1999). Alternatively, women in this group could be paid (or receive a subsidy for

their incurred out-of-pocket costs) for screening. A study found paying for transport improved attendance by 16% in ethnic minority groups (Bell & Branston, 1999).

The statistically significant predictors of class membership were ethnicity, employment status and concern about risk of breast cancer; this answered research question five. Other characteristics of the women (such as age or education) were not significant predictors of class membership.

Research question six sought to understand the MRS between attributes. The MWTP of women in preference-classes two, three, four, five and six was higher for an additional risk of unnecessary follow-up than for an additional cancer detected. This implies that women place more weight on the risks associated with breast screening than may have been previously thought. However, in preference-class one, the largest class, MWTP to detect a cancer was greater than that to avoid risk of unnecessary follow-up. Similarly, in preference-class one, women were about eight times more tolerant of the risk of unnecessary follow-up than women in the other preference-classes. The MRS highlights the heterogeneity between the distinct preference-classes.

The results from the heteroscedastic conditional logit model gave a WTP of £1783 (about £350 per screen). Private screening costs around £250 (BreastHealthUK, 2014) and allowing for some selection bias (women more interested in breast cancer and therefore more likely to attend screening completing the DCE) this is comparable to the market rate. Again allowing for some selection into DCE completion, the percentage uptake of around 85% also approximates the true uptake rates in England, with recent estimates indicating these to be around 75% (Breast Cancer UK, 2015).

4.5.2 Key implications

Three key implications for researchers conducting DCEs to elicit preferences for benefitrisk trade-offs were found: 1) pictorial framing of the risk attribute in addition to the percentage made no difference to either marginal utility or choice consistency; 2) risk preferences, in the case of screening for breast cancer, were highly heterogeneous; and 3) accounting for scale heterogeneity improved model fit.

What affects the valuation of risk in a benefit-risk trade-off DCE may not necessarily be due to the communication format of the attribute. An important factor affecting respondents' choices and valuation could be their concern, or perception, about their own risk. Research conducting DCEs for benefit-risk trade-offs should consider the sample's risk perception through additional survey questions and acknowledge the impact of this in their derived valuations.

There was significant preference heterogeneity in the model and the modelling data to allow for different preference-classes improved fit over a one-class heteroscedastic conditional logit model (BIC 19877.922 v 14595.29). Using latent-class models can also help understand preferences, when the drivers of heterogeneity are latent (or unknown). This is in line with other research which has found modelling preference heterogeneity with latent models (as opposed to interacted attributes and socio-demographic variables) is preferred, particularly in a healthcare context (Hole, 2008).

The study also highlighted the importance of accounting for scale heterogeneity in respondents' choices. Self-reported task difficulty was a significant predictor of scale-class membership. There is already existing evidence to suggest scale heterogeneity is a major issue, particularly in healthcare DCEs (Fiebig et al., 2010; Flynn et al., 2010). Allowing taste and scale heterogeneity to be confounded may result in biased estimates (Adamowicz et al., 2008), and if scale is not accounted for then apparent differences in estimates across groups may not be due to true differences in preferences.

4.5.3 Comparisons with existing literature

This section compares different relevant literatures with the results and key findings of the empirical study. In section 4.5.3.1, the literature surrounding women's preferences for the benefits and risks associate with a NHSBSP are explored and contrasted to the results of this study. In section 4.5.3.2, the findings relating to the risk communication format are contrasted with evidence from the review of Appendix 3.1. In section 4.5.3.3, other DCEs eliciting preferences for benefit-risk trade-offs are drawn upon to explore the robustness of the quantitative methods used in this study.

4.5.3.1 Literature on screening preferences

There have been no comparable DCEs looking at women's preferences for the benefit-risk trade-offs associated with screening for breast cancer (as identified in Chapter Three, section 3.4.6). However, a qualitative study investigating Australian women's views on over-diagnosis found that "*the lower and intermediate estimates [of over-diagnosis]* (1–10% and 30%) had limited impact on attitudes and intentions, with many women remaining committed to screening" (p.1; Hersch et al., 2013). Whilst this is in line with the results of the heteroscedastic conditional logit model (Model 2), the latent-class analysis did indicate substantial heterogeneity around these views. The findings from Hersch et al., (2013) do not reflect those of other large preference groups; at a 30% risk of unnecessary follow-up uptake would be very low (almost zero) in preference-classes three, five and six (accounting for around a fifth of women sampled). The quantitative exploration of

heterogeneity in women's preferences for a NHSBSP presents a contribution to the literature on screening preferences.

4.5.3.2 Risk communication literature

In the rapid review of risk communication studies presented in Appendix 3.1. Icon arrays were chosen as a risk communication format with the most empirical support, although one study communicating breast cancer risk (Brewer et al. (2012)) found that icon arrays were the least preferred method with 39% (n=46) of women saying it was risk format they liked the least (risk ladder was most preferred). Therefore, although the format had not been tested in a choice or trade-off setting, it was anticipated that IAP would be superior to PO. The hypothesis that IAP would improve choice consistency was not accepted. In addition, no differences were found between the utility parameters.

4.5.3.3 Studies considering framing effects in DCEs

There are some example DCE studies which have investigated the effect of different attribute frames. For example, Howard & Salkeld (2009) investigated the framing of risk attributes in a DCE for colorectal cancer screening. The DCE contained four framing formats for a risk attribute (true-positive/false-negative/true-negative/false-positive) in a forced choice experiment. The authors used a MXL and a MNL, comparing models on McFadden's pseudo R² and AIC. The authors interacted framing method with each attribute to identify differences due to the frame and compared the coefficients of the risk attribute by looking for overlap in the standard deviations from the four versions (no overlap indicated framing significantly influenced valuation). However, Howard & Salkeld (2009) made no account for scale heterogeneity between the survey versions, so their conclusion that the framing of risk did influence WTP (contrary to the results of this thesis) may be driven by differences in error variance rather than marginal utility.

In a conference paper, Buchanan et al., (2014) trialled a forced choice DCE with two risk communication formats (frequencies and icon arrays) with a mixture of postal, internet, clinic-based surveys completed by patients in NHS trusts or members of a charity. The authors then ran four models (MNL; a split sample MNL; MXL; and a latent-class model) on the pooled data, and found that respondents receiving the icon array had higher WTP values (by approximately 50%). However, these authors again failed to account for scale heterogeneity between risk formats (and survey modes), directly comparing coefficients despite the confounding of scale and preferences.

The absence of consideration of the scale parameter in studies comparing risk formats limits the ability to contrast the results of this study. Other studies comparing the effect of cost attribute framing have also failed to account for differences in scale heterogeneity (for example, Grutters et al. (2008)). It appears to be a consistent limitation in healthcare DCE analysis, with applied studies not accounting for sources of bias (most notably the scale parameter) identified in other research areas such as marketing and environmental economics (Louviere, 2006; Fiebig et al., 2010).

4.5.4 Strengths and limitations

As section 4.5.3.3 showed, this was an original empirical investigation, going beyond current standard of analysis in typical health DCEs. The application was also original in that there are no published examples comparing percentages and icon arrays in the framing of risk attributes. The DCE also possesses characteristics of both theoretical and external validity, with results in line with *a priori* expectations and comparable to revealed preference data.

However, the study had some limitations. A heteroscedastic mixed logit model could have been used as a further investigation into preference heterogeneity. However, it was believed that describing heterogeneity in terms of latent classes (rather than say a distribution from a mixed logit model) would be more useful for decision-makers. In addition, latent class models are arguably superior to mixed logit models in their statistical properties (Shen, 2009).

Non-linearity in preferences was investigated in the preliminary stages of model selection (see Model 1D). Investigations of non-linearity could not be conducted in the more complex models, such as the selected 3-Sc 6-Pc Model as this failed to converge. Non-convergence is likely to have occurred due to the increased number of parameters (from three to 14) and the small sample size in some preference classes, such as preference class six which had fewer than 50 respondents. However, there is a possibility that some, or all, preference classes may have non-linear preferences in some or all attributes, and this should be considered when interpreting the results presented in Table 4.11.

The conclusion that IAP offered no advantage to respondents in terms of reduced cognitive burden and had no effect on their preferences, may be specific to the context of this study (breast screening) or the magnitude of the levels used (percentages were whole numbers). In other scenarios, the results may differ, and further research is required to investigate the generalisability of the key findings described in section 4.7.1. Similarly, alternative methods (such as decimals, frequencies or fractions) may differ to percentages and icon arrays.

Furthermore, the quantitative analysis was based on observed data and unobserved behaviour (for example, what respondents were truly thinking) was not collected. It could be that this was affected by the framing of risk even if choice behaviour was ultimately the same.

4.6 Conclusion

This study has helped to understand more about the effect of different risk communication formats. It was identified that neither choice nor choice consistency were affected by the framing of the risk attributes, but that there was significant preference heterogeneity, with latent-classes of women deriving different values from different aspects of a breast screening programme. However, the analysis relied on a purely quantitative interpretation of women's trading behaviour.

In order to investigate risk communication effects further, alternative research methods may be required to understand more about women's interpretation of the attributes in a DCE context. The following chapter uses a qualitative research method called think-aloud to further investigate framing effects.

Chapter Five

Benefit-risk trade-offs for breast screening: a qualitative study

5.1 Introduction

This chapter describes a study designed to explore more about how respondents make choices in DCEs containing risk attributes. The chapter uses qualitative research methods as part of a mixed-methods approach to complement the purely quantitative approach of the previous study presented in Chapter Four. Qualitative research methods were used to understand more about how the format of risk might affect respondents' cognitive heuristics in a DCE setting and explore unobservable factors which might influence choice-making behaviour.

Chapter Five begins in section 5.2 by describing a systematic review of qualitative research alongside published DCEs and a survey to the authors of these DCEs eliciting their views and experiences of the usefulness of such methods. The research questions to be investigated are described in section 5.3. A description of the types of qualitative research methods available are presented in section 5.4, before a justification and explanation of the methods and analysis used in this study are presented in section 5.5. In section 5.6, the results are presented using the key concepts: risk and decision-making strategies. These concepts were identified through a literature review determining these *a priori* as important areas in addressing the overall aim of the PhD, which is to understand the drivers and consequences of risk communication format in a DCE most effectively. Chapter Five concludes with a discussion and description of the limitations of this study in section 5.7, before briefly introducing how the subsequent chapter (Chapter Six), an eye-tracking study, seeks to address some of these concerns in section 5.8.

5.1.1 Background to qualitative research methods

Qualitative research methods are used to understand *what*, *how* and *why* through explorations of people's views or lived experiences (Silverman, 2013). Chapter Four identified substantial heterogeneity in women's preferences for a NHSBSP. The quantitative analysis was limited in its ability to answer questions about *what* people think when they see the risk attribute, *how* they compare risks and benefits, and *why* they make such choices. In order to better understand how people balance risks and benefits, and to acquire a deeper-understanding of DCE respondents' valuation strategies, qualitative research methods were identified as an appropriate strategy.

5.2 Qualitative research methods and DCEs

While DCEs involve statistical analysis of quantitative choice data, other research areas, particularly in the social sciences, have used qualitative research methods to investigate people's beliefs and decision-making processes (Berg, 2007). Although not frequently employed by economists, there is a growing body of research supporting the use of qualitative methods in health economics (Coast, 1999; Coast et al., 2004). In the context of DCEs, qualitative research methods have a number of potential roles. This could include identifying attributes and/or their associated levels, refining the terminology, cognitive piloting of the survey, exploring the use of aids and vignettes, gaining an understanding of respondents' decision-making processes, or estimating preference heterogeneity before modelling the data (Coast & Horrocks, 2007; de Bekker-Grob et al., 2012; Kløjgaard et al., 2012).

Guidelines for conducting DCEs advocate qualitative methods with specific recommendations for the development of attributes and levels (Bridges et al., 2011b; Coast & Horrocks, 2007; Lancsar & Louviere, 2008). However, there has been no explicit investigation of how well these recommendations regarding the use of qualitative methods in the preparation and testing of DCEs have been translated into practice.

5.2.1 Systematic review of qualitative research methods alongside DCEs5.2.1.1 Systematic review aims

The aim of this component of the thesis was to systematically identify all published studies in healthcare that reported the use of qualitative methods to inform the design and/or the interpretation of DCEs. The objectives were to: summarise the proportion of DCEs using qualitative methods; assess the context in which the research was applied; identify the methods and techniques used; and, where possible, appraise the quality of the research conducted.

5.2.1.2 Systematic review methods

This study used systematic review methods as advised by the Centre for Reviews and Dissemination (CRD) to identify all healthcare DCEs published in the last decade (since, and including, a previous systematic review) (de Bekker-Grob et al., 2012). The systematic review focussed on identifying DCEs rather than other stated preference methods such as CJA, ACA or contingent valuation because, as discussed in Chapter Two, these methods are grounded in different economic theories and are therefore not directly relevant to this review or the overall thesis (Louviere et al., 2010). An electronic search of Medline (Ovid, 1966 to date) was conducted in June 2012. Although other databases could have been searched, the strategy exactly replicated that of published reviews of DCEs (Ryan & Gerard, 2003b; de Bekker-Grob et al., 2012). The search terms used were: 'discrete choice experiment(s)', 'discrete choice model(l)ing', 'stated preference', 'part-worth utilities', 'functional measurement', 'paired comparisons', 'pairwise choices', 'conjoint analysis', 'conjoint measurement', 'conjoint studies', and 'conjoint choice experiment(s)'. The term 'conjoint analysis' was included to identify studies which had used discrete choices rather than those which required respondents to rate or rank alternatives. No search terms were used to directly identify qualitative studies as this was deemed to be too restrictive.

The main foci of the review were to: 1) identify and quantify the proportion of DCEs using qualitative methods; 2) investigate the stages in the DCE at which qualitative research is employed; 3) understand the methods and techniques currently used; and 4) where possible, evaluate the quality of the reporting of research. The studies were initially categorised into three categories: 1) those which reported **no** qualitative research; 2) those which contained **basic** reporting which indicated some qualitative research may have been used; and 3) those which indicated an **extensive** qualitative component was conducted in direct relation to the DCE. This categorisation identified studies, in category three, which contained sufficient detail for critical appraisal.

Using pre-existing guides for appraisal of qualitative research methods in a traditional sense may have meant the extensive studies identified by this review would have been judged incorrectly or unfairly. It is, however, crucial that the qualitative research contained in the studies was formally assessed in a standardised and systematic way. Therefore a bespoke appraisal tool was developed to include broader issues which were not included in the traditional tools as advised by the CRD (CRD, 2008). The bespoke appraisal tool can be found in Appendix 5.5.

5.2.1.3 Systematic review results

One hundred and twenty four studies were already identified by a previous systematic review (de Bekker-Grob et al., 2012). The search resulted in 501 titles and abstracts since the previous review (2008 onwards) and 208 full papers were retrieved for further assessment and 148 papers met the inclusion criteria. Figure 3.1 shows the stages involved in screening and the reasons for rejection of the excluded papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) review guide (Moher et al., 2009). Therefore, in total, there were 277 papers included in the final review, which related to 254 empirical studies (because some studies were reported in more than one paper). A detailed description of the reasons for rejection can be found in Appendix 5.3.Overall, 111 (44%) studies did not report the use of any qualitative research methods; 114 (45%) reported minimal data on the use of qualitative methods; and 29 (11%) reported, or explicitly cited, the extensive use of qualitative methods.

Almost all (n=113, 99%) of the studies which reported basic qualitative research reported using it before the DCE was implemented, either in the design or piloting phase. A variety of applications of qualitative research methods were identified. In the design of the DCE, researchers were most commonly seeking to identify attributes and/or assign levels (n=70, 61%), or validate attributes and/or levels identified through other methods (n=31, 27%). Researchers also used qualitative research methods more specifically to check terminology, vignettes and descriptions (n=9, 8%) and to confirm translations (n=2, 2%). After the design phase, some studies also reported using qualitative research methods in the piloting of the DCE (n=24, 21%). In the pilot stage, the methods were specifically used to check for decision strategies (n=1, 1%) and also to determine an appropriate sample for the final DCE.

Of the studies reporting extensive details about the qualitative components, most studies (n=25, 86%) again reported the use of qualitative research methods to identify attributes and/or levels for use in the DCE. The most common data collection approach was interviews (n=18, 62%). These interviews were mostly semi-structured (n=12, 41%) and face-to-face, although two studies used telephone interviews. A number of studies also used focus groups or group interviews (n=12, 41%). Three studies (Cheraghi-Sohi et al., 2007; Ryan et al., 2009; Bridges et al., 2011b) used qualitative research methods to understand more about how respondents completed the choice task presented, and two of these (Cheraghi-Sohi et al., 2007; Ryan et al., 2009) used a type of verbal protocol analysis.

In terms of the analysis, most studies simply stated that they used thematic analysis (n=10, 34%) or content analysis (n=5, 17%) to categorise the qualitative data collected. with reducing the qualitative data to develop a few attributes and levels. Other analytical approaches included framework analysis (n=3, 10%) and a related approach to qualitative analysis called charting. Seven studies used some constant comparative analysis (n=4, 14%) or open-coding (n=3, 10%) at least in the initial stages.

5.2.2 A survey to authors

The systematic review presented in section 3.2 of this chapter was constrained by its reliance on the details reported in published academic papers. It is well established that a

bias exists in both medicine and social science, with studies exhibiting positive or interesting results being more likely to be published (Franco et al., 2014; Easterbrook et al., 1991) A survey was, therefore, designed to elicit information from authors of published DCEs in health about their views on the use of qualitative research methods alongside quantitative analytical methods.

5.2.2.1 Survey aims

The aim of this study was to try and reduce the effects of reporting bias through directly eliciting researchers' views and experience of conducting qualitative research alongside DCEs in a healthcare setting. The objectives of this study were to reveal more about researchers' experience of using qualitative research methods; their opinion of the usefulness of qualitative approaches in this context; and any explanation they had for the poor level of reporting found in the systematic review.

5.2.2.2 Survey methods

This study used an online semi-structured survey comprising closed and open questions. The survey was sent to authors who indicated, in a published study, that they had used qualitative research methods but only reported basic details. As described in section 5.2.1.3, total of 114 studies reported basic use of qualitative research methods and all authors were included in this review. As some corresponding authors had multiple studies included in the review, 91 individual authors were sampled. All of the journal articles provided an email address for the corresponding author. Therefore, the most feasible method of contacting authors and eliciting their views was to use an online survey.

The final survey (presented in Appendix 5.2) consisted of 12 questions and comprised a mix of closed-ended and free-text comment boxes. The questions asked authors about their **experience** and their **opinion** of: 1) using qualitative research methods alongside DCEs; and 2) communicating the qualitative work they conducted in a journal article. Additional questions included self-assessment of their and co-authors' expertise in qualitative research, the number of DCEs they had conducted, and whether they agreed with the key finding of the systematic review that qualitative research is not well reported in healthcare DCEs.

The survey data were downloaded from the online server and analysed in Excel® (Microsoft, 2010). The analysis involved simple production of descriptive statistics for each of the questions. The authors' free-text comments were not thematically analysed because of the limited textual data available (some authors chose not to comment).

5.2.2.3 Survey results

After the first email sent on 1st May 2013, 38 authors completed the survey (an initial response rate of 42%) within a month and before a second reminder email was sent. The questionnaire closed on the 30th June 2013, with a total of 53 completed or partially complete responses, resulting in an overall response rate of 58%.

The authors reported that the use of qualitative research methods added value to their experience of conducting DCEs in general, with 74% (n=31) stating it made a 'substantial improvement' to the study. The majority of survey respondents (n=42, 79%) agreed with the systematic review finding that the qualitative component was only briefly described in DCE papers. Half of the respondents (n=26, 50%) believed the amount of qualitative work conducted was accurately reflected in the published paper as reported. Over half of the respondents (n=26, 52%) considered qualitative research was too complicated to report in detail and one fifth of the respondents (n=10, 20%) reported that they felt it was too time consuming to conduct properly. Some authors (n=11, 16%) also stated that qualitative research would not be of interest to their peers, and n=4 (8%) also reported that they did not believe it was important to funders. Appendix 13 tabulates the authors' responses to each survey question.

5.2.3 Discussion of review and survey findings

The results of the systematic review and survey to authors identified qualitative research methods were being used by DCE researchers to answer multiple research questions, and that these methods add value to a DCE study. This finding suggested that qualitative research would be an appropriate method to use in this thesis.

The systematic review identified two papers (Cheraghi-Sohi et al., 2007; Ryan et al., 2009) which used concurrent verbal protocol analysis (called 'think-aloud') to understand more about people's choices. As this PhD thesis seeks to understand more about how people answer a DCE containing a risk attribute, the think-aloud method was considered appropriate to answer the key research questions.

There was a paucity of detail about the analysis of the qualitative data collected, however, when described, most studies used some sort of thematic analysis, which was most often a pre-defined framework to code the collected data from which themes were developed using NVivo® software. Therefore the analysis of the think-aloud data collected in this thesis followed a similar approach.

5.3 Research questions

The key research questions addressed in Chapter Five were formulated from the findings of the systematic review (section 5.2 and Appendix 5.1) and the results of the quantitative study (see Chapter Four). The key research questions addressed in this chapter were:

- 1) What do people say they think about when they complete a DCE?
- 2) How do people complete DCEs with risk attributes?
- 3) How do people interpret risk attributes in a DCE?
- 4) Does the communication format of risk attributes affect DCE respondents' qualitative accounts of choice making?

5.4 Introduction to qualitative research methods

Social science, and the research conducted by social scientists in particular, has been broadly split into two schools: positivisim (which tests correlation between variables) and interpretism (which is concerned with observation and description) (Silverman, 1993). Another way of distinguishing between these two schools of thought is whether hypothesis *validation* or hypothesis *generation* is sought (Johnson & Onwuegbuzie, 2004). For example, in generating a hypothesis, qualitative research methods could be used to begin an enquiry into a particular topic and quantitative research methods for confirmation of the topic's existence. The exact distinction between the two schools of thought depends on the epistemological (founding beliefs) and ontological (philosophies) view taken by the researcher. Table 5.1 briefly outlines some of the differences between the two schools of thought.

School	Implied Hypothesis	Definition	Example of research
Positivism	Hypothesis validation	Testing of correlation	Confirming a relationship
		between variables by	between variables;
		objective means and	investigating
		statistics	generalizability
Interpretism	Hypothesis generation	Observing and	Beginning a particular
		describing by subjective	enquiry into a research
		means and points of view	topic; investigating
		in an exploratory manner	transferability

Table 5.1: Schools of thought in	n research methodology
----------------------------------	------------------------

Although contemporary economics can be seen as a mathematical-based science, and hence uses quantitative research methods with a positivist perspective, some have argued that all variables have some element of social construction (Silverman, 1993). Qualitative research methods largely fit under the interpretivist paradigm, designed to describe and illuminate a particular topic, and thus offers the theoretical foundation to this study (Denzin, 2009).

Qualitative research methods collect and analyse textual data (in contrast to quantitative studies which traditionally seek numerical observations to test hypotheses) (Pope & Mays, 2008). The aim of qualitative research methods is to acquire a deeper understanding about a human action in reference to a particular topic or phenomenon (Huston & Rowan, 1998). To do this, qualitative research methods commonly employ interaction with people, and mostly use interviews or focus groups to initiate this interaction (Denzin & Lincoln, 2005).

Chapter Four empirically demonstrated that women's preferences were highly heterogeneous but that the communication format of the risk attributes did not affect these preferences or the consistency of their choices. However, the large quantitative study and analysis was limited in its ability to explain more about how risk is interpreted by DCE respondents. The econometric modelling allowed identification of groups of women answering in similar ways, as there was evidence strongly suggesting the presence of distinct (latent) classes. Although it was possible to predict class membership (with some significant preference class co-variates), this interview study builds upon the findings of the quantitative analysis by exploring quantitatively, the unobservable drivers of choice and behaviour in more depth by directly asking respondents about their thoughts.

Furthermore, in-depth analysis is warranted as it is impossible to be certain that those surveyed all uniformly interpreted both the choice tasks and the terminology used. In other words, differences in preference classes may reflect differences in women's interpretation rather than any actual differences in their attitudes or valuations. That is why qualitative research methods will be useful in revealing if, and how, respondents bring different meaning to the information presented to them.

The systematic review of the use of qualitative research methods in DCEs (see section 5.2 and Appendix 5.1) supported the use of qualitative methods to design and interpret DCEs. These uses are described in guidelines and in best-practice literature with the aim of improving DCE design and understanding more about DCE choices. The systematic review concluded that the design and analysis of qualitative research studies alongside DCEs were not well-reported. The supplementary survey to authors experienced in measuring peoples' preferences confirmed that qualitative research methods were being used and, more importantly, in the experience of those surveyed, did have a positive influence on study progression and quality. The few DCE studies that reported the use of qualitative research in more detail, complemented by the results of the survey to DCE

authors, illustrated the potential power of qualitative research methods to reveal influences on choice or decision strategies that may be difficult, or impossible, to measure in a traditional quantitative DCE survey alone. These results from the review of section 5.2 provide a pragmatic justification for the use of qualitative research methods whose prime purpose are to understand *what*, *how* and *why*, from an individual perspective.

In this study, the qualitative approach involved the collection of empirical data to explore, in-depth, the views of DCE respondents. From these data, it was possible to understand more about how risk was interpreted in a DCE setting and if this differed by the risk communication method. The qualitative study was built around the DCE task and the choice sets produced by the design described in Chapter Three. The qualitative data provided a snapshot of the effect of completing the DCE under an interview setting.

What makes good quality qualitative research, particularly for stated preference studies, was discussed in the development of the bespoke critical appraisal tool employed in the systematic review of section 5.2. The bespoke critical appraisal tool served a second purpose in this chapter, Chapter Five, by providing a checklist of steps and structure for the design of the proposed qualitative study. The tool (see Appendix 5.5) can be summarised using the following headings: 1) background and overview; 2) description of the context; 3) sampling methods; 4) data collection; 5) reflexivity (or a consideration of how the investigators themselves can influence qualitative research); 6) ethical aspects; 7) data analysis; and 8) presentation of results. These headings will serve as a structure for sections 5.5 and 5.6 of Chapter Five which detail the methods and results, respectively.

5.5 Methods

The previous section, section 5.4, introduced qualitative research as a research method. The techniques used in qualitative research are now explored and critiqued. Following this, an explanation and description of the methods used in the empirical study are presented.

5.5.1 Qualitative research methods

The aim of qualitative research methods is to generate a textual account of an event from the perspective of the people being studied (Silverman, 1993). The most commonly used qualitative method is interviews. However, qualitative research methods include other approaches ranging from raw observations in the form of anthropological-enthographies (which involve no interaction with the subject of interest) through to the analysis of free-text comments or social media updates (Branthwaite & Patterson, 2011; Reeves et al., 2008).

Interviews are broadly defined by Burgess (1984) as: "*conversation with a purpose*". They are not only the most popular method, interviews are also defined as the 'gold standard' of qualitative research methods (Silverman, 2013). Interviews are not one single method, rather a group of conversational approaches, which vary in the mode (phone, face-to-face, group settings) to the uniformity of the questions (structured, semi-structured, open-ended) (Denzin & Lincoln, 2005). Figure 5.1 summarises the possible research methods with the broad arrows containing possible conformations which could be employed with each approach.





Within the groups of conversational (interview) approaches, there are many different techniques which can be employed to conduct interviews. Possible approaches include group interviews, telephone calls, face-to-face settings or emails. The results of the systematic review presented in section 5.2 and Appendix 5.1 identified a verbal protocol method called concurrent think-aloud. This method has been used by a number of researchers (Cheraghi-Sohi et al., 2007; Ryan et al., 2009) specifically to understand more about people's choice task performance (rather than other uses such as attribute identification or level development). For example, Ryan et al., (2009) looked specifically at how respondents processed the DCE task and what led to 'irrational' responses in dominant choice sets.

Whilst telephone interviews or email exchanges could have improved the response rate (and saved money and time) it would not have been practical in this study which used a

survey. Face-to-face interviews are generally seen as the gold standard (Novick, 2008) and can be particularly useful in a survey context where seeing the interviewee pointing, their body language, and facial expressions may yield additional insights into their thoughts (Mason, 2002). Although these were not analysed in this study, their observation to aid the interview process and provide a stimulus for the appropriate timeliness of prompts was deemed crucial. Therefore, given the limitations of other techniques and the success of the verbal protocol analysis in previous DCEs, concurrent think-aloud was identified as the most suitable qualitative research method for this study.

Development of the think-aloud method is attributed to Ericson & Simon (1984) who introduced the method to generate scientific data in psychological experiments investigation information retention in people's short-term and long-term memory. Since then, the method has been widely used by both cognitive psychologists and interviewers seeking to understand people's approaches to problem-solving (Boren & Ramey, 2000). Unlike other interviewing techniques, and contrary to the definition above, the think-aloud method does not involve much 'conversation'. Instead, the dialogue is replaced by the interviewee speaking out-loud their thought processes with the interviewer prompting in periods of silence.

Semi-structured interviews lie in the midst of the qualitative methods spectrum, as they allow the interviewer to explore and probe areas of information that arise if the interviewee volunteers a topic; however, the presence of a schedule can assist in keeping the interviews structured enough to answer, rather than specifically develop, a research question. This study aimed to acquire understanding about people's experience of completing DCEs and, therefore, a series of pre-determined topics that needed to be covered were identified. These pre-determined topics meant that a semi-structured interview in the form of debriefing questions was deemed appropriate in addition to the think-aloud protocol.

5.5.1.1 Sampling

Sampling is a crucial stage of any study design but is often constrained by resources (time and money), access to potential participants, and ethical considerations (such as the sensitive nature of the interview topic). Although commonplace in positivist research, probability-based sampling which aims for an unbiased random sample, is not a requirement of qualitative research (Marshall, 1996; Luborsky & Rubinstein, 1995). This study did not seek to produce a statistically representative sample, as statistical significance of any kind was, and is not, a goal of qualitative research. Instead a trustworthy sample was desired based on the transferability criteria recommend by Lincoln & Guba (1985). Sufficient detail about the final sample, fieldwork, context and environment were provided to enable readers of this study to evaluate the similarity of another setting and decide for themselves whether the findings were transferable. Therefore a non-probability purposive sampling strategy which aimed to collect a sample of diverse women was devised. In contrast to many qualitative research studies, which rely on principles of inductive reasoning to construct an idea (Burns, 1989), the primary aim was not to achieve hypothesis generation, but to understand how risk was interpreted and traded-off in the DCEs. Although a more sophisticated sampling strategy could have been used, a specific sample was neither desired nor sought.

The sample selection criteria were limited to females, fluent in English and between 18 and 70 years of age for reasons discussed in Chapter Three. Seventy years was chosen as a maximum as this is the current cut-off for routine screening in England (Independent UK Panel on Breast Cancer Screening, 2012), meaning women over 70 years old would be neither current users nor potential users. No other sample restrictions were included as initial discussions with breast cancer experts in the design of the DCE revealed that they were interested in the views of all women (see Chapter Three). In light of this, the internet panel (see Chapter Four) surveyed women in the 18 to 70 year age category and therefore, for consistency, the sampling frame for this qualitative study maintained the same inclusion criteria.

In line with purposive theoretical sampling, advertisements were placed in an attempt to acquire a range of interviewees with a range of key characteristics covering various ages, jobs and ethnicities. The advertising strategy involved placing advertisements online at The University of Manchester's publically available volunteers' website (www.student.manchester.ac.uk/volunteers) and through emails to staff distribution lists. Paper advertisements were placed in shops, cafes, bars and public noticeboards. Examples of these advertisements can be found in Appendices 3.9, 5.14 and 5.15. The advertisement responders were only incentivised at first contact with the researcher, when they were informed about the £10 Amazon voucher, as per the recommendation of the ethics committee. At this first point of researcher contact, the potential participants were also provided with an information sheet containing frequently asked questions (Appendix 3.11). The advertisement responders were then asked to read the sheet, take time to think about the study and then, if they were interested, make contact again to arrange the interview.

5.5.1.2 Interview schedule

A preliminary interview schedule was used in pilot interviews (n=5) described in Chapter Three. After these interviews were completed, it was decided that warm-up questions were required to encourage interviewees to vocalise their thoughts. The final interview schedule (Appendix 5.16) was created through an iterative process of discussions with the supervisory team and comprised a warm-up exercise and debriefing questions. The main part of the interview involved prompting the participant to keep verbalising their thoughts through encouraging questions (*Why did you choose that? What are you thinking*?) and did not follow a pre-defined structure. This also limited response acquiescence or a 'yeasaying bias' by provoking interviewees to come-up with their own accounts instead of agreeing with the interviewer (Blamey et al., 1999).

The interview schedule helped to keep the interviews on topic, and although useful to examine participants' perspectives and perceptions of risk (including their family history and how this had influenced their responses) the structure allowed the interview to be brought back to topic whilst permitting exploration of emerging issues which were not anticipated and consequently not included in the guide. As discovered when developing the critical appraisal tool, care needs to be taken when applying qualitative research in a stated preference setting to ensure the questions do not lead to a pre-assigned idea or hypothesis and, instead, remain exploratory in nature (the so-called 'tourism effect' by Silverman (1993)). Similarly, there were no 'right answers' or *a priori* expectations to this research, and the flexible schedule allowed complete collection of (even disconfirming) data.

5.5.1.3 Data collection

Women who agreed to the interview were invited to either come to the university or meet in a location of their choice, as long as it was quiet and private enough to facilitate audiorecording. The participants were also reassured that they needed no prior subject knowledge. All participant contact and interviews were conducted by CV. Before the interviews started, interviewees were shown the information sheet they first received when they enquired about the study (Appendix 3.11) and the purpose of the study was reiterated. After the researcher had checked whether the interviewee had any questions, the interviewee completed a consent form permitting the use of recording and note-taking (Appendix 3.12). The digital voice recorder was then set to record and the interview began.

The interview started by explaining to the participant what was meant by think-aloud and that they did not need to talk to the interviewer *per se*, but just say whatever they were thinking. As identified to be useful in piloting, a warm-up exercise was conducted which

asked interviewees: "*How many windows do you have in your house*?". If women gave an absolute number straight away, the think-aloud concept was explained again and they were invited to reveal how they got to their answer, which usually resulted in a walk-through of their home. Interviewees were then reassured that this was all they would be required to do and were then invited to read over the training material for the DCE to familiarise themselves with the topic and task to be completed.

Interviewees then completed the DCE, thinking aloud whilst selecting their answers on an iPad. Women were randomised to a risk communication format and experimental design, consistent with the method used in Chapter Four. At the end of the DCE, a series of debriefing questions was posed to elicit anything else that might explain the interviewees' choices. The debriefing questions generally followed the schedule (see Appendix 5.16); however, interviewees who made interesting comments in the think-aloud task were probed further to explore a full range of ideas (for example, reference to particular familial experiences). At the end of the interview, interviewees were specifically asked to consider for a few minutes and see if anything else came to mind.

All the questions used were open-ended to encourage the interviewees to talk more descriptively and develop their own narrative. Prompts were used if the interviewee went silent. The debriefing questions were also used to generate more detailed data when the interviewee was less vocal in the think-aloud task. Throughout the data collection process, the topic guide outlined in the interview schedule remained unchanged.

All interviews were digitally recorded and notes were also taken when it was apparent the transcripts would not reveal sufficient information (for example, when interviewees pointed at attributes or alternatives).

5.5.2 Data generation

The interviews were transcribed verbatim and supplemented with field notes where appropriate (for example, when an interviewee had pointed to something then notes were used to identify the object – usually a particular level). For some interviews, the transcription was done immediately after the interview (n=5). Others were transcribed by a professional company with experience of the transcription of health-related interviews (n=14). The professionally transcribed interviews were then double-checked by CV for accuracy by listening to the recording and reading the interview reports, and, where appropriate, supplemented with field notes from the day.

Data were transcribed verbatim in order to maintain close contact with the raw data. When looking at cognitive processes and comprehension, any discourse, even pauses or

mumbles, were sources of useful qualitative data, adding realism to the quotes provided in support of the identified themes.

NVivo® qualitative software for research (QSR, 2012) was used to import the transcripts and also store the audio files in preparation for the analysis.

5.5.3 Process of analysis

The first stage of the analysis involved listening to and re-reading all the interviews and transcripts to ensure complete immersion in the qualitative data. This process cannot be explicitly reported here as it was a psychologically-internal procedure. However, familiarisation with the data simply by reading and re-reading transcripts and listening to audio files kick-started the initial analytical process. During each reading, notes were made and, as interviews were being conducted concurrently to the initial stages of the analysis, points of particular interest were also jotted down in subsequent interviews.

The data generated from the think-aloud exercise were pooled with the data from the debriefing questions and all were analysed using the same methods. This was largely because there was no pattern to the reporting of information in terms of which method (think-aloud or de-brief question) generated the data (for example, some women revealed in the think-aloud that they found the percentages confusing whereas others only mentioned this towards the end of the interview).

From these transcripts and memo notes, the coding and generation of themes began under a loose initial framework (Gale et al., 2013). A framework can be used to identify important areas a priori in order to facilitate the interpretation of qualitative data by providing a preliminary structure to the analysis (Smith & Firth, 2011). The initial codes were highlighted on the transcript text and involved a constrained form of open-coding under the framework. Although constrained open-coding may seem like a contradiction, the approach was 'open' in that all new themes were allowed to develop but 'constrained' in that they were restricted to the two distinct, but not mutually exclusive, key concepts: 1) decision strategies; and 2) risk. This approach was taken to focus the analysis and remove qualitative data that was not of interest (for example, comments on the warm-up exercise) but still allowed related themes to develop from the codes. Decision strategies related to any accounts of behaviour or heuristics adapted when completing the choice task. The topic of risk referred to any data relating to the probability attributes, whether this be their thoughts about likelihood, perceptions of risk based on experience, or visualisation of the numbers. The process was driven by the interview data and each transcript was systematically coded, with these codes later developed into themes.

The process of coding was both inductive and deductive. It involved looking at the transcripts inductively under the two broad themes of decision strategies and risk, but was deductive in that it was open to new themes which did not 'fit' and were then initially coded under new headings and incorporated into a new framework.

Within the two concepts of decision strategies and risk, coding involved an iterative process of developing new themes and ideas as the process advanced. It was then possible to stand-back and look at the snippets of interviews which seemed broadly related and thus re-define the themes in line with the original framework. This was unrestrictive: the same text could be coded multiple times and generate different codes which could possibly overlap. This stage was key in acquiring a familiarity with all interview data collected during the think-aloud and de-briefing questions. The analytical cycles of suggesting, checking and re-checking to expand and then contract themes continued until no new items or ideas were emerging.

The next step involved developing a coding tree stemming from the original framework (see section 5.5.2). The coding tree helped enable discussions with the supervisory team and facilitate feedback on the interview process, even when the interviews were ongoing. The initial broad range of themes were collapsed and re-organised by merging similar codes, moving or attaching others and creating sub-codes when necessary. In the initial stages, the tree was reflected a web of loosely related codes, with many branches of similar topics and tenuous links. After more meetings, interviews and further analysis, the codes were arranged into a series of more structured themes.

5.5.4 Qualitative boundaries and analytical limitations

A common criticism of qualitative research is the impact of the presence of an interviewer both in the collection of data and in its interpretation and analysis (Shenton, 2004). As described previously in section 5.5.1.2, steps were taken to reduce this source of bias in the recruitment and interview stages. It was suggested to women responding to the advertisement that the interviews could be held in their own home, and participants were explicitly told that they should not worry about talking to the interviewer and to only say what they were thinking.

Pilot work for the DCE (described in section 3.5.8.3 of Chapter Three) generated free-text comments from a small internet panel which suggested some DCE respondents were concerned about the confidentiality of the data (whether this would be used for a private company). Therefore efforts were made to reassure participants that anything they said and the choices they made would be completely anonymised in this PhD and any subsequent

publications. Again, interviewees were given an opportunity to read the information sheet and decline consent before the interview even began.

In the interviews, attempts were taken to remain as impartial as possible and minimise researcher bias. Although prompts were made, if interviewees looked for direction they were reassured that there were no right or wrong answers. With concurrent thought-verbalisation other criticisms of bias from interviews, stemming around recall and politeness, were minimised as interviewees were encouraged to speak their mind at all times. After 12 interviews, two meetings were held with the supervisory team to discuss the transcripts and acquire feedback on the preliminary analysis. Therefore the results presented in the next section are as honest, impartial and objective as is reasonably possible given the nature of qualitative research.

5.5.5 Summary of the analysis and methods

The overall process from sampling to conclusion is described visually in Figure 5.2. The process was fluid and involved assessments, reassessments, supervisory feedback and interviewee input before any conclusions could be substantiated. The content of the ovals represent key stages in the empirical study; the arrows show the processes leading to each stage; and the squares describe the methods used in each process.
Figure 5.2: Qualitative data collection processes



5.6 Results

A total of 76 women replied to the advertisements and 23 interviews were arranged. Two women cancelled and two failed to show-up. All women who replied to the advertisements and were sent the information sheets initially agreed to participate. However, some (n=2) had family commitments or became too busy to commit to the interview, and some (n=6) were ineligible to participate in this study because they made contact after all interviews were completed (these women were informed of subsequent research and many chose to participate in the study presented in Chapter Six). Appendix 5.17 describes the recruitment process for this chapter, Chapter Five, and Chapter Six, and the difficulties in attributing a response rate for either study.

5.6.1 Interviewees

The final sample comprised nineteen women. The sample size was guided by this previous research, but, as described in section 5.5.1.1, sampling would have continued if new themes emerged from the data; however, a point of saturation appeared to have been met. Although new participants could have been recruited (and women were actively responding to advertisements), the data generated by the accounts of the last few interviewees replicated that of previous participants so further work would have been redundant. This point of data saturation is not only common, it is also considered to be best practice in qualitative research (Silverman, 2013).

The 19 women were randomised to complete one of the two risk communication formats, with nine women receiving the IAP version and 10 women receiving the PO version. The randomisation occurred via an inbuilt function of the Sawtooth software (in an identical procedure to Chapter Four), therefore the process should not have affected the emerging themes or the qualitative data collected. A summary of the characteristics of each interviewee is described in Table 5.2.

Identity (ID)	Risk format	Age band	Occupation	Date and
	PO = %;			Location of
	IAP = 4			Interview
Female 1	%	50+	Analyst	Home
				18/09/2014
Female 2	%	35 to 44	Lecturer	University
				15/09/2014
Female 3	%	18 to 24	Social Worker	University
				15/09/2014
Female 4	ŧ	25 to 34	Teacher	Home
				19/09/2014
Female 5	ŧ	18 to 24	Unemployed	Home
				13/09/2014
Female 6	ŧ	25 to 34	Researcher	University
				17/09/2014
Female 7	ŧ	25 to 34	PhD Student	University
				18/09/2014
Female 8	ŧ	25 to 34	Researcher	University
				24/09/2014
Female 9	ŧ	25 to 34	Unemployed	University
				17/09/2014
Female 10	%	25 to 34	Administrator	University
				19/09/2014
Female 11	%	18 to 24	Undergraduate	University
			Student	19/09/2014
Female 12	%	50+	Information	University
			technology	06/10/2014
			(IT)/Administrator	
Female 13	%	50+	Administrator	University
				29/09/2014
Female 14	%	50+	Administrator	University
				30/09/2014
Female 15	%	25 to 34	Bar worker/ part-	University
			time student	30/09/2014
Female 16	%	45 to 49	Finance officer	University
				06/10/2014
Female 17	ŧ	25 to 34	Administrator	University
				06/10/2014
Female 18	ŧ	25 to 34	Engineer	University
				13/10/2014
Female 19	Ŧ	50+	Retired	Home
				15/10/2014

On average, the interviews lasted approximately 40 minutes (mean time of recording 37 minutes 50 seconds) with the shortest lasting 24 minutes and 54 seconds and the longest taking 57 minutes and 26 seconds. The interviews took place in September and October of 2014 and most (n=15) occurred in pre-booked meeting rooms on The University of Manchester's Main Campus. Interviewees could choose the location, but most decided to come to the university. The sample included a range of ages and professions as described in Table 5.2.

5.6.2 Coding tree

The initial coding tree (shown in Figure 5.3) included all themes and sub-themes identified in the initial stages of the analysis. This tree describes the "whole picture" and the full range of views expressed in the interviews before any substantive attempt at reduction. Themes were only loosely grouped together with subtle connections explaining potential links. The original framework is shown by black square boxes, and recurring themes are displayed according to their focus: circles representing feelings and emotions; capsules showing logic and reasoning; the diamonds demonstrating expressions relating to the interviewees' perceptions or experiences; and the white squares encapsulating opinions and beliefs. The arrows are suggested links and possible drivers of each phenomenon.



Proposed links

Feelings/emotions

Perceptions/experiences

Opinions/ beliefs

Logic/ reasoning



Figure 5.3: Intermediate coding tree demonstrating identified codes and themes

The initial coding tree was then simplified by reducing the themes into broadly related topics. This involved a back and forth approach of rearranging and establishing connections or patterns in behaviour to create a more coherent account of what occurred in the collected data. The themes are shown in the circles of Figure 5.4 and patterns in behaviour that linked themes are shown by arrows. Figure 5.4 also serves as the structure for the next section of this chapter which discusses the each of the final themes in more detail with supporting, and disconfirming, verbatim quotes as evidence.



Figure 5.4: Final coding tree identifying key themes and inter-theme relationships

Key issues in the empirical study

5.6.3 Experience of completing a DCE

In this next section, the themes summarised in Figure 5.4 are discussed in more detail. The section begins with an exploration with interviewees' general experience of completing a DCE survey. Section 5.6.4 examines how interviewees used decision strategies to make their choices. In section 5.6.5, interviewees' reaction to the risk attributes are explored in more detail with comparisons of the differences in the interviewees' accounts from the two risk communication formats: PO and IAP.

5.6.3.1 Experience of making choices

This section explores interviewees' general experience of completing a DCE task. Alongside each interviewee quote is an indicator of the risk communication format the woman received and the age bracket they fell into. The '%' identifies women who completed the DCE with risk communicated as a percentage only; and the ' **/** ' identifies women who also received visual communication of risk in the form of an icon array.

The interviewees were asked to start to think-aloud as soon as they started the DCE. Many of the initial comments were expressions of surprise or confusion. Interviewees, in both risk communication formats, reported that they had never done anything like the task before and, despite confirming everything was clear prior to commencing, a few asked for clarification of exactly what was expected:

CV: *Did you find that difficult?*

Participant: Yes, I did at first, you know, before I, sort of, realised what was going on there.....by the time I got to the end of it it was like you were almost looking at the same thing on each page. (*Female 19: 50+ years, ∮*)

Similarly, other interviewees sought confirmation or reassurance that they were doing the right thing with regards to both the choice experiment and the think-aloud task:

Participant: Sorry, is that wrong? *CV*: No, it's not wrong at all; there are no right or wrong answers. (*Female 12: 50+*, %)

Participant: Am I making sense? I'm worried! ...Hmm I think I will go for A. Am I making sense? (**Female 2: 35-44 years, %**)

Previous think-aloud studies have focussed on respondents' reactions to the dominance task (Ryan et al., 2009); however in this study, no women were confused by the dominant choice set incorporated as a test of internal validity, and all participants gave the expected response:

Participant: So that's all a win, win thing there. (Female 13: 50+ years, %)

Participant: Oh well that's really easy... that's pretty obvious. (*Female 18: 25-34* years, *↓*)

When interviewees were asked whether they would make the same choices in real life, the majority confirmed that they probably would:

CV: Ok, so do you think if you were to make these decisions in real life, your choices would be the same?
Participant: I think, I think that I just chose the ones that personally I felt I would go for. (Female 2: 35-44 years, %)

Participant: Yep, I'm not 100% but I think I'd be quite confident. I'm trying to imagine, if I got a letter, what would I do? I think this is what I would do. So I'm quite confident. (*Female 4: 25-34 years, ≇*)

However, a few interviewees found it hard to get around the hypothetical nature of the DCE:

Participant: You see, that's why it's difficult, because you're not actually presented with it now, as in real life kind of thing, it's difficult. (**Female 16: 45-49 years, %**)

Participant: Oh, ok, I just think until something actually happens.. I mean even with role play.. you don't actually know until it happens.. you'll have lots of other things to think about... because it's real and it's happening...you might think.. oh..oh well.... oh that's why I hate role play. (**Female 3: 18-24 years, %**)

The above quotes highlight that a DCE is not an easy thing for people to complete and that it takes respondents a while to familiarise themselves with the survey despite reading the pre-information and having an interviewer present to answer any questions. This DCE was 'simple' in that there were only three attributes, but women still expressed a need to take time to comprehend exactly what was asked, understand the good described and imagine the reality of their hypothetical choices.

5.6.3.2 Experience of completing a think-aloud task

Some women appeared to find it difficult to think-aloud. It was not possible to distinguish whether this was down to confusion with the task, because they could not verbalise their responses, or because they genuinely felt as if they had nothing to say. As detailed in the methods section of this chapter (section 5.5.1.2), prompts were used extensively in some cases but, even then, some women struggled to verbalise:

CV: What about in this case?
Participant: [Pause]. Same again. I'd go for programme B.
CV: Okay. Is that the same...?
Participant: For the same reasons. (Female 12: 50+ years, %)

Women who were hesitant or quiet in the think-aloud task, often revealed more about their choices and strategy in the debriefing interview. One interviewee waited until the very end of the interview to explain more about what was influencing her choices:

Participant: Now we've finished, I should tell you that I recently started HRT [hormone replacement therapy] and I know that increases your risk of breast cancer so that probably made me always choose screening. It's a risk, and I know that, but at the end of the day I'm prepared to take it because it's made me feel so much better. (**Female 1:** 50+ years, %)

5.6.4 Evidence of economic phenomena

In their responses, many women expressed behaviours that have resonance with established economic phenomena. The economic behaviours were not reported in the qualitative accounts of the studies identified in the systematic review of DCEs presented in section 5.2 and Appendix 5.1, and so were unexpected and entirely driven by the qualitative data collected. The women did not talk in economic terms, but their self-reported choice strategies could be aligned with recognised economic theories. Whilst these behaviours have not been extensively explored within a health economics context, or in a DCE framework, they have been identified by environmental economists and in studies using the contingent valuation method (Hanley et al., 2003).

5.6.4.1 Opportunity costs

Opportunity cost is a key concept underpinning health economics' overall aim of ensuring the efficient allocation of resources. This concept is also key to the supporting theories of DCEs which require respondents to trade-off in order to estimate their underlying preferences. Although not explicitly stated in academic-economic terms, many women expressed that they were making choices consistent with the idea of opportunity cost in terms of financial analogies:

Participant: You don't miss, you know, £5 here, £5 every time you go for breast screening, or the ticket cost on the bus or car parking. You don't miss that, because you're having a service done... you're asking as you pay, like you would to go shopping or to go to town.... Bus fare, car, you know, car parking, it's nothing. You wouldn't think twice of doing that to go shopping, so why not do it to have your life saved. (**Female 12: 50+ years, %**)

Participant: I think it's very important and spend more money on healthcare than other things in life probably, but that's just probably me personally. (*Female 11: 18-24 years, %*)

Another emergent theme was that women seemed to be thinking in terms of 'efficiency' of, or the opportunity costs to, the NHS more generally. In a publically funded health care system, many women seemed to be concerned about a programme being wasteful:

Participant: Well it seems to me like programme B sounds like it's much more efficient in what it's doing, if there's no women detected who had unnecessary follow up. (Female 14: 50+ years, %)

Participant: See this is a waste of a lot of resources. I think I would go for none of these... This is too expensive and the other is too wasteful. 20% of people coming back is too high, it's not so good... I'm evaluating the programmes so first I see which would I go for then my second point of view is if it's useful or not useful. If programme B is calling half the people back with follow up then it means there's a lot of wastage on the system. (Female 2: 35-44 years, %)

Two women also expressed ideas about 'regret' contributing to their decision-making:

Participant: It doesn't really matter about the cost because obviously if you have breast cancer and you haven't gone to Programme A or B then you'd be kicking yourself, you know, if you were to end up terminally ill... (**Female 11: 18-24 years, %**)

Participant: I think there's quite a few things, but for me I think a big concern would be being treated unnecessarily because you can't really reverse it once you've started the treatment. (*Female 6: 25-34 years, ‡*)

5.6.4.2 Discounting

It is well established in neoclassical economics that people have a preference for money today rather than in the future and thus discount returns at a later-date (Frederick et al., 2002). It is less clear how people treat non-financial benefits such as gains to health (Torgerson & Raftery, 1999). In this study, there was evidence that some younger women were discounting the benefits of breast screening:

Participant: But also I think - I know I'm only sort of, like 19 years away from starting such a programme, but it seems like a long way away. Dying doesn't seem to me - that's not something that's going to happen to me anytime soon. ... (Female 18: 25-34 years, 1)

One contributing factor to a discount rate appeared to be the uncertainty surrounding both their future income and future health (ultimately uncertainty in their future budget constraints and the future utility they could acquire from participating in the screening programme):

Participant: Obviously you don't know if you're going to have £1,000 just to spend over 20 years. (*Female 11: 18-24 years, %*)

Participant: You know, where I've been in and out of hospital, had surgeries, and different things, lots of different things. So, you never know how you're going to be at any age really, do you? (Female 19: 50+ years, #)

When consulting breast cancer experts in the original design of the DCE, they expressed an interest in understanding the drivers behind women's decision to attend screening. This was deemed to be useful in terms of predicting adherence and explaining why some women coming to the end of the programme chose not to attend (Edwards & Jones, 2000).

Some women's statements indicated that their choice was probably only reflecting their preferences for early screens, and as they reached the end of the programme they may no longer choose to participate. This is not discounting, but rather women updating their priors about the expected utility from screening (the probability of a positive test result) given their previous results:

Participant: I think it would also depend on if I've had one or two before and they've turned out okay, I think I would be willing to pay less for further ones but I would be willing to pay probably the most for the first few. So that's how it would be... I think my views might change if I've had three and it turns out they're all alright and I'm almost...yes, I'd be almost 60 (**Female 6: 25-34 years, #**)

Participant: I would always go, but dependent on what it would say would influence my decision to go back. (*Female 4: 25-34 years, ↓*)

5.6.4.3 Decision process and model selection

The systematic review of qualitative research methods (reported in section 5.2 and Appendix 5.1) found that some authors had used qualitative data to inform the quantitative analysis of choice data (see Pitchforth et al., 2008). Although this is not touched upon in any of the DCE guidelines (Bridges et al., 2011b; Lancsar & Louviere, 2008), some respondents revealed behaviour that could inform model selection through their treatment of the 'No Screening' alternative. Some interviewees reported that they first considered whether to opt-out of screening before even considering the two screening alternatives:

Participant: So no screening, yes, okay. I think Programme B is probably preferable to that one, yes. (*Female 10: 25-34 years, %*)

Participant: For this one I'm not really sure because I think that like Programme A the smallest one has ten per cent who are treated unnecessarily compared to the twenty but then now it's a case of whether it's better to not have a screening at all than to be one of those ten... (Female 6: 25-34 years, **f**)

5.6.4.4 Information attendance

In terms of attendance to the information presented in each choice set, it was clear that some interviewees were not taking everything into account. One woman expressed complete surprise when prompted about her thoughts of the 'No Screening' alternative having previously ignored it:

CV: Ok, do you consider this [opt-out] when you make your choices?
Participant: I didn't even see this! So what.. so what is this? No cancers detected..
Wait...so this... what? I don't get this. So this...?.. [paused] so you have no cancers detected...because you wouldn't be screened?
CV: Yes...

Participant: That none-option doesn't even come into it. (Female 3: 18-24 years, %)

For another woman, the opt-out offered a way-out of making a choice at a point of indifference:

Participant: I think I'm probably in the not sure and I'd probably in this one go no screening for this one. (**Female 6: 25-34 years**, **#**)

Many women expressed that they would always opt-in to a breast screening programme. This was not surprising given the large and positive ASCs identified in five out of six of the preference-classes described in Chapter Four. The quotes below further support this quantitative finding:

Participant: That wouldn't be an option for me at all...Yes, sorry, I hadn't even bothered with that one because you know, I would definitely go, if it was me I would definitely go for screening, no matter what, even if they didn't...it was unnecessary and I had chemotherapy and whatever, I would still have gone for it anyway. (*Female 14:50*+ *years, %*)

Participant: So there's no screenings at all? No I'd never go for that....So this third option is always going to be with no screening?
CV: Yeah.
Participant: Well I'm never going to go with that then, I can't do that. (Female 13: 50+ years, %)

Non-attendance was also present in the form of ignoring attributes. This specific phenomenon of attribute non-attendance is explored in more detail in Section 5.6.4.8.

5.6.4.5 Information asymmetry

It is well established that information asymmetry is common in healthcare markets (Bloom et al., 2008). Whilst DCEs are often cited as a way round this, trying to specifically understand patients' or the public's demand for a healthcare interventions, many women expressed that their choices were completely dependent on the advice they received from government agencies or 'experts':

Participant: So I would be guided by whether I thought there was a reason. If I was invited, I would definitely go... But again, I'm thinking if it was offered to me then there's a reason it would be offered to me. I would, I think I would want to go. If the NHS said oh, we think you should go, then I would be like ok... I would trust them. Only because I don't know anything about it myself, if I had a better judgement then I would make a choice for myself. But I don't know much so I would be guided by them. (Female 4: 25-34 years, *‡*)

Participant: Obviously the NHS has decided for a reason that 50-70 is a good age to go so I think I would go for it... One thing I find is, you don't often receive information on how accurate any screening is or how effective the medication is, so it's difficult for people to think - like when the doctor tells them something you don't really know is this going to work? Is this giving me the result I want? Because you're putting so much trust in that and most people don't have the time to go off and research it themselves. (**Female 18: 25-34 years, #**)

Participant: You don't think about it being necessary or not. And you just, I suppose, with breast, you just kind of do what your GP tells you, don't you, a lot of the time. And you might have limited information about what you're supposed to do, and it's just kind of, well, I should go for that, so I should go for it. (**Female 16: 45-49 years, %**)

Participant: Yeah, I think basically because it's just the normal thing to do as well, people do it and it's just like... (Female 15: 25-34 years, %)

Whilst many women expressed feelings of trust, some women were surprised at the statistics presented and felt that they had not acquired sufficient information about screening:

Participant: I thought the percentages of detection were quite low, yeah, it just sounds quite a low number isn't it to go through all that and just, it's quite...so I'm quite

surprised about that, I thought...Yeah, but yeah, I'm surprised the detection levels aren't higher and, yeah...mm. Yeah. I've learned stuff that I didn't know about. I didn't realise that they took a...up about the unnecessary follow-ups actually. (Female 15: 25-34 years, %)

5.6.4.6 Familiar good effects

In the ecological literature, there is empirical evidence that valuation of animal species can be influenced by people's prior attitudes towards the wildlife or natural resource and its familiarity (Czech et al., 1998). Potentially, higher valuations could be a consequence of social, political and media attention to the subject (Czech et al. 1998). Whilst this has been briefly discussed in some health economics articles (where the success of breast cancer charities in their campaigns and subsequent influence on policy has been noted (Pennery, 2007)) it has not been specifically investigated in a health DCE context. However, cancer, and breast cancer in particular, appears to have some socially-constructed virtue ascribed to it by some of the respondents in this study. Whilst these media effects and charitable campaigns are established in terms of revealed actions (Evans et al. (2014) highlight the 'Angelina Jolie effect' on breast cancer referrals) it was unclear whether this influenced hypothetical choices. At the time of the interviews, a celebrity was diagnosed with breast cancer and a few women drew upon the celebrities and/or the media in their debriefing questions:

Participant: Like Angelina Jolie, isn't it? If you compare with that. (*Female 13: 50*+ *years, %*)

Participant: Michaela Strachan... She's just had them both off, hasn't she, because of breast cancer. She's not that old, is she? (Female 12: 50+ years, %)

5.6.4.7 Healthcare perspectives and non-use values

As introduced in section 2.2 of Chapter Two, a key stage in understanding how people value health care is to understand the perspective they take and whether the valuations derived are welfarist or extra-welfarist (Brouwer et al., 2008). In this DCE, women expressed sentiments implying they were completing the DCE from a population-health (rather than individualist) perspective, thinking of the greater benefit to the general public:

Participant: the higher, the more people we can, you can find, the more people you could treat, the higher the chances of survival, so... I think that on a humanitarian level, that you've got to open it to everybody, and try and save as many people as possible. (*Female 13: 50+ years, %*)

Participant: if it's like a mum with children, if you've got kids or you've got a big family then you've got to look after your health and you've got to try and get medical advice and see if there's any problems. Or if they couldn't... some people have it really bad have got families, one parent families, having to do loads of stuff in debt and it's just like they literally can't afford to take a day off work. So that's...yeah, that would be a bit of a factor with people that actually come and get these tests I think. (**Female 15: 25-34 years, %**)

There was also evidence to suggest that the respondents were not just considering other people's health whilst they were completing the DCE, but they were specifically drawing their own utility from these altruistic provisions. In the environmental economics literature (Hanley et al., 2001; Perman et al., 2003), these philanthropic motives are identified as potential 'existence values' which contribute to the non-use values people may have from a good or service:

Participant: So it's difficult to say, because although you don't want lots of people having unnecessary treatment, at least if it's more successful in detecting it, I'd probably for A there then. (**Female 16: 45-49 years, %**)

Participant: It's tricky because until you're in that position I don't ever want to have to put 20 per cent of people through any unnecessary follow ups.. hopefully not submitting people to, as many people as possible obviously, hopefully not submitting the highest number to any unnecessary procedures but outweighing that and balancing that off against what would be mostly detected. It's not the perfect solution, that's the frustrating thing, you can never tell... I suppose part of my influence is also the wider community. (Female 10: 25-34 years, %)

There was further evidence of existence values in line with 'warm-glow' theories (Kahneman & Knetsch, 1992). Receiving a warm-glow suggests some women gained utility from the moral satisfaction of having an inexpensive service that anyone could use:

Participant: Would you get more people going in for Programme A despite the fact there's 20 per cent unnecessary follow up. There is a higher rate of detection but is that balanced off the fact that it's only £20 per screen. We were estimating if 100 people going to each of these programmes... if we were to look at the general population as a whole and offer the population these two programmes only a certain number of people would be able to afford Programme B...because Programme A's cheaper you've got more people going for screening on Programme A and therefore you're finding a higher rate of cancers detected in any case... At a cost of £20 it would hopefully be far more of an incentive to go along and spend that money to have the screen...Thinking about...certainly people that I work for whether they'd have the disposable income to do it as well is slim to none. Therefore the lower the cost the more likely it is that they would go. Certainly if you're talking about hundreds of pounds per screen I think that would be far outside the reach of most people. (**Female 10: 25-34 years, %**)

Participant: Even though I personally would still pay £200 to get the same thing, I think that for the population, and considering the demographic of the people who get the screening, that £200 might be too much for people of that age maybe. So to encourage more people to go for it I would go for programme B in that case. (*Female 8: 25-34 years, 1*)

Non-market valuation methods could be susceptible to bias from respondents gaming the survey to give a desired answer, or refusing to complete the survey because they disagree with the principal entirely (Smith, 2000). In environmental economics, one potential reason for these strategic responses is that respondents have a strong belief that something should be preserved (Perman et al., 2003). In health, the NHS and the NHSBSP are the items to be protected. Whilst there were no explicit statements expressed by any of the respondents in this study, some women did indicate protective views of the NHS:

Participant: I mean, I honestly believe in the NHS. They've done so many good things for me in the last 20 years and I would prefer to support the NHS every time and, obviously, there's a few cases where you've got to jump the queue because you're desperate ... and I think the NHS are suffering at the moment. (Female 19: 50+ years, #)

Strong beliefs about protection could lead to another bias: protest responses. A protest response reflects a choices given on principal and therefore may not reflect the

respondent's true value (Jorgensen et al., 1999). One statement indicated a temptation to give a protest response:

Participant: I'm also trying to think about, for example, my gran having to go for one and having to pay £200 each time. She probably would, but I just think that it's quite mean. She shouldn't have to do that, kind of thing. (*Female 8: 25-34 years, ‡*)

5.6.4.8 Non-compensatory preferences

A number of key axioms support this rational choice model and estimation of utility requires respondents to express reflexive, complete, transitive and continuous preferences (Varian, 1992). If a respondent exhibits non-compensatory preferences and no amount of money can compensate them for not having screening, they will choose to go for screening at any cost. When respondents are completely unwilling to trade-off this cost attribute they are exhibiting attribute-non-attendance. This violation of the continuity axiom through a reluctance to trade-off the cost attribute was common (particularly in women aged 50+):

Participant: I still think that having the screening totally outweighs what the cost would be anyway....well because I don't think you can really put a cost on your health and your life, your survival. (**Female 14:50+ years, %**)

Participant: So I just can't see how you can go wrong by spending money on something because money against life [laughter] it's no comparison... There's no comparison... it's not a matter of money. And I do stress that's not because I'm well off ... generally speaking, to be told with great relief that you haven't got cancer, you can't buy that. (Female 13: 50+, %)

Participant: Well, where it's saying cost to you I don't think the cost should come into it... it's your life you're talking about. (*Female 19: 50+ years, #*)

Participant: I just don't think you can put a price on your health. (*Female 8: 25-34 years, 1*)

Participant: It's a lot of money but then I think it's hard to put a price on cancer being detected, isn't it?... just think that with stuff like that, money's not a factor..you've got to pay for screening then I'd pay it. (**Female 17: 25-34 years, #**)

Even when women took into account the cost attribute, many would exhibit lexicographic preferences by simplify the task by rank-ordering the attributes. The most common rank-ordering involved making decisions with the preference: probability of detecting a cancer>>risk of unnecessary follow-up>>cost of screening:

CV: Did you find yourself sort of concentrating on one of the characteristics or ...? *Participant*: Yes, the percentage of detected more than anything else, really....And then the unnecessary follow-ups and then the money was my third kind of thing, but yes, it was governed by percentage of detected from the screening. (*Female 17: 25-34 years, f*)

CV: So, you're just were you going for the thing with the highest detection rate, best at detecting?

Participant: Absolutely, yes. (Female 19: 50+ years, #)

Whereas other respondents were able to interpret the cost attribute and made choices in line with the underlying economic theories:

Participant: It sounds silly because it's obviously your health and you can't put a figure or price on it but people will do because that's the day to day life that we lead. If you've got to go home, feed the kids, clothe the kids or whatever and if they haven't got the money to do that then they're not thinking much beyond that I don't think. (**Female 10**: **25-34 years, %**)

However, worryingly, a few women also associated the cost of the screening programme with its quality:

Participant: Oh I think there's an element of saying, if it's £20 a screen then I'm bound to think that's not serious: it's a no-good screening programme that's basically just, you know, making random decisions and I don't think it's working. (Female 2: 35-44 years, %)

Participant: Even the unnecessary treatment I'd rather be on the safe side than think I've gone and I've got something and it's not been detected because it's cheaper or because of another reason. (**Female 11: 18-24 years, %**)

Participant: It's like why wouldn't I maybe pay a bit more to be seen, sort of, in a more beneficial and professional way perhaps? (*Female 18: 25-34 years, ‡*)

One woman also drew upon her experience of paying for healthcare to add realism to the hypothetical choice and estimate how much she would really be willing to expend:

Participant: Funnily enough I was looking into like the HPV [human papilloma virus] vaccine but they said if your cervical smear test didn't come back abnormal they wouldn't test you for HPV, so you could go home and buy the vaccine which was £400... but they couldn't tell you whether or not it was going to work ... and I thought, mm, £400 for something that I've no idea whether it's going to work or not is just too much money for me. But if it was I'd say up to about £80 I probably would have done it anyway, so it's kind of like similar... (Female 7: 25-34 years, ≇)

5.6.4.9 Stability of preferences and anchoring

In some stated preference studies, the axiom of complete preferences is tested with a stability check (Czajkowski et al., 2014). In this study, some women reported that their preferences changed or became more defined as they completed the experiment (potential learning effects):

CV: Do you think your choices changed as you went along through the different choices?

Participant: Yes, I think so. I think I started to think more about the cost and whether I would prioritise being treated unnecessarily more than the chance of it being detected and treated. Most of the time I was on the treated unnecessarily, that was the main thing for me, but then when I started to think about the fact that if you have cancer then it might be missed that became more of a priority as well. So, yes, I think it changed a little bit. (*Female 6: 25-34 years, ∮*)

Participant: I think, maybe at the start, I was more focussed on the price, but then actually having gone through a few and weighing it up it's actually more important to look at the other factors rather than the money. (*Female 9: 25-34 years, #*)

Whilst other women felt that they were consistent and did not change their decision rules:

Participant: I think I was consistent. I think it got difficult when one was really expensive and one was quite a lot cheaper. Again, for the same reason that I can't expect everyone to do what I would do, kind of thing. (*Female 8: 25-34 years, #*)

Participant: I think I based it on the same reasoning but in weighing up the options it became easier to understand what the options could be. I think I stuck to the same logic all the way through though. (Female 10: 25-34 years, %)

The DCE training materials were extensively piloted (as described in section 3.5.8 of Chapter Three) and included a description of all the attributes and potential levels prior to starting the DCE. However, some respondents appeared to anchor their choice based on the levels which had appeared in previous choice sets:

Participant: Oh, is that [£1,000] the highest? **CV**: Yes

Participant: Oh ok the earlier one where I chose a different one even though it had a high cost. But now this high cost is what's stopped me here.
CV: Yep. Do you think you change your mind as you see new choices?
Participant: Yes! Not massively, but it's making me think more. Like £1,000 is massive! And it makes me want to choose that one ... Weird how it makes you think.
(Female 4: 25-34 years, #)

5.6.5 Experience of completing a DCE with risk attributes

A key aim of this PhD is to investigate how DCE respondents' understood risk attributes. This next section looks at the think-aloud interviewees' reaction to the risk attributes specifically, before comparing how these differed between the communication formats.

5.6.5.1 Computing Behaviour

For some respondents, the use of risk attributes appeared to induce behaviour indicative of calculating and a desire to 'work out the answer', and this was more apparent in women who received the PO version:

Participant: I was trying to do the maths...[paused]
CV: Ok, so when you try and 'do the math' what do you mean?
Respondent: Well... I think if you actually worked out the maths, I might be slightly out.
(Female 3: 18-24 years, %)

Participant: Oh because I need to compare, I'm making decisions to make them comparable. So this is 7% so I'm halving it to 3.5% which is nearer the 3% and then I'm halving the rest to make it equal and then I can see it's half and half. Then I'll look at the cost - and the 7 - 5 ratio - and then the cost. (**Female 2: 35-44 years, %**)

Participant: So actually yes, so if you've got - I'm just trying to work this out. You'd have approximately a 3 in 10 chance - no you wouldn't at all, you would have a 1 in 3 chance, sorry, of getting a false positive there and you'd have a 0% chance there - 7% or 10%. (*Female 18: 25-34 years, 1*)

Some of these calculations were more apparent with women receiving the IAP version, suggesting the format might aid trade-off decisions:

Participant: It's like £50 versus £150 and then the follow up looks like about half, there's only five per cent who'll be treated unnecessarily whereas it's double the amount of Programme B... Then looking at the number of people who have had it but then at the same time the cost is a quarter of what it would be. Programme A is a quarter of the cost of Programme B but because the amount's so different in terms of the financial cost for this one I would pick A. (**Female 6: 25-34 years,** *‡*)

Participant: I suppose because I work in engineering I do tend to consider things quite carefully and I'm quite logical in how I make choices. As I say, I'm probably more logical than most other people. I'm sure that things like this can probably trip me up too. (*Female 18: 25-34 years, ‡*)

Two women (who both received the PO version) reported that they were attracted to the largest number first (whether that was probability of detecting a cancer or risk of unnecessary follow-up) and would make a decision from there:

Participant: the higher the number, I was looking for there. (Female 14:50+ years, %)

Participant: Ermmm the high percentage.. and then I think hmm .. then the money... and then that doesn't come into it. (*Female 3: 18-24 years, %*)

One woman reported flipping the negative risk into a positive one to make the comparison between attributes easier:

Participant: Errr.. actually..well 80% wouldn't have an unnecessary follow-up...so I'm just thinking..[paused]..oh actually... (Female 3: 18-24 years, %)

5.6.5.2 Experience and analogies

Many interviewees drew upon their previous experiences to determine their own perceived risk. Women who had no history seemed to believe they were at low risk:

Participant: I suppose, like, cancer is not in my family, either breast cancer or any kind of cancer is not in my family at all, so it's something that I kind of don't think about (*Female 17: 25-34 years, 1*)

Participant: No-one as far as I'm aware has any history of breast cancer in my family. So I don't know if what I would do is different to someone who does. (*Female 6: 25-34* years, *f*)

But women who had a family history also reported that they did not feel at increased risk:

Participant: I mean my grandmother died quite young of breast cancer, but I still think my risk is probably average because I think that it's quite a common thing to happen. (*Female 18: 25-34 years, ‡*)

Participant: I don't think it influenced my choices. And at the moment I really don't think I'm at risk at all. (*Female 4: 25-34 years, #*)

The quotes above illustrate a phenomenon known as 'optimism bias', where prior information does not affect perceptions of vulnerability to a particular risk (Weinstein, 1984). A noticeable exemption was when one woman described how her mum's experience with breast cancer did influence her choices:

Participant: My mum went to a van...can I talk to you about that? *CV*: Yeah.

Participant: My mum went to a van, just from work, they had that, and they detected her cancer And three years later the van came round again, and she went. And it was a young girl who'd never...like she was new to it. And she said to my mum, I don't know whether I should do you because you've no breast there. So she said, well, rather than waste your time coming back, I'll do what little bit you've got and then you know I've done it...saved her life that, because apparently they don't do it...and she'd got cancer again in it. And they said, you owe...this is at the Nightingale Centre, and they said, you owe this girl your life. And they caught it, and she had to have it completely removed, lymph glands. So it was worth everything just for that..

CV: Do you think that influences your choices?
Participant: Yeah, a lot. Because she could be dead but for that young girl who didn't know. (Female 12: 50+ years, %)

However, the experience of close friends appeared to have influenced interviewees' preferences for screening:

Participant: A lady I know, she had a mammogram earlier than 50 ... and they detected breast cancer. ...And she has been really ill with it and lost her hair and everything else. And she got very depressed and everything. She was a very strong woman. So I think when you see somebody like that, it does worry you a lot more, thinking, because it does bring it home to you that it could happen to you. (**Female 14:50+ years, %**)

Other women felt that risk was just 'a fact of life' and which could happen to anyone:

Participant: Life is a risk. Everything we do is a risk. (Female 13: 50+ years, %)

Participant: So sometimes there's good, and sometimes there's bad. So I think it's kinda like, it goes along with the screening, you know. (Female 2: 35-44 years, %)

In addition to breast cancer, one interviewee looked at her experience of other ill-health to think about her preferences for the benefits and risks of screening:

Participant: Well, I don't know. I'll tell you something I have actually got on my mind. My boyfriend's mum and sister have got Coeliac disease, so gluten intolerance, and he thinks he's got it, and he won't go to the doctors... like you can't put a price on your health, and I know it would be rubbish to have Coeliac but you'll feel so much better knowing and you can deal with it then. So I guess that is just another example of where I think it's better to know so you can deal with it... (Female 8: 25-34 years, #)

Some women also drew upon their experience of previous breast screening or other screening programmes in order to make their choices:

Participant: I used to have a little bit of a phobia about going for a smear test... It's like if I can avoid it with good reason and it seems like a good reason to - it's like I can justify it. Well this is unreliable or I would worry then that - although it's a bit of a crappy reason, it's still a justification and it's like I feel like I could back that up in my own mind. (Female 18: 25-34 years, #) Participant: I suppose, I don't really know why it is but I, sort of, like - it's like coming back to the smear test thing. It's like I don't really like seeing doctors very much. It's like if I can avoid it with good reason and it seems like a good reason to - it's like I can justify it. Well this is unreliable or I would worry then that - although it's a bit of a crappy reason, it's still a justification and it's like I feel like I could back that up in my own mind. (Female 4: 25-34 years, #)

To aid their interpretation of risk, some women used analogies to compare the magnitude of the different levels. Interestingly, one woman decided 20% was sizeable as she would be happy with a 20% discount in a shop:

Participant: It's royalties on a book. It's 20 per cent off something as in discounts, it's a fairly high percentage compared to most discounts that you're offered. (*Female 13: 50+ years, %*)

For other women, their interpretation of the magnitude of risk was quite difficult to verbalise:

Participant: So I like to, sort of, think about let's just say you're more likely to be eaten by a shark than struck by lightning or whatever, because it's fun. I like things like that. **CV**: Did you find you compare it to any sort of likelihood of other events occurring or do you just...

Participant: Not really but I did notice that like I say, some of - I mean all of the figures seemed fairly low to me. So 3% seemed extremely insignificant, but I would say is that if it was the chance, I don't know why I thought like this, but if it was the one where it's a chance of discovering actual cancer, the lower figures didn't bother me... I can't explain that. They just seemed to unnerve me a bit more and I don't really know why. (Female 18: 25-34 years, #)

5.6.5.3 Visualisation of risk

Women in both the PO and IAP versions of the DCE often read out-loud the risk attributes as natural frequencies (such as '1 in X'):

Participant: Oh well I'm thinking obviously they're both 7 per cent but 1 in 10 people will have unnecessary worry in A, whereas 1 in 20 in B (Female 11: 18-24 years, %)

Participant: One in a hundred people who will have an unnecessary follow up, ten in a hundred, yes. (*Female 10: 25-34 years, %*)

However, this was largely done for the risk of unnecessary follow-up which contained easier levels, rather the detection which contained more complex probabilities (particularly 3% and 14%) which were more challenging to convert into a 1 in X number:

Participant: Because it's like 1 in 5 people are going to end up being told, oh we think there's something like suspicious or not quite right...this one [pause], oh well I'm thinking obviously they're both 7 per cent but 1 in 10 people will have unnecessary worry in A, whereas 1 in 20 in B, yet B is like four times more expensive than A. So B due to money. (Female 11: 18-24 years, %)

Participant: I think to avoid like unnecessary stress and unnecessary procedures which was one in five women, would go through, then I would pay extra money for it... I think 20% is really pushing the limits, because that's one in five people who have to undergo unnecessary surgery. (Female 7: 25-34 years, #)

Participant: I think saying that something happens a fifth of the time that's quite - it's a more meaningful statement to me I would say. (*Female 18: 25-34 years, ‡*)

Without prompting, many women justified the translation of percentages into frequencies as an easier way of understanding the risk and aiding visualisation:

Participant: I'm just trying to picture it. Sounds silly, but like the people. So if it was 100 people it would be three people who found out they had it and it would be 20 people potentially going for chemotherapy, which would be really stressful for them. (*Female* 8: 25-34 years, **#**)

Participant: I just imagined literally one in ten. There's ten and then you're just like one of them and the odds of it all, one in 20 or one in five that kind of thing. So I literally imagined just the number of people rather than how much it would work out in terms of hundreds and thousands of people. So yes, that's how I thought about it in my head. (Female 6: 25-34 years, #)

Participant: So I could visualise little people! But...like in children's books or whatever, but numbers I do have a problem with, especially when the numbers, you know, like 14

per cent and ten per cent, doesn't seem that much of a difference between them...Yes I do find that difficult... Even though this looks like simple numbers, but me trying to...because I think in pictures I think. I'm trying to visualise it. And I'm finding that difficult.. I don't know, I mean for me I'm trying to look at it as though they were people. Little pictures of people there, that's like if there was ten people and five people...(Female 14:50+ years, %)

For two women this was very personalised, and one of these imagined herself in a group of five close friends:

Participant: Yeah, like because I've got a group of five friends, you see, we're in a five, so I always think in that case one of us. That comes to my head quite a bit. (Female 11: 18-24 years, %)

Participant: I think because I'm thinking of seeing people, I know personally, who've had breast cancer detected and are undergoing treatment, and to me three per cent sounds like a very low number of people detected with having cancer. (**Female 14:50**+ **years, %**)

It was very apparent that for many respondents receiving the PO version, there was an initial alarm over the numbers presented, but some women felt that they were able to overcome this as the task progressed:

Participant: I'm thinking, oh God, numbers, percentages! ... I don't find numbers easy to deal with.. For me, I mean, I've never actually been that good at maths or anything, and perhaps I've got a bit of a number blindness I don't know, but I feel a bit stressed when I see numbers .. I knew what my options were and I was quite clear in what I wanted to choose, but it was the numbers that I found difficult to work with... (Female 14:50+ years, %)

Participant: Yeah... but then I was trying to work it out... I was getting confused with the percentages... (Female 3: 18-24 years, %)

Participant: It's just personal to me I think, but it was quite a lot of figures and when you sort of look at one and then look at the other, and then it clicked with me and I was fine, but I think I needed perhaps a little bit more time to understand the percentage. (Female 13: 50+ years, %)

One woman who received the IAP explicitly commented on how it aided her visualisation of risk:

Participant: It's easy to visualise it where you've got the pictograms here because you can quite clearly see it's one line out of five. (*Female 18: 25-34 years, ‡*)

The next section of this chapter will discuss, in more depth, the results presented in sections 5.6.3, 5.6.4 and 5.6.5.

5.7 Discussion

The primary objective of this chapter was to understand more about how individuals complete DCEs which contain a risk attribute and whether this differs between risk communication formats. In section 5.7.1 the main results of the qualitative interviews are discussed in more detail, drawing upon existing literature and exploring the implications of the findings. Ultimately all of the findings rely on the validity of the think-aloud as a method to accurately elicit people's thoughts. The strengths and limitations of the study, particularly the qualitative approach taken, are presented in section 5.7.2.

5.7.1 Key findings

The key findings from the interviews described in this qualitative study will be discussed in the following sections: in section 5.7.1.1, results relating to the method, think-aloud interviews are discussed; in section 5.7.1.2, the key insights about respondents' choice strategies are explored; and in section 5.7.1.3, findings related the risk communication formats are presented.

5.7.1.1 Think-aloud Interviews

The final sample size of nineteen women was in line with other think-aloud studies identified in the systematic review presented in section 5.2 and Appendix 5.1 (see Cheraghi-Sohi et al. (2007) who used a sample of twenty; and Ryan et al. (2009) who used a sample of eighteen). Respondents could choose the location, but most decided to come to the university. This could possibly reflect the anonymous nature of the recruitment and the reluctance of interviewees to invite a stranger into their own home.

The think-aloud method proved useful for some women who were able to continuously verbalise their thoughts, however, for others this was more difficult and the debriefing questions were vital in eliciting their thoughts. The think-aloud results provide further confirmatory evidence that the DCE had validity with many women trading-off the attributes presented and understanding the task at hand, although a few women (in both risk communication formats) reported that it took them a while to familiarise themselves with the task. The unfamiliar nature of the choice task was also reflected in women's comments on the hypothetical situation and the need for reassurance that they were doing the 'right' thing. This serves as evidence of the importance of effective training materials explaining exactly why their choices are of interest and what the task will involve.

5.7.1.2 Choice Strategies

A key assumption in the supporting theories to DCEs is that respondents will behave rationally and choose to maximise their utility. A number of key axioms support this rational choice model and estimation of utility requires respondents to express reflexive, complete, transitive and continuous preferences (Varian, 1992). If a respondent exhibits non-compensatory preferences (no amount of money can compensate them for not having screening), they will choose to go for screening at any cost.

Respondents unwilling to trade-off the cost attribute exhibit attribute-non-attendance. It is satisfactory for cost to have a negligible impact on utility, but statements made by some interviewees implied they were actively ignoring it (even at levels of £1,000). This violation of the continuity axiom through non-attendance to the cost attribute could result in upwardly bias WTP valuations. This is not new; DCE respondents' reluctance to trade-off attributes in a health setting has been identified in other studies (Lagarde, 2012). Similar behaviour has also been reported in environmental DCEs (see Carlsson et al. (2010)) and studies in health DCEs have found insignificant coefficients on cost (Watson et al., 2009). However, this result was surprising given that the estimated utility parameters for the cost attribute were significant in all preference-classes described in Chapter Four. The discrepancy with the quantitative results could be attributed to researcher-presence and social-desirability bias with respondents suggesting that cost was unimportant because it was what they believed was 'right' or in line with 'social norms'.

A few women who did take into account the cost of screening interpreted it as an indicator of the programme's quality. This association could also lead to biased valuation estimates as respondents associate higher cost with more utility. The finding could explain why some (see Augustovski et al. 2013) studies have found the cost attribute to have a positive coefficient in their analysis.

Lexicographic preferences were another simplifying heuristic that some women introduced to make their choices. As with attribute non-attendance, the rank ordering of attributes is a violation of the continuity axiom. Not accounting for lexicographic preferences in the modelling of DCE data may result in bias estimates as the decision strategy introduces systemic errors (Campbell et al., 2006). As this DCE had only three attributes, it is difficult to verify the existence of this simplifying technique.

Another violation of RUT occurred when women made choices using other people's budget constraints. Concern with the affordability of the programmes to other people was a recurring theme in the qualitative analysis. If DCE respondents employ stricter budget constraints when making their choices, this could result in a conservative estimate of true WTP. This result highlights the difficulty of framing cost in a health DCE. However, the effect could have been exaggerated by the presence of an interviewer and a desire to appear generous or philanthropic.

Linked to feelings of altruism was the existence of non-use values where women expressed that they received utility from the programme that was not related to their direct consumption. For example, women stated they felt it was important to have a screening service that was accessible to others and this affordability (determined by the level of price) also had impact on their choice. This is a key finding as it implies women were making choices to maximise the population's health, not just their own.

5.7.1.3 Interpretation of risk

Some women expressed feelings of confusion when the respondents were presented with the PO version, and overall the task appeared to be more daunting. One woman even stated: "*I'm thinking, oh God, numbers, percentages*!". Icon arrays appeared to make the prospect of the DCE task more attractive and engaged women in the survey. Similarly, a few respondents also felt the need to make a calculation and produce a 'correct' answer when the DCE was presented with percentages; despite the training materials and interviewer explicitly stating that there were no right answers. This should be accentuated to DCE respondents possibly throughout the task.

Respondents who received the IAP version expressed statements supporting the use of icon arrays with one woman explicitly reporting that their presence improved her ability to trade-off between attributes. As visualisation of risk was common in interviewees receiving the percentages only version, the addition of an icon array appeared to relieve the cognitive burden of 'imagining women' in an already imaginary scenario. This is in line with previous studies which have found icon arrays aid people's processing of screening information (Hess & Siegrist, 2011).

Some women also created analogies to aid their interpretation of the risk attribute (and one woman compared this to a 20% discount in a shop). It is established that people find risk a difficult concept to understand partly because it is unfamiliar and point-probabilities are not often obvious in day-to-day life. As the structured review of risk communication formats described in section 3.2 of Chapter Three and Appendix 3.1 found, studies employing parallel statistics to assist understanding has had mixed success; generally working most effectively in highly-numerate populations (Galesic & Garcia-retamero, 2013). Providing some equivalent likelihoods as examples in the training materials might have proved useful to some respondents, however, there is also a danger of information overload leading to further confusion.

The structured review of risk communication formats also found that risk experience and perception were important factors in the communication of risk, and women drew on these aspects when trading-off the attributes 'probability of detecting a cancer' and 'risk of unnecessary follow-up'. Concern about breast cancer was also a significant preferenceclass covariate in the quantitative results of Chapter Four. There was some evidence of an optimism bias where by women, even those with a family history, perceived themselves to be at lower than average risk. This is in line with other studies which have looked at perceptions of risk (Weinstein, 1984). Katapodi et al. (2004) reviewed 42 studies examining women's perceived risk of breast cancer and comparing it with an objective estimate. The authors found that not only were they inaccurate; many women were conservative in their estimations. Extensive piloting to ascertain whether an optimism bias exists (or even a reverse tendency to overestimate) would be useful in order to develop a DCE tailored to the particular case study. Attempts to reduce the optimism bias have had limited success (Weinstein, 1998).

5.7.2 Strengths and limitations

The study represents an original empirical piece of work. Whilst other studies have sought to understand how people understand risk and how framing can affect its interpretation (see key examples: Henneman et al. 2013; Sprague et al. 2012; Garcia-Retamero & Cokely 2011), and qualitative research methods have been used to explored women's perception of the risks and benefits of breast screening (Hersch et al., 2013), none have done so in a DCE setting. Similarly, a few studies have used think-aloud methods to gain a deeper-understanding of respondents' choices in a DCE setting (Cheraghi-Sohi et al., 2007; Ryan et al., 2009), however, this is the first empirical study which has used qualitative research to understand how people balance risks and benefits in a DCE setting.

Criticisms of interviews and in particular the think-aloud method, highlight that the method relies solely on what people say, rather than what they really think (Boren & Ramey, 2000). Although an interview schedule was designed (see section 5.4.1.2) to limit sources of bias as much as possible, other limitations such as researcher-presence biases (reflexivity), nonveridicality and reactivity could have influenced the study.

The nature of qualitative analysis means there is no 'right' answer. Instead the approaches taken in this study have tried to be as objective as possible using a clear coding framework, consistently in NVivo® software. The processes of coding and the details of theme development are complex and cannot be articulated in their entire. The details presented in the previous sections demonstrate that the process was transparent and many steps were taken to minimise sources of bias. However, the qualitative analysis in this study was conducted solely by CV and, although extensive discussions with the supervisory team took place, this is acknowledged as a limitation. Quantitative research is not purely objective in its nature either. In choosing to quantify something, the choice of important variables or the nature of what is rational/ irrational is imposed, subjectively, by the researcher. As a consequence, it is in inevitable in both qualitative and quantitative research that the researcher will employ some 'common-sense' that will influence study progression and, as a consequence, the results.

Silverman (1994) describes the notion of 'romanticism' as the influence of society and culture on the subject that is almost impossible for the interviewee or respondent to remove. As many women reported seeing media articles and public health campaigns for breast cancer, it is possible that society's view on the subject could have influenced theirs. Romanticism is distinctly different from researcher bias where the respondent is purposefully disguising their true feelings to align themselves with perceived 'social norms', and is instead a subconscious attention to the subject matter through indirect routes. Additionally, the inclusion of a cost attribute may have made the topic politically-and socially-sensitive. Whilst respondents were encouraged to think of costs as out-of-pocket expenses, there were inevitably some views about the preservation of existing services and affordability for the interviewee.

These sensitivities were not necessarily problems of the qualitative study; they are findings of people's thought patterns when they are completing a risky-DCE. Therefore they are acknowledged, instead, as considerations when thinking about the results. The susceptibility of this study to these sensitivities were considered before embarking on the interviews but, as discussed in Chapter Three, there are many reasons to pursue breast screening as a case study. It was felt that the familiarity of breast cancer allowed a greater

investigation of women's thoughts of the choice task and risk attributes. Choosing a lesser known condition would likely result in other issues (such as increased cognitive burden from clinical explanations). Therefore these limitations are recognised and mentioned in both the analysis and results section of this chapter.

The recruitment of women to this study was constrained by resources and time pressures; however, the sample was varied in terms of key characteristics (age, employment). The interviewees in this study self-selected into breast screening research so for a woman to see an advertisement and respond shows some degree of inherent altruism. A selection of women who were particularly caring could have been interviewed. In addition, these women might have a pre-existing interest in breast cancer which could have, even subconsciously, influenced their choices or behaviour. Similarly, the study relied on women having free time to become actively engaged in an academic study which could have induced some social-class bias. However, the advertisements were placed in many locations and worded to attract all women from a range of backgrounds.

For these particular research questions, think-aloud provided many benefits over other qualitative research methods; most notably the verbalisation of current cognitive processes and concurrent perceptions revealed thoughts in the working memory. The method does not rely on any recall data, which is beneficial given the disadvantages associated with retrospective methods (Lundgrén-Laine & Salanterä, 2010). A serious concern is whether the act of thinking-aloud disturbs the natural decision-making behaviour and therefore generates data that is not the 'usual'.

It is also acknowledged that the transcripts generated by the interviews contained many pauses and mumbles and some respondents required repeated prompting in the experiment. This could have been generated by two sources: 1) respondents struggling to complete the DCE task; and/or 2) respondents struggling to verbalise or communicate their thoughts. The combination of an unusual survey, about a subject they might not be familiar with, and the additional complication of verbalising their thoughts are important limitations of the think-aloud method. The warm-up exercise attempted to reduce the presence of these issues by introducing the think-aloud concept at the start of the interview and allowing the interviewee to 'practice' before the real task.

However, even the most articulate respondent might not be able to say their thoughts outloud. Some processes are not easy to verbalise, and there is a certain amount of subconscious input and impulse in anyone's behaviour (Conrad & Blair, 2009). As a result, in theory, respondents could fictionalise in response to prompts. This problem of 'reactivity' is not unique to the think-aloud method; any interviews encouraging someone to articulate something they find difficult are susceptible to these biases (Leow & Morganshort, 2004; Russo et al., 1989).

The completion time for the DCE was longer than the online study indicating that verbalisation was causing some change in behaviour. Whilst the respondents might just have been 'slowing-down', it could also be encouraging them to consider and deliberate over items that they would otherwise have missed and might not truly reflect typical DCE behaviour. This phenomenon is referred to as 'nonveridicality' and is defined by Russo et al., (1989): "*a protocol is considered nonveridical if it does not accurately reflect the underlying primary process*" (p.760).

These limitations do not discount the findings of this qualitative study. Despite the cognitive burden of both the DCE task and the think-aloud task, there is evidence to suggest that this additional thinking actually improves the generated qualitative data. Taylor & Dionne (2000) recommend that tasks should not be too simple because anything that is familiar or involve routines will not be noticed by the short term memory and respondents will therefore not verbalise their thoughts. Instead, they actually recommend tasks which involve deliberation and a final goal (or choice, in this case) as they believe that this will generate rich and insightful data: "*tasks should be novel and moderately difficult, so as to elicit conscious processing, but not so difficult as to stymie reporting*" (p.415). Further research is required to investigate respondents' choice making behaviour and reduce (if not entirely eliminate) some of the limitations mentioned above.

5.8 Conclusion

This study has helped to understand more about how DCE respondents make choices and trade-off risk attributes. It addressed the key research questions by revealing what people report thinking about in a DCE and how risk attributes are seen and interpreted by DCE respondents. The risk communication format did not substantially alter women's accounts of their choice making but it did appear to relieve some of the cognitive burden in the initial choice sets by aiding the visualisation of risk.

However, qualitative research methods, in particular think-aloud, are not without their limitations. There are important considerations for the interpretation and explanation of these results. The following chapter explores an alternative method, eye-tracking, which is limited in other assumptions but helps to overcome some of the issues (namely reactivity, nonveridicality and refelexivity) identified with the qualitative research presented in this study.

Chapter Six

Benefit-risk trade-offs for breast screening: an eye-tracking study

6.1 Introduction

In the previous chapter, Chapter Five, a qualitative study using the think-aloud method was conducted, where respondents completed a DCE whilst verbalising their thoughts. This attempted to understand how the format of risk may affect respondents' choices or decision-making strategies. However, this qualitative study raised a number of limitations with the think-aloud method. There was evidence to suggest that the act of speaking concurrently may have distracted the DCE respondents. Furthermore, the effort to fully verbalise decision strategies could have *changed* the way the respondents attended to the attributes. Eye-tracking is an approach which may remove some of these issues, providing a useful alternative, or supplementary, method to understand respondents' decision-making processes through the collection of 'objective' (and countable) data.

Eye-tracking could offer some advantages over qualitative research methods in understanding how respondents complete a DCE. As the data provided is not textual, it requires less subjective interpretation. The qualitative study also relied on the respondent's account of their actions. Instead, quantitative eye-tracking data can reveal information about a respondent's overt attentional processes. Eye-tracking data has been used in many areas of research from neuroscience and psychology to computer science and marketing (Duchowski, 2002; Bialkova & van Trijp, 2011).

This chapter presents an investigation into how eye-tracking can be used to understand respondents' attention to information presented in a DCE. The chapter first outlines the research questions to be answered in section 6.2. Section 6.3 describes eye-tracking as a method with reference to the technology's history, its foundations in psychology, and the methods used in published studies. In section 6.4, the psychological theories described in section 6.3 are used to refine the research questions of 6.2 and develop specific hypotheses to be tested. In section 6.5, an empirical study involving an eye-tracking experiment with the DCE designed in Chapter Three is then presented, with the results and discussion described in sections 6.6 and 6.7, respectively.

6.2 Research questions

The study sought to address four key research questions which built upon the overall aim of the thesis (which was to understand if, and how, risk communication formats affect DCE valuations). Eye-tracking was used to inform understanding of the decision strategies, rather than the choice component, of the valuation. The key research questions addressed in this study were:

- Does DCE respondents' attention to attributes differ with the risk communication method?
- To what degree does subjects' retrospective account of attribute non-attendance (ANA) match their visual attention in the DCE?
- 3) Does the way in which respondents seek information in a DCE differ between framing methods?
- 4) How does respondents' cognitive burden differ between risk communication formats?

The research questions were answered using different data, collected using an eye-tracking study with the DCE designed in Chapter Three.

6.3 Introduction to eye-tracking as a method

In order to answer the research questions, different eye-tracking data were collected. The following sections describe how these data were collected, with section 6.3.1 explaining how the technology works based on the physiology of the eye. In section 6.3.2, the key terms and concepts related to eye-tracking are explained and defined. Section 6.3.3 presents a description of the key psychological theories which support eye-tracking as a scientific method, and explains how different measures can proxy different behaviours. In section 6.3.4, existing literature using eye-tracking methods to investigate choice is summarised.

6.3.1 Eye physiology and key concepts

The eye is pulled both horizontally and vertically by extraocular muscles, which allow fixating and tracking of stimuli even when the head remains still (Purves et al., 2001). When the muscles supporting the eye contract, the focus of gaze shifts in a movement called a 'saccade' (Rayner, 2009). Although the eye is never completely still, relatively short saccades or 'jitters' are called 'fixations' (Rayner, 1998). Most eye movement data come under the broad classification of either saccades or fixations (Purves et al., 2001).

Saccades are easily identifiable as the eye moves quickly in response to or in search of visual 'stimuli' or objects of interest. Saccadic behaviour rarely indicates information processing as the movements are so rapid that the brain is unable to consciously realise everything that is scanned, a process known as 'saccadic suppression' (Rayner, 1998). Instead, saccades most often represent a search for information (Kowler et al., 1995).
Saccades are distinctly different to 'micro-saccades', which are involuntary movements whilst an individual is attempting to fixate, and the involuntary movements which occur when an individual blinks (Otero-Millan et al., 2014). Saccades are usually described by the related variables of speed, amplitude and direction (van Beers, 2007).

What constitutes a fixation varies from study to study and is dependent on the stimulus presented (Holmqvist et al., 2011). For example, a familiar picture may be processed quicker than text, and a new diagram may be somewhere in between. Although complex algorithms exist for the identification of fixations in eye-tracking data (Salvucci & Goldberg, 2000), most studies define a threshold for a fixation as a less than one degree of movement (a measure of distance) for between 50 to 200 milliseconds (Manor & Gordon, 2003). Aggregation of the total time spent fixating, including recurrent fixations, is defined as the 'dwell time' to a stimulus.

Eye-tracking data provide a highly detailed record of all the locations that a user has looked at, and reducing these data to a level that can be easily analysed is challenging. One possible approach in the analysis of eye-tracking data involves segmenting coordinates to defined 'areas of interest' (AOI) (Vansteenkiste et al., 2014). AOI can be defined either prior to the experiment or post-experimentally once eye-movement data have been collected. Another approach to reducing the data is the generation of a 'scan path' describing the overall sequence of movements in terms of both saccades and fixations of a respondent, either imposed on a background image of the stimulus or as a colour-coded heat map (Holmqvist et al., 2011).

6.3.2 Technology and equipment

For many years, researchers in psychology have used eye-tracking to investigate cognitive processes (Rayner, 1998). In its most basic form, the research has involved researcherindividual observation of a participants' eyes and manual notes on pupil dilation (Kahneman, 2012). However, more sophisticated methods have since developed in line with changes to, and availability of, technology. In the 1950s, magnetic search coils were used to track people's eye movements which involved placing two coils on the eye, with one circling the iris on a contact lens (Duchowski, 2007). Nowadays, most eye-tracking involves less invasive equipment, commonly with a camera recording data on a computer and complex algorithms to calculate the location of the individual's gaze (Rayner, 2009).

To track eyes, almost all modern devices record the corneal reflection on a camera positioned towards the individual's pupil (Holmqvist et al., 2011). The corneal reflection is a glint, usually in the iris, which allows the machine to calculate the direction of the gaze

using the distance from: 1) the camera to the eye and; 2) the eye to the screen. From the corneal reflection, the X and Y (horizontal and vertical) coordinates, which provide the location of current focus on the screen, are then recorded. The number of times this is logged a second is referred to as the speed of the tracker (Duchowski, 2007). As the eye moves from one position to another, the magnitude of the movement is measured in visual degrees (θ), rather than millimetres, as studies may involve moving stimulus and so the distance between eye and object would change. In a fixed setting, a typical computer monitor of between 17 to 20 inches has a width of 20 to 30 visual degrees (Raney et al., 2014). Figure 6.1 describes how visual degrees (θ) are measured, and how the X and Y coordinates provide a position of gaze.

Figure 6.1: Measurement of visual degrees



Eye-trackers are usually distinguished by their speed and, as a general rule, a good eyetracker has a high sampling frequency and high resolution camera (Holmqvist et al., 2011). A higher sampling frequency allows a more accurate estimation of the fixation duration, as the start of the fixation is revealed earlier and the end revealed later. There is a consensus in the literature that a sampling frequency of 500 Hz is sufficiently powerful to accurately determine fixations and saccades, although anything higher is beneficial (Raney et al., 2014). An eye-tracker equipped with a high resolution camera will be able to detect the corneal reflection more easily and more accurately. Another determinant of a good eyetracking device is its 'latency', which is the time taken to for the computer to make a recording. A substantial volume of processing from headset to screen to recording is required, and for some devices there is a measurable delay in this process (Holmqvist et al., 2011). It is crucial that the eye-tracker is calibrated for each individual to ensure the eye-tracker is recording correctly (Nyström et al., 2013). The calibration procedure involves collecting fixation data from simple points on the screen in order to ascertain the true gaze position of the individual before the experiment begins (Holmqvist et al., 2011). The points are often shown as dots or crosses which move around the screen whilst fixation data are collected. A test of the calibration can be conducted by re-running the sequence and comparing the secondary fixations to the tracker's prediction based on the first calibration data.

The calibration should involve points in all corners of the screen to ensure that the tracker is able to record in all areas (Holmqvist et al., 2011). In the corners and edges, the corneal reflection can disappear, which therefore invalidates the computer's calculations as well as result in missing data. Similarly, for individuals with visual aids (glasses, contact lenses) or heavy eye-makeup, the far corners can often induce another reflection which may confuse the recording and create anomalous data.

Tracking devices can either record both eyes (binocular) or a single eye (monocular) (Holmqvist et al., 2011). When both eyes are recorded, an average of the horizontal and vertical coordinates from each eye are taken. However, most people generally have an 'active' and 'lazy' eye and literature suggests that the active, dominant eye should only be tracked (Nyström et al., 2013). If a participant performs poorly in the calibration, then an alternative eye should be tried.

There are three broad categories of modern eye-tracking devices: 1) head mounted; 2) remote; and 3) head-supported towers. Head mounted eye-trackers, such as smart glasses or helmet cameras, offer participants some freedom, but these are harder to calibrate and can be cumbersome to wear. These eye-trackers are often used to understand how objects are attended to in a dynamic situation, for example, whilst the participant is engaged with a shopping activity (Holmqvist et al., 2011). Remote eye-trackers involve no head restraint and instead let the participant move freely, with algorithms used to detect non-eye movements (Bohme et al., 2006). However, the additional calculations to distinguish head and eye movements are a burden to the processing capacity of the computer, and generally result in a lower frequency and, as a consequence, have decreased precision (Holmqvist et al., 2011).

Head-supported towers involve the use of a forehead and chin-rest. Whilst being contactless, these can be uncomfortable and unnatural for some participants. These devices are also often immobile, due to their heavy processing power, and require stability of the head because of their high frequency. However, head-supported towers are the most accurate and precise equipment available for researchers. For studies where the individual is not required to move and the stimuli are stationary (such as a survey), a head-supported tower eye-tracker provides the best quality data (Holmqvist et al., 2011; Raney et al., 2014).

6.3.3 Psychological foundations

Cognitive processes are incredibly complex and it may be impossible to measure what any individual is actually thinking at any one time. However, psychologists have studied eyemovements in order to understand more about people's brain functioning and how information is processed (Rayner, 1998). The 'eye-mind hypothesis', suggested by Just & Carpenter (1980), provided the underpinning to most psychological analyses of eyetracking data. The following sections describe the three key behaviours expanded to answer the research questions stated in section 6.2. The theoretical foundations supporting visual attention, as acquired through fixation data, are described in section 6.3.3.1. In section 6.3.3.2, information searching behaviour as exhibited through fast, saccadic movement, are described. In section 6.3.3.3, 'cognitive pupillometry', a measure of task difficulty through pupil dilation, is described.

6.3.3.1 Visual attention

The logic of using eye-tracking data as criterion data is based on a generally accepted assumption that where people look is indicative of their thoughts. Just & Carpenter (1980) assertively stated that: *"there is no appreciable lag between what is fixated and what is processed"* (p.331). Therefore attention, as measured in fixations, can be thought of as a quantification of an individual's information processing. Although it could be argued that individuals who are day-dreaming may inadvertently fixate on something whilst their thoughts are somewhere else, this behaviour is often shown through attention to areas of white-space with no information, or off the screen entirely (Rayner, 1998). This dreaming behaviour is also known as 'covert attention' and is distinctly different to information-acquiring fixations to stimuli (Henderson et al., 1989).

6.3.3.2 Information searching

Data generated between fixations are generally classified as saccades. These rapid movements occur when individuals are searching for information but are not taking account of the stimulus they are scanning (Kowler et al., 1995). Saccadic patterns in eyetracking data have been used to explain visual responsiveness, such as how quickly an individual can find an answer (Findlay, 2009). In the context of choices, saccades have been used to understand how individuals seek information to make a decision, with research suggesting vertical movements in line with EUT (Arieli et al., 2011, 2009).

6.3.3.3 Cognitive pupillometry

In addition to eye movements, some eye-trackers have the capacity to collect other data on the participant's eye. Pupillometry is the measurement of the pupil, and cognitive pupillometry refers to the change in pupil size due to task burden (Hartmann & Fischer, 2014). When individuals are thinking hard about something which is difficult or requires significant memory load, the pupils may dilate (Laeng et al., 2012). In early studies, this was an incidental finding discovered when observing participants in laboratory tasks (Kahneman, 2012). However, there are now many data-driven studies which have empirically demonstrated a relationship between task complexity and pupil dilation (Binda et al., 2014; Beatty & Wagoner, 1978).

6.3.4 Research to date

Whilst eye-tracking has grown as a method, and is frequently used by psychology departments in academic institutions (Rayner, 1998), there are very few examples in either health economics or economics. The results of the systematic review presented in Appendix 5.1 did not identify any published healthcare DCEs which had used eye-tracking methods.

A DCE published in 2015, that elicited preferences for food consumption, used eyetracking data to investigate visual ANA as measured by the number of times respondents re-attended attribute levels in a choice set (see Balcombe et al. (2015)). A study by Bialkova & van Trijp (2011) looked at respondents' attention to food labelling in a choice survey set up in a similar way to a DCE. In this study, the analysis lacked detail but the authors used fixation and response-time data to analyse the effect of different labelling of yoghurt packages on consumer choice. Despite parallels in these choice-based eye-tracking studies, neither looked at a risk attribute in a DCE, and neither involved a healthcare good or service.

The next section of this chapter, section 6.4, presents an empirical study with an innovative application of eye-tracking to the DCE designed in Chapter Three.

6.4 Hypotheses

Section 6.3 identified outcomes of eye-tracking data that potentially have value in this PhD. Linking the psychological theories and research questions of section 6.2, four hypotheses were generated, each matching the respective research question.

Hypothesis one: The 'eye-mind hypothesis' introduced in section 6.3 suggested attention to stimulus indicates information is processed by the participant. Therefore, an individual

completing a DCE and attending an attribute more (longer) would be conducting more information processing resulting in a better, or more informed, choice. Therefore it was hypothesised that participants presented with the IAP version of the DCE would pay more attention (and make more informed choices) to the risk attributes in comparison to those who received the percentages only. Visual attention was measured by: 1) number of fixations; 2) longer fixations; and 3) in total, a longer dwell time.

Hypothesis two: Individuals completing a DCE who reported not attending an attribute would pay less visual attention to these attributes. Attention to an attribute was measured as total fixation duration (dwell time) to the attribute.

Hypothesis three: The communication of risk could affect the way respondents' searched for information in order to make their choice in the DCE.

Hypothesis four: It was hypothesised that women who received this risk communication format would have reduced pupil dilation in the task as these would aid choice-making and make the task less cognitively burdensome.

6.5 Methods

The following sections describe the methods used in an eye-tracking experiment, where respondents completed the DCE survey designed in Chapter Three whilst their eyes were monitored.

6.5.1 Recruitment

The sample selection criteria were limited to females, fluent in English and between 18 and 70 years of age (for reasons discussed in Chapter Three). As stated previously in section 3.4.1 of Chapter Three, 70 years was chosen as a maximum age as this is the current cutoff for routine screening in England (Independent UK Panel on Breast Cancer Screening, 2012). No other sample restrictions were included as initial discussions with breast cancer experts in the design of the DCE revealed that they were interested in the views of all women (see Chapter Three). In light of the expert advice, both the internet panel (see Chapter Four) and the qualitative study (see Chapter Five) sampled women in the 18 to 70 age categories and therefore, for consistency, the sampling frame for this eye-tracking study continues with the same inclusion criteria.

The advertising strategy replicated that of the qualitative study described in section 5.4.1.1 of Chapter Five. Examples of the advertisements can be found in Appendices 6.1 and 6.2. Following the recommendations of the University's ethics committee, responders to the advertisements were only incentivised at first contact with the researcher, when they were

informed about the £10 Amazon voucher, as per the recommendation of the ethics committee. At this first point of researcher contact, the potential participants were also provided with an information sheet containing frequently asked questions (Appendix 6.3). The advertisement responders were then asked to read the sheet, take time to think about the study and then, if they were interested, to arrange a visit to the eye-tracking laboratory located in the Zochonis Building at The University of Manchester.

The study aimed for a sample size of 40 women, on the basis of a similar choice study by Bialkova & van Trijp (2011), who also used a sample of 40 and the results of a pilot study conducted in May 2014 (see Appendix 6.4 for a description). The number of subjects was also restricted by time constraints (as the eye-tracking laboratory was shared, time using the equipment was limited) and resource constraints (if the £10 Amazon voucher thank-you was maintained). Therefore recruitment continued until 40 successful experiments had been completed.

6.5.2 Apparatus

The head-mounted EyeLink® 1000 was used in this experiment (SR Research, 2012). The eye-tracking device calculates the participant's gaze position using a camera to detect the corneal reflection due to an infrared illuminator. The device recorded the eye position a thousand times a second, every millisecond (ms). The head rest was positioned 43 centimetres from the screen, as per the manufacturer's recommendations, and this distance was re-measured for every participant. Whilst the machine had a capacity for binocular recording, monocular recording of the dominant eye was conducted for the reasons discussed in section 6.3.2. The experiment took place in a dark, window-less room with minimal luminosity. Choices for the DCE were made via a handheld games controller.

The training materials and choice sets of the DCE survey designed in Chapter Three was programmed for the eye-tracker with assistance from EyeLink® Experiment Builder software (SR Research, 2012). The DCE was identical to that used in Chapters Four and Five except that no video was included in the training material, as the video file was too large to be loaded into the experiment. Eye-tracking data relating to the video would have been disregarded in the analysis as data relating to a dynamic stimulus would have required completely different analytical methods. The eye-tracking DCE also included three additional warm-up questions, which allowed participants to acquire a familiarity with the handheld controller buttons. All images in both risk formats were exactly the same size to the nearest pixel. The set-up of the eye-tracking apparatus is shown in Figure 6.2.



A shows the screen which displayed the training information and choice sets. B represents the camera recording the participant's eye. C is the infrared illuminator which allowed recording in the dark room. D points to the handheld controller that participants used to make their choice.

6.5.3 Data collection

Once a woman had arranged to visit the laboratory, she was asked again to read through the information sheet and, if she felt happy, consent to the study by signing the form shown in Appendix 6.5. The women who consented were randomly assigned (using a random number table) to receive either the PO version from design block one or two, or the IAP version from design block one or two. The participants were then asked to place their chin on the head rest, make themselves comfortable, and refrain from speaking. The calibration process then began. If the calibration was 'good' (the corneal reflection was consistently recorded), then it was validated through an additional re-calibration which ensured all corners of the screen were recordable before the survey began. If the calibration or validation failed, the recording eye was changed and the procedure restarted. In the event that neither eye could be calibrated, the participant was thanked for their time and informed that the experiment could not be completed. Whilst answering the DCE, a between choice set calibration occurred called 'drift correction'. This was to correct for any 'drift' which had occurred and to improve the accuracy of the collected data. Drift occurs when the position of the eye moves slightly and the calibration is no longer perfect. In the event that the drift correction indicated the participant had moved their head or the tracker had stopped recording, a complete re-calibration and validation procedure was conducted. The eye-movements were recorded for each of the eleven choice sets as separate independent 'trials'.

Background questions were completed by the participant on an iPad after they had finished the DCE with the eye-tracker. Questions included socio-demographic details and selfreported ANA. The iPad-based survey used in this experiment can be found in Appendix 6.6.

6.5.4 Analysis

The effect of risk attribute format on a range of eye-tracking-based outcomes was investigated to answer the research questions stated in section 6.2, and test the four hypotheses presented in section 6.4. Section 6.6.4.1 describes the first stage of preparing the collected data for analysis. Section 6.6.4.2 explains how outcomes of interest were collapsed to an individual-trial level, and in section 6.6.4.3, the regressions conducted to analyse the eye-tracking data are specified.

6.5.4.1 Data preparation

The following section describes how the eye-tracking data was prepared for the analysis, with generated outcomes to answer the research questions stated in section 6.2. The first stage of data reduction, defining AOI, is explained in section 6.5.4.1.1. The key outcomes of interest: fixations, saccade and pupil size, are described in sections 6.5.4.1.2, 6.5.4.1.3 and 6.5.4.1.4, respectively.

6.5.4.1.1 Areas of interest

The first stage of analysis involved reducing the data to pre-defined AOI. This process reduced the X and Y coordinates and pupil size data into a series of variables of interest and reduced the file to a workable size. AOI reflect the participant's attention to a particular segment of a task. These AOI quadrants were defined for each choice set on the screen, based on the sections of the task that might be stimulating for the participants. Figure 6.3 shows the AOI defined in each choice set, for example, the attribute titles, the levels presented, and the response options. Of particular interest in this study were the outlined segments; the AOI for the attribute levels. The areas outside the AOI were used to

measure the amount of gaze in the 'white space', and attention to this white space provided an approximate measure of validity.





6.5.4.1.2 Fixations

Fixations were measured in milliseconds and were conservatively defined as a less than 1° movement for 75 milliseconds. This fixation duration threshold was defined based on previous studies, such as Bialkova & van Trijp (2011) who assigned a threshold of 80 milliseconds to respondents completing a choice set of images and numbers. If a fixation was under this threshold, and another fixation occurred within 1° of the original fixation, then the fixations were merged together. Merging adjacent fixations allowed identification of fixations which may have been missed due to measurement errors.

Collecting fixation data allowed analysis of information processing via the respondents' attention to the choice task. Fixation data can be analysed in terms of: 1) the number of fixations; 2) the average length of a fixation; and 3) the total dwell time (the sum of all fixation durations) to an AOI. These three outcomes were of interest in the analysis of the eye-tracking data and can be useful in testing hypothesis one, as they each provide different measures of visual attention.

Self-reported ANA was captured as a dummy variable depending on the respondents' answer to the iPad-based survey question (see Appendix 6.6, p.495). These data were used to compare mean dwell time to each attribute which can be thought of as measuring the

latent scale from which individuals assign ANA in the self-reported question. These data were used to test hypothesis two, which theorised respondents who report ANA would have a shorter mean dwell time to those attribute AOI.

6.5.4.1.3 Saccades

No amplitude threshold was defined for saccades and any observations that were not fixations constituted movement. Saccades were measured by their direction (angle of movement). Amplitude and velocity were not of interest in this study because they would yield little information about respondents' decision-making strategies.

The EyeLink® 1000 tracker records degrees of direction between -180° and 180°, with 0° being a perfectly horizontal movement to the right. A rightward saccade was defined as a movement between -45° and 45°; a downward saccade was defined as a movement between -135° and -45°; a leftward saccade was defined as a movement between -135° and 135°; and a upward saccade was defined as a movement between 45° and 135°. Figure 6.4 describes each directional movement and its definition in degrees.

Figure 6.4: Definition of direction from saccade angles



Blinks were identified by the EyeLink®'s in-built software (SR Research, 2013). When a participant blinks whilst their eyes are being tracked, the gaze shifts down in what could be interpreted as a saccade. However, a blink was identifiable as it was immediately followed by a missing pupil image on the camera as the eyelid closes. These data were acknowledged but disregarded from the analysis.

The saccade data of interest was the number of saccades made, and the direction of these saccades. These data tested hypothesis three which sought to understand how respondents seek information in a choice set before selecting their preferred alternative. No specific hypothesis existed about the number of saccades a respondent may make, but more up-

down movements could suggest the choices were made in line with EUT and Lancaster's Theory as the respondent was taking into account the alternative as a whole.

6.5.4.1.4 Pupillometry

Pupil size was calculated by the eye-tracker which counts the number of black pixels on the camera image of the eye to identify a measurement of the pupil diameter. The EyeLink® 1000 is generally regarded as a good measurer of pupil size, as alternative trackers use an ellipse fitting highly affected by noise. Pupil dilation was calculated as the difference between the minimum pupil size and maximum pupil size. This was kept as an absolute measure rather than as a percentage which can be inflated when baseline pupil size is small (Wang, 2010).

With the EyeLink® 1000, pupil size data were not calibrated, and units of pupil measurement typically vary between studies. Pupil size was recorded as an integer number, based on the number of pixels but measured in arbitrary units meaning that results cannot be compared across studies, or even within studies if there was inconsistent luminosity or the stimuli appeared at different locations on the screen, as this can affect the measure of pupil size. However, in this experiment the choice set stimuli, either percentages or icon arrays, occurred in precisely the same location on the screen and the head-mount, screen and camera were identically located for each subject. In addition, all experiments took place in the same laboratory with identical equipment set-up and light sources, maintaining the same luminosity.

Pupil size data was analysed as: 1) the average pupil size per individual fixating to an attribute in a choice-set (trial); and 2) the average change in pupil size per individual fixating to an attribute in a trial. The pupil size data were used as a measure of cognitive burden and was used to test hypothesis four. Hypothesis four suggested that larger pupils indicate a more demanding task, and a large change in pupil size could be the result of a question that is more cognitively challenging.

6.5.4.1.5 Response time

Response time was measured from the end of the drift correction to the participant pressing a controller button to make their choice, and was recorded to the nearest millisecond. A shorter response time could indicate that respondents were able to make their choices more quickly because the information was easier to acquire.

Variables derived from the background data collected from the iPad-based survey at the end of the experiment were not used as covariates in the analysis of the eye data. As respondents were randomised to receive either risk format of the DCE, any observable differences in characteristics would have occurred by chance and there was no hypothesised-relationship between socio-demographic characteristics and visual attention.

6.5.4.2 Collapsing data to derive outcomes of interest

For most outcomes, data were available at a saccade or fixation level as made by each participant, in each choice set. To analyse the effects of risk format, the eye-tracking data of all participants completing all choice sets were collapsed to an individual-trial level. This allowed the results to be interpreted as a statistic per choice set for a respondent receiving the icon arrays or percentages only. For mean outcomes, this was achieved by taking an average of outcomes at the fixation/saccade/pupil size/response duration level for each individual in each choice-set (trial). For aggregate outcomes (dwell time) a sum of outcomes was taken. Number outcomes (number of saccades, number of fixations) were calculated through counts. For binary outcomes (direction of saccade), this was calculated as the proportion of times at the saccade level (in a trial) where the binary variable equalled unity (1).

To illustrate the collapsing of data, a highly simplified example of fixation-level data can be found in Figure 6.5. Figure 6.5 shows hypothetical-participant 1 making 9 fixations in choice set 1 and attending all attributes except the cost of alternative B. The participant attended the detect attributes 4 times, the risk attributes twice, and the cost attribute once. The participant also looked at the white space once for 81 milliseconds and the text about detection in the 'No Screening' option for 150 milliseconds.

ID	Fixation duration	Pupil size	AOI	Choice set
1	78	1000	RiskA	1
1	123	1100	DetectA	1
1	85	1220	DetectB	1
1	196	1190	CostA	1
1	178	1400	Detect B	1
1	93	890	RiskB	1
1	182	850	DetectA	1
1	81	1000	•	1
1	150	1100	None_detected	1
1	113	1340	RiskA	2
1	76	950	RiskB	2

Figure 6.5: Example of eye-tracking fixation data

The collapsed fixation data analysed would then take the form of Figure 6.6, with the key outcomes of interest presented per participant, per trial. With key variables including

number of fixations (NOF), mean fixation duration (MF), dwell time (DT), mean pupil size, change in pupil size and self-reported ANA at an individual-trial level. The average fixation duration to the detection attributes was 142 milliseconds, to the risk attributes 85.5 milliseconds and to the cost attribute was 196 milliseconds. The total dwell time was 568 millisecond to the detect attributes; 171 milliseconds to the risk attributes; and 196 milliseconds to cost. The mean pupil size for the whole trial was 1092 and the change in size was 550.

ID	AOI	Choice set	NOF	MF	DT	Mean pupil size	Change in pupil size	ANA detect	ANA risk	ANA cost	Risk format
1	Risk	1	2	85.5	171	945	110	0	0	1	%
1	Detect	1	4	142	568	1143	550	0	0	1	%
1	Cost	1	1	196	196	1190	0	0	0	1	%
1	Risk	2	4	126	420	945	15	0	0	1	%
1	Detect	2	6	156	672	1126	360	0	0	1	%
1	Cost	2	2	78	158	1191	70	0	0	1	%
1	Risk	3	3	90	246	949	110	0	0	1	%
1	Detect	3	5	156	736	1104	442	0	0	1	%
1	Cost	3	3	98	282	1193	50	0	0	1	%

Figure 6.6: Example of collapsed fixation data to key AOI

6.5.4.3 Regression-based analysis

The collapsed data were analysed using ordinary least squares (OLS) in a linear regression model attempting to fit the observed data by minimising the sum of the squared residuals (Wooldridge, 2008). This method was chosen to investigate the effect of the risk communication method on the outcomes of interest. For all outcomes, the regression specification in equation 6.1 was estimated separately for the detection, risk and cost attributes:

$$y_{i,c} = \alpha + \beta IAP_i + \varepsilon_{i,c}$$
 [6.1]

 $y_{i,c}$ denotes the outcome for individual *i* in choice set *c*. *IAP_i* is a binary variable which equals zero if individual *i* completes the DCE where risk was framed as a percentage only, and equal to unity if individual *i* completes the DCE where risk was formatted with an additional icon array. $\varepsilon_{i,c}$ is a zero mean error term, assumed uncorrelated with the method of risk formatting. Clustering of standard errors at an individual level was included to allow for observations by the same individual to be correlated. The regression of equation 6.1 produced the estimated coefficients shown in equations 6.2 and 6.3:

$$\hat{\alpha} = \bar{y}^{PO} \qquad [6.2]$$

$$\hat{\beta} = \bar{y}^{IAP} - \bar{y}^{PO} \qquad [6.3]$$

The estimated constant, $\hat{\alpha}$, is the average of individual-trial level outcomes for individuals who completed the DCE with risk framed as a percentage only, \bar{y}^{PO} and $\hat{\beta}$ is the average difference in average outcomes between individuals who completed the DCE with risk attributes formatted as both a percentage and icon array (\bar{y}^{IAP}) and individuals who completed the DCE with risk communicated as a percentage only (\bar{y}^{PO}).

To study the effects of self-reported ANA, the outcome was measured as the mean dwell time to an attribute. To estimate this effect, the regression specification in equation 6.4 was estimated separately for the detection, risk and cost attributes:

$$y_{i,c} = \alpha + \beta ANA_i + \varepsilon_{i,c} \qquad [6.4]$$

 $y_{i,c}$ denotes the total fixation duration for individual *i* in trial *c*. *ANA_i* is a binary variable which equals zero if individual *i* reported that they attended to an attribute, and equal to unity if individual *i* reported not attending to an attribute. $\varepsilon_{i,c}$ is a zero mean error term, assumed uncorrelated with self-reported ANA. The regression of equation 6.4 produced the estimated coefficients of 6.5 and 6.6:

$$\hat{\alpha} = \bar{y}^A \qquad [6.5]$$
$$\hat{\beta} = \bar{y}^{ANA} - \bar{y}^A \qquad [6.6]$$

The estimated constant, $\hat{\alpha}$, is the mean dwell time per attribute per trail for individuals who reported attending an attribute (\bar{y}^A) , and $\hat{\beta}$ is the difference in average (mean) dwell time between individuals who reported not attending to an attribute (\bar{y}^{ANA}) and individuals who reported attending to an attribute (\bar{y}^A) .

Table 6.1 summarises the different types of eye-tracking data, how the outcomes are defined and their units of measurement, the meaning of \bar{y} and $\hat{\beta}$ for each OLS regression, and the interpretation of the regression results.

Label	Outcome	Definition	Туре	Unit	\overline{y}	Â	Interpretation
					in OLS regression	in OLS regression	_
Experi- ment	Response time (RT)	Time between choice set appearing and choice being made	Mean	Milli- seconds	Mean RT from starting a trial to pressing a button	Difference in mean RT when risk is presented with an icon array	Hard experiments take longer to complete
Fixation	Number of fixations (NOF)	Number of separate occasions an individual fixates on information	Count	NOF	Mean NOF by an individual to an attribute in a trial	Difference in mean NOF when risk is presented with an icon array	More fixations mean more information processing
Fixation	White space fixations	Number of separate occasions an individual fixates on areas containing no information	Count	NOF	Mean NOF by an individual to a white space in a trial	Difference in mean NOF when risk is presented with an icon array	More fixations to white space means confusion/ no interest
Fixation	Average fixation duration	Mean time of a fixation	Mean	Milli- seconds	Mean duration of a fixation to an attribute by an individual in a trial	Difference in mean duration when risk is presented with an icon array	Longer fixations mean more information processing
Fixation	Total dwell time (DT)	Total time of all fixations	Sum	Milli- seconds	Mean DT to an attribute by an individual in a trial	Difference in mean DT when risk is presented with an icon array	Longer DT means more information processing
Fixation and ANA	ANA dwell time	Total time of all fixations	Sum	Milli- seconds	Mean DT to an attribute by an individual in a trial	Difference in mean DT when respondents self-reported ANA	Self-reported ANA should mean shorter DT as attributes ignored
Saccade	Number of saccades (NOS)	Number of separate occasions an individual searches for information	Count	NOS	Mean NOS by an individual in a trial	Difference in mean NOS when risk is presented with an icon array	No hypothesis
Saccade	Saccade direction	Direction of eye-movements in information seeking	Count	Binary (0=up-down; 1=left-right)	Proportion of saccades to an attribute in a trial, which are left-right	Difference in mean proportions when risk is presented with an icon array	More up-down movements mean more whole-good calculations
Pupil	Average pupil size	Mean pupil size in a fixation	Mean	Pixel-units	Mean pupil size of an individual when fixating to an attribute in a trial	Difference in mean pupil size when risk is presented with an icon array	Larger pupil size indicates more cognitive burden
Pupil	Pupil dilation	Change in pupil size whilst fixating (largest-smallest)	Mean	Pixel-units	Mean difference in pupil size of an individual when fixating to an attribute in a trial	Difference in mean pupil size change when risk is presented with an icon array	Larger changes in pupil size indicates more cognitive burden

Table 6.1: Outcomes of interest in the eye-tracking data

6.6 Results

The results section starts with a description of the participants' responses and the validity of the experiment in sections 6.6.1 and 6.6.2, respectively. The results of the regressions relating to the visual attention and ANA are presented in sections 6.6.3 and 6.6.4, respectively. The results relating to the saccades made by participants are described in section 6.6.5 and the pupil size regression results are presented in section 6.6.6.

6.6.1 Participants

In total, 42 women were recruited, although two women were excluded because their eye movements could not be accurately recorded due to calibration failure. In total, forty women aged between 18 and 63 years old were included. Twenty women were randomised to each risk version of the DCE. Within each version, ten women were randomised again to each of the two design blocks.

The complex recruitment strategy confounded with women from the qualitative interviews presented in Chapter Five makes it difficult to attribute a precise response rate for this particular study. For more information on the recruitment results, see Appendix 5.4.

The whole experiment, including the calibration, choice questions, and the completion of background questions on the iPad, took approximately 20 minutes. The average response time for a single choice was 13388 milliseconds (13.4 seconds), as shown in Table 6.2. This was slightly shorter (half a second) for the IAP format; however, the difference was not statistically significant.

 Table 6.2: Mean response time in milliseconds from starting the choice set to pressing a button, and
 differences between risk formats

Duration
13387.67 ms
(685.21)
12942.85 ms
-444.82 ms
(968.54)
440

Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

6.6.2 Experimental validity

Figures 6.7 and 6.8 show examples of the eye-movements of a respondent from the IAP and PO DCEs, respectively. A visual examination of the scan paths of participants indicated that the eye-tracking experiment had face validity, with almost all fixations to the

pre-defined AOI. Table 6.3 shows that, on average, each participant made only two fixations to areas of white space (with no information) per choice-set question. There was no significant difference between the two risk formats.

 Table 6.3: Mean number of fixations to white space per choice set, and differences between risk

 formats

Mean number of white space fixations
2.379
(0.68)
2.403
0.024
(0.02)
440
-

Standard errors in parentheses, * *p*<0.05*,* ** *p*<0.01*,* *** *p*<0.001

6.6.2.1 DCE validity

Results of heteroscedastic conditional logit models using choice data from the eye-tracking participants can be found in Appendix 6.7. All coefficients possessed the expected sign (negative coefficients on the attributes 'risk of unnecessary follow-up' and 'cost of screening programme'; and a positive coefficient on the attribute 'probability of detecting a cancer'). However, the coefficient on the cost attribute was statistically insignificant. Two participants failed the internal validity check (a failure rate of 5%).

Figure 6.7: Example scan path from a respondent completing the DCE with icon arrays and

percentages



Figure 6. 8: Example scan path from a respondent completing the DCE with percentages only



6.6.3 Visual attention: fixation and dwell times

Hypothesis and research question one sought to understand if, and how, visual attention differed by risk framing communication format through an examination of the effects on fixations. As shown in Table 6.4, the risk and detection attributes had the highest number of fixations, attracting almost twice as many fixations as the cost attribute. The average number of fixations made when looking at a choice-set was not significantly different between the two risk versions of the DCE.

Table 6.4 also shows that not only did DCE respondents fixate more to the risk and detection attributes, the fixations to these attributes were also longer in duration. Although respondents made more fixations to the risk attribute, the mean fixation time was shorter. There was no significant difference in fixation duration between the risk communication formats.

When the number of fixations and the fixation duration were aggregated to measure the complete dwell time to an attribute, results were maintained. The eye-tracking participants spent a total of 1.8 seconds, 1.9 seconds and 0.8 seconds on the attributes detect, risk and cost, respectively. Eye-tracking participants who received the IAP spent longer processing all attributes, although the difference was not significant.

Table 6. 4: Visual attention to each attribute in a trial (hypothesis one)

	Attribute						
Risk format	Probability of detecting a	Risk of unnecessary	Lifetime cost of				
	cancer	follow-up	screening programme				
Mean number of fixations to each attribute in a trial, and differences between risk formats							
ΡΟ % [α]	7.605	8.500	4.045				
	(0.71)	(0.84)	(0.68)				
IAP 🕴	8.137	9.832	4.400				
Icon array difference $[\beta]$	0.532	1.332	0.355				
	(0.97)	(1.46)	(1.04)				
Mean duration of a fixati	on to each attribute in a trial i	n milliseconds, and diff	erences between risk				
formats		,					
ΡΟ % [α]	237.267 ms	205.846 ms	157.733 ms				
	(14.95)	(9.08)	(12.16)				
IAP 🛊	261.727 ms	201.465 ms	164.731 ms				
Icon array difference $[\beta]$	24.460 ms	-4.38 ms	6.998 ms				
	(20.61)	(11.94)	(16.54)				
Mean dwell time (total du	ration of fixations) to each at	tribute in a trial in milli	seconds, and differences				
between risk formats			·····				
ΡΟ % [α]	1789.427 ms	1919.718 ms	829.427 ms				
	(219.59)	(292.12)	(162.30)				
IAP 🕴	1982.659 ms	2072.573 ms	897.536 ms				
Icon array difference $[\beta]$	193.232 ms	152.855 ms	68.109 ms				
•	(275.11)	(396.19)	(235.14)				
Observations	440	440	440				

*Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001*

6.6.4 Attribute non-attendance

Hypothesis and research question two sought to understand if, and how, self-reported ANA matched ANA in the eye-tracking data through a comparison of fixations. When the eye-tracking dwell-time data were compared with self-reported ANA, it was found that the mean dwell time was significantly lower for people who reported non-attendance to these attributes (for' probability of detecting a cancer' and 'risk of unnecessary follow-up'). The results in Table 6.5 shows the difference was considerable, with non-attendance to risk resulting in a 25% lower dwell time to the risk attributes, and non-attendance to detect resulting in over 40% shorter dwell times. However, there was no statistically significant difference between dwell-times to the cost attribute between those who reported attending it and those who stated ANA.

Table 6.5: The effect of self-reported ANA on total dwell-time per trial in milliseconds (hypothesis two)

ANA	Probability of detecting a cancer	Risk of unnecessary follow- up	Lifetime cost of screening programme
No self-reported ANA	1970.293 ms	2010.305 ms	881.676 ms
[∞]	(145.70)	(203.04)	(113.22)
Self-reported ANA	1127.796 ms	1443.909 ms	808.900 ms
Difference	-842.497 ms***	-566.396 ms**	-72.776 ms
[6]	(231.37)	(203.04)	(344.48)
Observations	440	440	440

Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

In this study, there was complete visual attention to the attribute 'probability of detecting a cancer' by participants who received the PO version. However, there were some participants who did not attend to an attribute at all in some choice sets, exhibiting complete visual ANA. As Table 6.6 shows, when risk was framed as a percentage only, complete non-attendance to the risk attribute occurred in seven choice-sets, compared to only once when risk was framed with an icon array. Table 6.6 also shows that in terms of complete visual ANA, the cost attribute was most neglected by the participants with 24 choice sets completed with no visual attention to the attribute, at all.

Risk format	Detect	Risk	Cost
IAP 🕴	1	1	15
	(0.23%)	(0.23%)	(3.41%)
PO %	0	7	9
	(0.00%)	(1.59%)	(2.05%)
Total	1	8	24

As a percentage of all choice-sets in parentheses

6.6.5 Saccades

Hypothesis and research question three sought to understand if, and how, respondents searched for information in a choice-set differed by risk communication formats. Information searching was evaluated through analysis of the saccade data, with more upward-downwards eye-movements in line with EUT and Lancaster's Theory. There was no significant difference between the two risk communication formats in terms of the average number of saccades a participant made in each trial (around 48 movements). When it came to the direction of the saccade, participants completing the DCE in both risk versions made more horizontal (left-right) saccades than vertical (up-down). As table 6.7 shows, participants who received the IAP version made significantly more upward-downwards (48.9%) movements compared to the participants who received PO (43.5%).

Table 6.7: Number and direction of saccades in a trial (hypothesis three), and difference between risk

formats

Risk format	Number of saccades	Percentage of saccades moving vertically	Percentage of saccades moving horizontally
ΡΟ % [α]	47.955	43.5%	56.5%
	(2.24)	(0.01)	(0.01)
IAP 🕴	48.228	48.9%	51.1%
Icon array difference [β]	0.273	5.4%*	-5.4%*
	(3.17)	(0.02)	(0.02)
Observations	440	440	440

*Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001*

6.6.6 Pupil dilation

Hypothesis and research question four sought to understand if, and how, the cognitive burden of the task differed by risk communication format through an examination of pupil size. The average (mean) pupil dilation for the overall task and to individual attributes are presented in Table 6.8. The number of observations reflects the fact that some attributes had complete visual ANA (as shown in Table 6.6). For all attributes, the mean pupil size was smaller for respondents completing the icon array version of the DCE, this was also the case for the change in pupil size for the detection and cost attributes. However, none of these differences were statistically significant.

		A 44						
		Attribute						
Risk format	All	Probability of	Risk of unnecessary	Lifetime cost of				
		detecting a cancer	follow-up	screening programme				
Mean pupil size per trial over all fixations to the choice set and fixations to each attribute, and								
differences bet	ween risk forma	its						
ΡΟ % [α]	1228.90	1248.89	1186.24	1343.14				
	(92.24)	(94.44)	(90.46)	(99.75)				
IAP 🛊	1184.18	1195.84	1157.78	1264.57				
Icon array	-44.72	-53.05	-28.46	-78.57				
difference [β]	(115.55)	(118.59)	(115.63)	(126.33)				
Change in nun	il size ner trial o	wer all fivations and fi	vations to each attribut	e and differences				
between risk fo	rmats			c, and uniter chees				
PO % [α]	390.33	216.87	207.69	130.30				
	(33.14)	(21.30)	(25.20)	(19.23)				
IAP 🛊	375.93	203.9	210.35	108.41				
Icon array	-14.40	-12.97	2.66	-21.89				
difference [β]	(45.05)	(28.16)	(33.10)	(27.61)				
Observations	440	439	432	376				
Standard arrors in parantheses $* n < 0.05 ** n < 0.01 *** n < 0.001$								

Table 6.8: Measurements (in pixel units) of cognitive pupillometry (hypothesis four)

Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

6.7 Discussion

The aim of this study was to explore eye-tracking as a method to assess DCE respondents' choice strategies. The eye-tracking data generated contributed to the results of this thesis by further exploring the difference risk format can have on the decisions respondents make in a DCE. The study is an original contribution to the literature and marks one of the first investigations into using eye-tracking as a method alongside healthcare DCEs. The next section, 6.7.1, discusses the key findings of the study before considering its strengths and limitations in section 6.7.2.

6.7.1 Key findings

The eye-tracking experiment appeared to be well calibrated with very few fixations to areas of white space. This could also suggest that the respondents were engaged in the task. The equipment was also able to record eye-tracking data for 40 of the 42 women who were invited to the experiment, a success rate of 95%.

Data on the number of fixations indicated formatting risk as either a percentage and icon array or a percentage only had no effect on the visual attention to these attributes. This indicated that there was no difference in information processing, and therefore the ability to make an informed choice. However, the attributes 'risk of unnecessary follow-up' and 'probability of detecting a cancer' both attracted more attention than the cost attribute. This could suggest that the attributes 'risk of unnecessary follow-up' and 'probability of detecting a cancer' required more information processing or that the information was more important for the study participants. When risk was presented as an icon array rather than as a percentage only, the time spent attending to attributes was greater. This could mean participants who received IAP made a more informed choice, as information processing occurred for longer. However, a larger sample size (approximately 50 women to each risk format given a power of 0.8 and 95% confidence level) would be required to investigate the significance of this finding further.

The ANA results presented in this study matched the results found in Balcombe et al., (2014) who state: "*most respondents visually attend most attributes most of the time*" (p.447). Self-reported ANA was reflected in the visual attention to the 'risk of unnecessary follow-up' and 'probability of detecting a cancer' attributes. When a respondent reported that the 'risk of unnecessary follow-up' and 'probability of detecting a cancer' attributes attention to these attributes (as measured by dwell-time).

There was no difference in dwell time between participants who reported attending the cost attribute and those who did not. The lack of a link between self-reported ANA and ANA suggested by the eye-tracking data for the cost attribute could be because respondents could have subconsciously taken account of the cost information in their choice making. Alternatively, the final questions were completed in the presence of a researcher, which could have induced a social desirability bias. The presence of this bias was also supported by the insignificant coefficient on the cost attribute in the modelling of the choice data from the eye-tracking participants, in contrast to the results of the large internet panel in Chapter Four where cost was a significant attribute in all latent preference-classes. The results of this study suggest that self-reported ANA, as measured through supplementary questions to the DCE, may be unreliable for cost attributes.

Results from the saccade data indicated that communicating risk using an icon array significantly increased the proportion of vertical eye-movements made by an individual. These differences in movements indicate that the deliberation process which led to a choice was clearly affected by the risk format. In a simple choice set experiment, with one risk attribute and one cost attribute, Arieli et al. (2011) concluded that a high proportion of vertical eye movements was indicative of an expected pay-off calculation and a high proportion of horizontal eye movements would indicate that the risk and cost were considered separately. Hypothesis three made no *a priori* expectations about the direction of saccades in a multi-attribute choice experiment and there was no existing hypothesis

about the effect icon arrays would have on the movement. The results showed that whilst respondents in both risk versions of the DCEs exhibited more vertical movements than horizontal, the IAP version spent more time considering each alternative separately to make their choice.

More vertical movements could indicate that the respondent was weighing up each alternative as a whole in line with Lancaster's Theory (Lancaster, 1966), and decision-making in line with EUT calculations. The implication of these combined economic theories is complicated (see section 2.5.4.1 of Chapter Two). As this study provided a preliminary investigation into ascertaining whether risk format affected decision strategies, there is only scope to state that a significantly different pattern occurred before a choice was made. However, the nature of the drivers of the difference in saccade patterns can only be speculated and requires further research (for example, changing the orders of the attributes, introducing non-risk attributes).

The results of section 6.5.6, indicated that pupil dilation over the whole trial was smaller for the IAP version of the DCE. This could mean that the task was less cognitively burdensome and easier for respondents to complete. However, the result was insignificant and a larger sample size would be required to investigate this finding further.

6.7.2 Strengths and limitations

This study used sophisticated eye-tracking equipment which recorded many variables which allows for analysis of more variables than conducted in existing published studies (Balcombe et al., 2014; Bialkova & van Trijp, 2011). However, the study had some key limitations.

Within eye-tracking, there is a trade-off between using precise equipment and apparatus that is comfortable for participants. The EyeLink® 1000 required participants to keep their head fixed and involved a forehead and chin-rest set-up. This provided an unnatural reading position and could have reminded the respondents that their eye-movements were being monitored, causing them to pay more attention to the screen than they would have otherwise. However, it is unlikely that the effects of this would differ between respondents randomised to the different risk formats.

The drift correction procedure implemented between trials to check calibration generated precise records of gaze. However, the correction procedure may have induced a starting point bias in the middle of the screen. As all participants in each risk version completed the between trial correction, the effects of risk communication format on eye-tracking outcomes should have been unaffected.

The sample size of twenty women in each condition could ideally have been larger and may possibly led to more statistical significance for some findings. With a larger sample size, covariates could have been added to the regression specifications to investigate how individual effects can influence eye-movement data. However, as described in section 6.5.1, both time and resource constraints meant that a larger sample size was not possible.

When making multiple comparisons between the two risk communication formats using the eye-tracking data, there was a risk that a difference may have occurred in their fixations, saccades or pupil dilation, by chance. The more tests conducted, the bigger the risk of this happening. This means there is an increased risk of a type I error, an incorrect rejection of the null hypothesis that there is no difference between the formats. When comparing risk communication formats, this experiment had only one statistically significant result, difference in the direction of saccades. The Bonferroni correction (Holm, 1979) for multiple comparisons adjusts the significance level to $\frac{\alpha}{n}$, where *n* is the number of tests. In this experiment which involved nine comparisons between the two risk communication formats, the corrected critical p-value of $0.005 \left(\frac{\alpha}{9}\right)$ means this finding is no longer statistically significant. It is important for researchers wishing to use eye-tracking methods in the future to consider the problems associated with multiple comparisons and correct accordingly with techniques such as the Bonferroni correction or more sophisticated approaches.

A conservative limit of 75 milliseconds for a fixation was implemented as a threshold for the analysis. Other studies have used algorithms to calculate the optimal fixation threshold (Salvucci & Goldberg, 2000), but this was not conducted for this study. Similarly, sensitivity analysis using different thresholds could be used to test for robustness in the results.

In addition, randomisation of the order of choice sets or attribute location could have tested the robustness of some of the findings further. The direction of saccades could have been investigated through randomisation of the order of attributes and choice set appearance, and this would allow further investigation of a top-bottom, left-right bias.

As pupil size is also dependent on the location of the stimuli on the screen and luminosity in the lab, the cognitive burden of studies cannot be compared on either an absolute or relative scale using this metric alone. As there are various causes of pupillary responses, care must be taken to distinguish the exact activator of this response. Similarly, the luminance of the lab was not recorded for every participant and additional equipment would have allowed for a more precise measure of pupil dilation. However, the room was windowless and all light sources were consistent between and within experiments, which would suggest a constant level of luminescence in the environment. Given the experiments were conducted under the same settings for all research participants and the two DCE stimuli occurred in exactly the same location on the screen and within the same pixel space, it is valid to make comparisons using pupil measurements in this experiment.

To some extent, participants had control over where they look at and, in theory, could game the situation by looking at particular attributes more or less in the presence of a researcher. However, pupil dilation is a measurement that respondents are unaware of and would find difficult to control. It therefore provided one of the most objective measures of mental intensity available from eye-tracking data (Laeng et al., 2012).

Extraneous variables, such as the time of day and the participants' previous activities were not collected or controlled for in the experiment. Studies have shown that alcohol, a lack of sleep and reading may affect participants' eye-movements. Future work, in a larger study, could collect this information from study participants. However, as respondents were randomised to different risk formats, these are unlikely to have affected the results.

Ideally, concurrent think-aloud analysis would have been conducted alongside the eyetracking study to directly compare the two methods. Using an eye-tracking system which allowed freedom of movement and talking would have resulted in a less accurate account of visual attention. As technology progresses, future research may be able to simultaneously record eye-movements in a think-aloud interview. Alternatively, retrospective think-aloud data can be coded in line with AOI making up the themes to compare participants' self-reported accounts and their attention (Guan et al., 2006).

6.8 Conclusion

This chapter presented an exploratory eye-tracking study which looked at the effect risk communication format had on DCE respondents' choice strategies. The study was innovative in its design, as very little previous research has been conducted using eye-tracking alongside the completion of a DCE. This study found that eye-tracking offered a promising method to understand more about how DCE respondents viewed the choice task. The study also found that visual attention to attributes was greater, and cognitive burden lower, when risk was presented with an icon array; however these differences were not statistically significant (probably due to sample size). Self-reported ANA was only reliable for the non-cost attributes, suggesting respondents may not be truthful about their attention to these attributes. There were statistically significant differences in how participants searched for information in the choice set to make their decision between risk formats.

However, research combining eye-tracking and choice surveys is very much in its infancy, and thus there is potential to answer many further research questions and other risk format effects in future work.

Chapter Seven

Discussion

7.1 Introduction

The thesis examined how the communication of risk attributes in a DCE eliciting women's preference for breast screening affected their choices and decision-making heuristics. Systematic review methods, discrete choice modelling, qualitative research methods and eye-tracking methods were used to investigate how DCEs with risk attributes were completed by respondents and the subsequent values derived. The use of mixed research methods aimed to provide a comprehensive approach to meeting the overarching aim; how to communicate risk in a healthcare DCE.

In section 7.2, the key findings from each chapter are summarised and discussed with regards to their relative contributions. The limitations of the thesis are described in section 7.3, with a discussion of the lessons for future research and researchers using DCEs. In section 7.4, the results are contrasted with relevant existing literature. The key implications and generalisable findings which may be useful for other researchers or policy-makers are presented in section 7.6. Finally, in section 7.7, the chapter reflects on the contribution of this thesis.

7.2 Key results

The first research question concerned the extent to which qualitative research was reported in healthcare DCEs and how useful these research methods were when used alongside DCEs. The systematic review of qualitative research methods alongside healthcare DCEs, and the survey to authors of these studies, identified qualitative research as an underreported yet useful method. Out of the 254 healthcare DCEs identified, only 29 studies reported the details of the qualitative component in-depth. There was no qualitative research reported in 111 of the identified studies. Despite the lack of reporting, the findings of the survey to authors indicated that all authors believed that using qualitative research methods added value to their study. For identifying attributes and levels, the most common approach were interviews or focus groups. For understanding more about thought processes, decision strategies and respondents' understanding of the task, the 'think-aloud' method was identified as most useful. In reviewing the quality of qualitative research conducted alongside DCEs in the 29 studies which reported details, it became apparent that conducting qualitative research alongside stated preference methods was a unique, untraditional, application. As a result, existing appraisal tools for qualitative research

210

methods were found not to be applicable in this setting. A bespoke appraisal tool was therefore developed in order to extract this information consistently and in a systematic, but fairer, manner than existing tools.

The second research question sought to identify the most effective risk communication formats which may aid respondent decision-making in a DCE. To answer this research question, Chapter Three (section 3.2) and Appendix 3.1 presented a structured review of risk communication formats. The review identified icon arrays as the most promising risk communication format which was found to be most effective, compared to other formats, in 10 of the 29 empirical studies which investigated its use as a risk communication tool. However, the structured review also found that the evidence supporting the potential formats was highly heterogeneous. Understanding the relative merits of the different formats, the populations in which they could be most effective, and the best criteria for assessing efficacy is a point for further research. For example, investigations using randomised designs directly comparing formats with appropriate outcomes for appraising the efficacy would be useful. This chapter also identified other challenges in communicating risk information and assessing the effectiveness of the communication format. The structured review of formats generated some incidental findings with regards to the predictors of individuals' ability to interpret risk information, namely: people's perceived risk; their experience of the risk or hazard; and their numeracy skills.

The third research question sought to understand if, and how, the risk communication format of attributes in a DCE affected the valuations derived. In order to investigate this research question, a case study was selected: women's preferences for breast screening. In Chapter Four, a large internet panel was surveyed to elicit female members of the public's preferences. Women were randomised to two risk communication formats: 1) percentages only (the most commonly used format in healthcare DCEs); and 2) icon arrays and percentages. Estimation of a series of random utility choice models indicated that the risk communication format did not affect either women's marginal utility or choice consistency. However, women's preferences were highly heterogeneous (some women most sensitive to cost; others to the probability of detecting a cancer; and others to the risk of unnecessary follow-up) with six latent preference-classes identified. Significant predictors of class membership were women's concern about their own risk of developing breast cancer and their ethnicities. The finding that the valuations of breast screening attributes derived from this DCE were insensitive to the risk communication format was surprising given the findings of the structured review presented Appendix 3.1. However, a

key implication was the importance of understanding DCE respondents' attitude towards the risk associated with the good, service or intervention being valued.

Chapter Five utilised qualitative research methods as an alternative approach to traditional quantitative choice modelling, to investigate the sensitivity of DCE respondents' choices and decision strategies to the risk communication method in the DCE. Qualitative research methods have not previously been used to investigate attribute format effects in DCEs. The think-aloud interviews revealed that when risk was communicated as a percentage only women often visualised the risk, creating their own icon array and relieving part of the task's burden. There were often sentiments of initial 'panic' when the percentages only version was completed by the respondents, and the icon array was much better received.

The fourth research question aimed to understand how effective eye-tracking methods were in evaluating respondents' understanding of DCE attributes and levels. Chapter Six used eye-tracking methods to understand more about individuals' decision strategies and whether risk communication format affected these. The eye-tracking experiment possessed characteristics of validity, with successful calibration and little attention to areas of no information. The eye-tracking results revealed significant differences between the DCE versions, with women receiving icon arrays making significantly more up-down eyemovements when they completed the choice task. The consequence of this finding on the underlying economic theories, which support the use of DCEs as a valuation method, requires further research.

Results for other outcomes suggested that women receiving icon arrays also paid more visual attention to attributes (indicating more information processing) and found the DCE less cognitively burdensome. However, differences between risk formats were not statistically significant for these outcomes. Interestingly, although women who self-reported non-attendance to the risk and detection attributes visually attended these attributes significantly less than those who reported attending them, this result was not found for the cost attribute The mismatch of actual and self-reported attribute non-attendance to the cost attribute (and not in other attributes) could indicate some bias due to researcher presence, such as social desirability. In this respect, eye-tracking data could prove more reliable than self-reported behaviour.

Each of the three empirical studies contributed to the overall aim which was to understand if and how DCE respondents' choices and decision strategies were affected by the risk communication format. In the case of understanding whether choice was affected, the insignificant parameter on the scale term in the heteroscedastic conditional logit and the finding that communication format did not predict scale or preference class membership suggests that the addition of icon arrays made no difference. Alternative research methods (qualitative and eye-tracking) confirmed some of the findings of the quantitative study but also suggested that icon arrays relieved some of the cognitive burden in the choice task.

7.3 Limitations

In this section, the limitations of each study conducted in this thesis will be discussed. Possible remedies to the issues will also be discussed where possible.

The systematic review of qualitative research conducted alongside DCEs which was presented in Chapter Five (section 5.2) and Appendix 5.1 used standardised systematic review methods and published guidelines. Two potential limitations were identified: 1) the reliance on material published in the articles; and 2) the lack of an appropriate appraisal tool. These limitations were rectified with a survey to authors and the creation of a bespoke appraisal tool. However, the survey to authors only involved contacting authors who had reported basic qualitative research in their studies and did not include those who did not report any. Authors who did not report qualitative research in their study frequently mentioned the absence of qualitative methods as a limitation, suggesting they felt it would have added value. Authors who reported extensive details about the qualitative aspect were assumed to have presented the research as it was an important component of their study. In hindsight, it may have been useful to extend the survey to all authors of the review (particularly to understand the views of authors who reported no details). As a consequence, the survey results may be biased and authors who did not report qualitative details (and hence were not contacted) might have done so because this element was not important or useful.

Another limitation of the systematic review presented in section 5.2 and Appendix 5.1 was that the data extraction was also only conducted by one author (CV). In retrospect, it would have been useful to have a second reviewer extracting data. However, the recent publication of a systematic review employing the same strategy (Clark et al., 2014) in part validates the quality of the systematic review presented in Chapter Three, which identified more DCE studies.

In Chapter Three (section 3.2), the review used a specific search strategy to rapidly identify a range of risk communication formats. The advantage of using this approach was that the review was conducted promptly. However, this meant that non-health risk communication was not covered. For example, future research could explore environmental risk communication or literature in health and safety, or engineering. The analysis of the extracted data was also limited. Counting the best performing risk communication formats resulted in more frequently used formats appearing more popular. In hindsight, a more sophisticated meta-analysis of each empirical study could have been conducted with a thorough ordering of the formats rather than a best-worst rating. It is not clear if, or how, using this more sophisticated approach would have changed the results of the review.

In the large sample DCE study described in Chapter Four, an internet panel provider was chosen to yield responses. Internet panels were chosen to acquire a large sample size, paramount for the investigation of heterogeneity, quickly and relatively inexpensively. However, there are limitations to this sampling approach. The respondents to internet panel DCEs are likely to be computer-literate and therefore could be biased against the older-age ranges. Furthermore, the respondents were paid to complete the survey which could have effected their engagement in the task and causing them to provide 'answers' to things they would not chose in real life. There is evidence to suggest internet panels condition respondents whose preference and behaviour may change due to participation in the panel (Dennis, 2001). The advantages and disadvantages of using internet panels for DCEs have yet to be explored; however, other health valuation studies have found they provide good quality data compared to other methods such as postal surveys, telephone interviews (Mulhern et al., 2015). It is, however, acknowledged that the views found in this study may not necessarily be representative of the general public's which may consequently hinder the interpretation of the quantified benefit-risk trade-offs and even the existence of the distinct latent preference-classes identified.

Although other discrete choice models could have been fitted (such as the MXL or nested logit), the study used sophisticated econometric techniques beyond that which are typically used in healthcare DCEs (Clark et al., 2014). It was believed that describing heterogeneity in terms of latent class models would be most useful for decision or policy-makers. In addition to providing details about preference heterogeneity across groups, rather than across all individuals, latent class models have also outperformed mixed logit models in most tests of statistical properties (Shen, 2009).

The qualitative interviews presented in Chapter Five used a verbal-protocol method which was particularly susceptible to three key issues: reactivity, nonveridicality and reflexivity. The study presented in Chapter Five recognised the impact of the interviewer's presence on the interviewees' choices, and accounts of making choices, and hence the analysis and interpretation of the qualitative data collected. Efforts were made to minimise the researcher bias with the flexible interview schedule which largely comprised prompts and the deductive analytical approach of open-coding. The framework to hone themes into the

categories of 'risk' and 'strategy' were to focus the analysis but did not influence the emergence of new themes. However, it is impossible to say that, even an unconscious, awareness of the background research question did not influence the reading of transcripts.

Women who participated in the qualitative interviews made contact in response to an advertisement (email, website, or noticeboard). It could therefore be that the sample in this study was biased toward middle-class and those who were intrinsically motivated. This problem is not unique to this study; the challenges of recruiting diverse research participants are well established, particularly for studies with a health focus (Yancey et al., 2006; Shavers-Hornaday et al., 1997).

The eye-tracking study presented in Chapter Six suffered from a small sample size chosen based on previous studies (Bialkova & van Trijp, 2011), of which there were very few. Retrospective sample size calculations suggest that approximately 100 women would be required to identify statistically significant differences in outcomes between the risk communication formats. The sample size in this study was also restricted by time (due to the shared laboratory) and resource constraints. The recruitment of eye-tracking participants was particularly challenging as women had to visit the university's campus during core weekday hours. This was unavoidable due to the laboratory set-up and sharing arrangements. As a consequence, women who participated in this study may not be representative on the general public. This is acknowledged as a limitation of this study. More time and resources could have resulted in an increase in the sample size and therefore an increase in the meaningfulness of the results through statistical significance.

The device used the in the eye-tracking study required respondents to rest their chin and forehead for stability. This unnatural set-up may have interfered with the respondents' behaviour or choices. However, less intrusive devices, recording with a lower frequency, would have compromised the accuracy of the eye-tracking data. Furthermore, many remote devices do not record data on pupil size, meaning an examination of the effect of risk communication format on cognitive burden could not have been conducted.

Comparisons between the results of the three empirical studies of this thesis was challenging due to the differences in the epistemological and ontological assumptions under which the research was conducted. Similarly, comparing conflicting findings requires an opinion of which method was most valid. In health services research, there is a lack of consensus even with regards to the synthesis of mixed-methods findings (Brannen, 2006). As a consequence, the thesis is restricted in its ability to compare findings across the empirical chapters.

All empirical studies employed the same DCE containing just three attributes. In a more complex experiment, the effect of icon arrays on respondents' marginal utilities and decision-making heuristics may have been different. Other attributes, such as process attributes, could have been included; however, these three attributes were chosen based on literature reviews and discussions with experts in breast cancer. Including more attributes may have increased the cognitive burden for respondents and detracted from the risk attributes. A simple DCE also resulted in large areas of interest which enabled a clear analysis of the eye-tracking data. Similarly, choosing a familiar example was a key criterion for selecting a case study in Chapter Three in order to limit any other cognitive burden to the respondents.

The use of round numbers for the risk attribute (1%; 5%; 10%; and 20%) are arguably easy for women to visualise even when presented only in percentage form. Therefore a greater test of the benefits of presenting an additional icon array may have been investigated through smaller percentages (<1%) or presenting them to the nearest one thousandth (for example, X.X%). Furthermore, the empirical studies in this thesis only explored two risk communication formats out of the numerous pictorial, graphical, numerical and qualitative approaches identified in the structured review of section 3.2, Chapter Three.

It is acknowledged that the event or hazard described can affect the interpretation of its likelihood. It has been found that a 1 in 100 risk of death feels more likely than a 1 in 100 chance of a cold even if, statistically, they are the same (Sunstein, 2003). This is known as 'probability neglect' and highlights the importance of the event on people's interpretation of its likelihood (Zeckhauser & Sunstein, 2009). With a different event and even the same probabilities, the results of this thesis could be very different. For example, if the risk of unnecessary follow-up was changed to risk of mortality (keeping the probabilities the same), the effect of the risk communication format may have been more prominent.

Stated preference methods are limited in that they rely on respondents saying what they would do in a hypothetical situation. However, in a publically-funded healthcare system there are rarely market situations from which revealed preference data can be generated. Therefore, a robust and established stated preference method was chosen using a choice-based format. Novel, less established, methods may have induced other limitations. The hypothetical nature of stated preference valuation methods and new alternative choice experiment designs, such as BWS, are recognised as a limitation of this thesis.

Furthermore, all stated preference methods involve the use of surveys. The empirical studies of this thesis relied on a survey based method which therefore implicitly resulted in
a literate sample. The different mode of administration (web-based, face-to-face and eyetracking) and the different choice results (for example cost becoming insignificant in a face-to-face setting) suggest that surveys are affected by their administration. This is in line with other survey-related literature (see Hyman (1944); Preisendörfer & Wolter (2014)). As a survey design is inherent to stated-preference studies, and revealed preference data was not available, the issues arising from a survey-based approach in this context are unavoidable.

7.4 Comparisons with existing literature

There was little existing research investigating the communication of risk attributes in DCEs. The results of other studies that have only explored quantitatively the effects of risk communication on choices (Howard & Salkeld, 2009; Buchanan et al., 2014), had major limitations in that they failed to account for differences in the scale parameter when comparing the risk communication formats. The analysis conducted in Chapter Four extensively accounts for differences in scale, and explores preference heterogeneity, another area also under-investigated in existing literature (Clark et al., 2014).

The risk communication literature reviewed in Chapter Three (section 3.2) was supportive of the use of icon arrays as a risk communication format. The results of this study found that icon arrays were well received by DCE participants, even if they did not directly affect the valuations. Other components of important aspects of risk communication identified in the literature were important. The quantitative study, in particular, highlighted the importance of women's perception (or concern) on their risk preferences.

Literature on women's preferences for screening has suggested that women are tolerant of the risk of unnecessary follow-up, even at rates of 30% (Baum, 2013). The quantification of preferences in this thesis showed that although some women were very tolerant, others were not. The heterogeneity in preferences match other literature which suggests large variation in people's preferences for healthcare (Hole, 2008).

7.5 Key implications

The culmination of the findings from each of the studies in this thesis provides important implications for researchers and policy-makers. These implications are now described.

For researchers seeking to use DCEs for eliciting individuals' preferences for benefit-risk trade-offs, the method was not sensitive to the risk communication formats used in this study. The qualitative study identified that respondents felt more engaged with the task when risk was presented with an additional icon array. As a result, this PhD suggests that

icon arrays in addition to percentages are preferred to percentages alone when presenting risk attributes in a DCE.

Another key implication for researchers was the usefulness of qualitative research methods alongside DCEs. The survey to authors found all researchers' experience of the methods added value to their DCE. Given the lack of qualitative research conducted alongside DCES, previous calls for qualitative research in health economics (Coast, 1999) and alongside DCEs (Coast et al., 2012) appear to have remain unanswered. The bespoke appraisal tool developed as part of this thesis and presented in Appendix 5.5 may be a useful checklist for health economists seeking to conduct qualitative research alongside their stated preference study. The appraisal tool may also be incorporated into existing guidelines, such as those published by ISPOR, or utilised by researchers in the form of standardised reporting criteria.

In this thesis, eye-tracking was successfully used alongside a DCE and was found to be a valid and insightful method for examining respondents' decision-making heuristics. A recommendation of this thesis is that eye-tracking should be used in future research to understand more about choice-making behaviour and attention to attributes. Researchers seeking to use eye-tracking methods alongside DCEs can do so, but should be aware of the large sample sizes required to achieve significance and the relative advantages and disadvantages of different types of eye-tracking devices.

Women's preferences for an NHSBSP were highly heterogeneous. For researchers, models which allow for such preference heterogeneity need to be used when modelling DCE choice data. Particularly researchers in health, where applications of these methods are limited (Clark et al., 2014). Additionally, the scale parameter must also be considered if there are reasons to believe respondent characteristics might influence choice consistency. Failure to account for differences in scale may prevent accurate investigation of preference heterogeneity. Therefore researchers modelling healthcare choice data should consider greater use of scale-adjusted latent-class and heteroscedastic conditional logit models.

There are also implications for policy-makers wanting to implement an effective NHSBSP. Policy-makers must consider that there is a substantial group of women who were less tolerant of the risks of screening, and recognise that ethnic minority groups place little value on the benefit of detecting a cancer. If policy-makers wish to improve up-take in minority groups, then the benefits of screening need to be better communicated or incentives should be in place to encourage programme participation. However, 'informed choice' is a rising consideration for policy-makers providing healthcare services (Forbes & Ramirez, 2014); and increased participation may not be an important objective.

Policy-makers should also realise that some women received a very high utility from just participating in a NHSBSP, regardless of the benefits, risks and costs. This finding suggests women placed a high value on non-health benefits. Decision-makers may wish to consider alternative valuation methods to extend or complement the results of a traditional QALY based cost-utility or cost-effectiveness analysis, to incorporate alternative sources of utility. As NICE currently makes decisions using an extra-welfarist view-point and the results of CUA and/or CEA, the NHSBSP may be undervalued under this framework.

Recent changes to breast screening policy have included an extension of the screening ages by six years, inviting women between 47 to 73 years old in some areas of England (NIHR Trial ISRCTN33292440). Policy-makers should be aware of women's preferences for the increased risk of unnecessary follow-up when screening younger and older women. For women between 50 and 70 years, how preferences change through the programme (as the probability of detecting a cancer reduces and the risk of unnecessary follow-up rises) also needs to be considered by policy-makers (Barratt, 2015).

7.6 Future research

The studies presented in this thesis can be taken forward in several ways.

The critical appraisal tool developed for assessing the quality of the qualitative research reported alongside published DCEs should be trialled with different stated preference studies such as BWS and with more contingent valuation studies. Successful application of this tool in more research areas would provide scope to transform the tool into best practice or reporting guidelines.

How the analysis of DCE choices can incorporate other expected utility theories is still under development and requires further research. EUT can be combined with random utility theory which models the uncertainty in conclusions drawn about the unobservable error component of utility (Polak & Liu, 2006). Exploration of this in a healthcare setting has identified this as an important consideration (Robinson et al., 2015). In addition, alternative theories for analysing choice data, such as regret minimisation, may be appropriate (Thiene et al., 2012).

The effect of different risk frames for attributes in a DCE (for example, survival versus mortality) has been investigated previously (Howard & Salkeld, 2009). Given the existence of research already investigating framing-effects, this thesis concentrated on the

effect of risk communication formats specifically. Future research could consider the effect of risk communication format and risk frame, for example randomising respondents within each format to receive the 'probability of receiving a true positive' (in contrast to attribute 'probability of unnecessary follow-up').

Not only does the risk attribute need to be appropriately communicated, but for WTP calculations, cost should be communicated appropriately too. The think-aloud method revealed other strategic behaviour (such as using others' budget constraints) and the presence of non-use values in women's decision-making indicating that current communication methods may be improved. Further investigation of the presence of these biases and the appropriate method of communicating cost in a healthcare setting is warranted. This could be conducted via think-aloud methods or other qualitative approaches to ascertain the presence of such phenomena.

The use of eye-tracking as a method alongside DCEs was very exploratory and there is scope for further research in this area. Future studies should consider using a larger sample size to investigate visual attention to information in a DCE choice set. Other research could extend the use of pupil dilation as a proxy for task complexity in a DCE setting by comparing pupil size in choice-sets containing alternatives of similar utility (a difficult choice) where correct communication is likely to be most important. Further research may involve the incorporation of visual attention data into the quantitative modelling of choices, or a comparison of eye-tracking attention to qualitative accounts of behaviour collected in retrospective think-aloud interviews.

7.7 Conclusion

This thesis presents a contribution to multiple research areas including: the applied DCE literature; eye-tracking and psychology research; and literature utilising qualitative research methods alongside stated preference methods. Whilst few definitive answers can be given, an insight into the usefulness of different research methods and the effect of risk communication format on choices and choice-making in a DCE context has been extensively tested. It appears that DCEs can be used to value benefit-risk trade-offs. These benefit-risk trade-offs were not sensitive to the risk communication format of the attributes in the selected case study. However, providing an icon array is highly recommended with the results of the qualitative study suggesting this is beneficial for DCE respondents. In addition, the results and implications of this thesis are likely to be of interest to both researchers and policy-makers.

- Abdullah, A., Peeters, A., de Courten, M. and Stoelwinder, J. (2010). The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. *Diabetes research and clinical practice*. 89 (3). 309–19.
- Abellan-Perpinan, J.-M., Martinez-Perez, J.-E. and Pinto Prades, J.L. (2005). Biased risk perception in health state utility measurement: a new type of framing effect. *Working Paper: Universidad de Murcia, Departamento de Economía Aplicada.* 1–22.
- Adamowicz, W., Bunch, D. and Cameron, T. (2008). Behavioral frontiers in choice modeling. *Marketing Letters*. 19. 215–228.
- Alcaly, R. and Klevorick, A. (1970). Judging quality by price, snob appeal, and the new consumer theory. *Zeitschrift für Nationalökonomie*. 30 (1-2). 53–64.
- ALLHAT (2002). Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs usual care: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA*. 288 (23).
- Anderson, W.F., Jatoi, I., Tse, J. and Rosenberg, P.S. (2010). Male breast cancer: A population-based comparison with female breast cancer. *Journal of Clinical Oncology*. 28 (2). 232–239.
- Appleby, J. (2013). Spending on health and social care over the next 50 years: why think long term? *The King's Fund.*
- Arieli, A., Ben-Ami, Y. and Rubinstein, A. (2009). Fairness Motivations and Procedures of Choice between Lotteries as Revealed through Eye Movements. *Eitan Berglas School of Economics*. 1– 23.
- Arieli, A., Ben-Ami, Y. and Rubinstein, A. (2011). Tracking decision makers under uncertainty. *American Economic Journal: Microeconomics*. 3 (4). 68–76.
- Athanasakis, K., Souliotis, K., Tountas, Y., Kyriopoulos, J. and Hatzakis, A. (2012). A cost-utility analysis of hypertension treatment in Greece: assessing the impact of age, sex and smoking status, on outcomes. *Journal of hypertension*. 30 (1). 227–34.
- Augustovski, F., Beratarrechea, A., Irazola, V., Rubinstein, F., Tesolin, P., Gonzalez, J., Lencina, V., Scolnik, M., Waimann, C., Navarta, D., Citera, G. and Soriano, E.R. (2013). Patient preferences for biologic agents in rheumatoid arthritis: a discrete-choice experiment. *Value in Health*. 16 (2). 385–93.
- Austin, M., Hutter, C.M., Zimmern, R.L. and Humphries, S.E. (2004). Familial hypercholesterolemia and coronary heart disease: a HuGE association review. *American journal of epidemiology*. 160 (5). 421–9.
- Avineri, E. and Prashker, J. (2004). Violations of expected utility theory in route-choice stated preferences: Certainty effect and inflation of small probabilities. *Transportation Research Record*. 1894. 222–229.
- Baker, R., Freeman, G., Boulton, M., Brookes, O., Windridge, K., Tarrant, C., Low, J., Turner, D., Hutton, E. and Bryan, S. (2006). Continuity of Care: patients' and carers' views and choices in their use of primary care services. *National Co-ordinating Centre for NHS Service Delivery and Organisation Research and Development*. SDO/13b/20.
- Baker, R. and Robinson, A. (2004). Responses to standard gambles: are preferences 'well constructed'? *Health Economics*. 13 (1). 37–48.
- Baker, R., Robinson, A. and Smith, R. (2008). How do respondents explain WTP responses? A review of the qualitative evidence. *Journal of Socio-Economics*. 37 (4). 1427–1442.

- Balcombe, K., Fraser, I. and McSorley, E. (2014). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*. 30 (3). 1–27.
- Bansal, N., Bhopal, R.S., Steiner, M.F.C. and Brewster, D.H. (2012). Major ethnic group differences in breast cancer screening uptake in Scotland are not extinguished by adjustment for indices of geographical residence, area deprivation, long-term illness and education. *British journal of cancer*. 106 (8). 1361–6.
- Barratt, A. (2015). Overdiagnosis in mammography screening: a 45 year journey from shadowy idea to acknowledged reality. *BMJ British Medical Journal*. 350.
- Baum, M. (2013). Harms from breast cancer screening outweigh benefits if death caused by treatment is included. *BMJ British Medical Journal*. 346. f385.
- Beatty, J. and Wagoner, B. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*. 199 (4334). 1216–1218.
- Bech, M. (2003). Politicians' and hospital managers' trade-offs in the choice of reimbursement scheme: a discrete choice experiment. *Health Policy*. 66 (3). 261–275.
- Bech, M. and Gyrd-Hansen, D. (2005). Effects coding in discrete choice experiments. *Health economics*. 14 (10). 1079–83.
- Bederman, S.S., Mahomed, N.N., Kreder, H.J.H., McIsaac, W.J., Coyte, P.C. and Wright, J.G. (2010). In the eye of the beholder: preferences of patients, family physicians, and surgeons for lumbar spinal surgery. *Spine*. 35 (1). 108–115.
- Van Beers, R.J. (2007). The sources of variability in saccadic eye movements. *The Journal of neuroscience*. 27 (33). 8757–70.
- De Bekker-Grob, E.W., Hol, L., Donkers, B., Van Dam, L., Habbema, J.D.F., Van Leerdam, M.E., Kuipers, E.J., Essink-Bot, M.L. and Steyerberg, E.W. (2010). Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value in Health*. 13 (2). 315–323.
- De Bekker-Grob, E.W., Ryan, M. and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics*. 21 (2). 145–172.
- Bell, T. and Branston, L. (1999). Interventions to improve uptake of breast screening in inner city Cardiff general practices with ethnic minority lists. *Ethnicity and Health*. 4 (4). 277–284.
- Berg, B. (2007). Qualitative Research Methods for the Social Sciences. 7th Ed. Pearson Education.
- Berry, D., Cronin, K., Plevritis, S.K., Fryback, D.G., Clarke, L., Zelen, M., Mandelblatt, J.S., Yakovlev, A.Y., Habbema, J.D.F. and Feuer, E.J. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *The New England journal of medicine*. 353 (17). 1784–1792.
- Bialkova, S. and van Trijp, H.C.M. (2011). An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Quality and Preference*. 22 (6). 592– 601.
- Binda, P., Pereverzeva, M. and Murray, S.O. (2014). Pupil size reflects the focus of feature-based attention. *Journal of neurophysiology*. 112 (12). 3046–52.
- Blakemore, F. and Williams, A. (2008). British tourists' valuation of a Turkish beach using contingent valuation and travel cost methods. *Journal of Coastal Research*. 246 (246). 1469–1480.
- Blamey, R., Bennett, J. and Morrison, M. (1999). Yea-saying in contingent valuation surveys. Land Economics. 75 (1). 126–141.

- Bleichrodt, H. and Johannesson, M. (1997). Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of health economics*. 16 (2). 155–75.
- Bloom, G., Standing, H. and Lloyd, R. (2008). Markets, information asymmetry and health care: towards new social contracts. *Social science & medicine*. 66 (10). 2076–87.
- BNF (2015). Adverse reactions to drugs. British National Formulary. 62.
- Boden-Albala, B. and Sacco, R. (2000). Lifestyle factors and stroke risk: exercise, alcohol, diet, obesity, smoking, drug use, and stress. *Current atherosclerosis reports*. 2 (2). 160–166.
- Bohme, M., Meyer, A., Martinetz, T. and Barth, E. (2006). Remote eye tracking: State of the art and directions for future development. *The 2nd Conference on Communication by Gaze Interaction*. 2–6.
- Boote, J., Baird, W. and Beecroft, C. (2010). Public involvement at the design stage of primary health research: a narrative review of case examples. *Health policy*. 95 (1). 10–23.
- Boren, T. and Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions* on *Professional Communication*. 43 (3). 261–278.
- Boxall, P.C., Adamowicz, W., Swait, J., Williams, M. and Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological economics*. 18 (3). 243–253.
- Boyle, K.J., Holmes, T.P., Teisl, M.F. and Roe, B. (2001). A Comparison of Conjoint Analysis Response Formats. *American Journal of Agricultural Economics*. 83 (2). 441–454.
- Brannen, J. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*. 6 (1). 97 113.
- Branthwaite, A. and Patterson, S. (2011). The power of qualitative research in the era of social media. *Qualitative Market Research: An International Journal*. 14 (4). 430–440.
- Brazier, J., Deverill, M., Green, C., Harper, R. and Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment*. 3 (9). 1–176.
- Brazier, J., Rowen, D., Yang, Y. and Tsuchiya, A. (2012). Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European journal of health economics*. 13 (5). 575–87.
- Breast Cancer UK, . (2015). Breast screening statistics.
- BreastHealthUK (2014). Private breast screening and diagnostic service prices. *www.breasthealthuk.com*.
- Bresnahan, B., Dickie, M. and Gerking, S. (1997). Averting behavior and urban air pollution. *Land Economics*. 73 (3). 340–357.
- Brewer, N.T., Richman, A.R., DeFrank, J.T., Reyna, V.F. and Carey, L. (2012). Improving communication of breast cancer recurrence risk. *Breast Cancer Research and Treatment*. 133 (2). 553–61.
- Bridges, J.F., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L., Regier, D. a, Johnson, F. and Mauskopf, J. (2011a). Conjoint analysis applications in health-a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health*. 14 (4). 403–413.

- Bridges, J.F., Kinter, E.T., Schmeding, A., Rudolph, I. and Muhlbacher, A. (2011b). Can patients diagnosed with schizophrenia complete choice-based conjoint analysis tasks?. *The patient*. 4 (4). 267–275.
- Brouwer, W., Culyer, A., van Exel, N. and Rutten, F. (2008). Welfarism vs. extra-welfarism. *Journal of health economics*. 27 (2). 325 338.
- Bryan, S. and Dolan, P. (2004). Discrete choice experiments in health economics. For better or for worse? *The European journal of health economics*. 5 (3). 199–202.
- Bryan, S., Williams, I. and Mciver, S. (2007). Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. *Health Economics*. 16 (2). 179–193.
- Buchanan, J., Wordsworth, S. and Schuh, A. (2014). Presenting risk information in discrete choice experiments : do graphical techniques influence willingness-to-pay ? *Health Economists' Study Group Meeting*.
- Bundey, S. and Crews, S.J. (1984). A study of retinitis pigmentosa in the City of Birmingham. II Clinical and genetic heterogeneity. *Journal of medical genetics*. 21 (6). 421–8.
- Burgess, R. (1984). In the field: An introduction to field research. London: Routledge.
- Burns, N. (1989). Standards for Qualitative Research. Nursing Science Quarterly. 2 (1). 44-52.
- Campbell, D., Hutchinson, W.G. and Scarpa, R. (2006). Lexicographic preferences in discrete choice experiments: consequences on individual-specific willingness to pay estimates. *Fondazione Eni Enrico Mattei*. (I).
- Cancer Research UK (2010). Breast cancer UK incidence statistics.
- Cancer Research UK (2013). NHS breast screening: Helping you decide. *NHS Cancer Screening Programmes, Public Health England, Department of Health Publications.* 1–3.
- Carlsson, F., Kataria, M. and Lampi, E. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental and Resource Economics*. 47 (1). 65–89.
- Carson, R.T. (1994). Contingent Valuation Surveys and Tests of Insensitivity to Scope. Netherlands: Springer.
- Carson, R.T. and Louviere, J. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*. 49 (4). 539–559.
- Carson, R.T., Louviere, J. and Wasi, N. (2009). A cautionary note on designing discrete choice experiments: a comment on Lusk and Norwood's 'effect of experiment design on choice-based conjoint valuation estimates'. *American Journal of Agricultural Economics*. 91 (4). 1056–1063.
- Carter, J.L., Coletti, R.J. and Harris, R.P. (2015). Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ British Medical Journal*. 350. 1–18.
- CASP (2006). 10 questions to help you make sense of qualitative research (Critical Appraisal Skills Programme). *Public Health Resource Unit.*
- Charles, C., Gafni, A. and Whelan, T. (1997). Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Social science & medicine*. 44 (5). 681–692.

- Chen, W., Hong, H., Liu, Y., Zhang, L., Hou, X. and Raymond, M. (2004). Recreation demand and economic value: An application of travel cost method for Xiamen Island. *China Economic Review*. 15 (4). 398–406.
- Cheraghi-Sohi, S., Bower, P., Mead, N., McDonald, R., Whalley, D. and Roland, M. (2007). Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique. *Family Practice*. 24 (3). 276–282.
- Cheyne, B.H., Terry, R., Niven, C., Dowding, D., Hundley, V. and Mcnamee, P. (2007). 'Should I come in now ?': a study of women's early labour experiences. *British Journal of Midwidery*. 15 (10). 604–609.
- Chilton, S.M. and Hutchinson, W.G. (2003). A qualitative examination of how respondents in a contingent valuation study rationalise their WTP responses to an increase in the quantity of the environmental good. *Journal of Economic Psychology*. 24 (1). 65–75.
- Choice Metrics (2012). Ngene User Manual 1.1.1.
- Cicchetti, C.J. and Wilde, L.L. (1992). Uniqueness, irreversibility, and the theory of nonuse values theory. *American Journal of Agricultural Economics*. 74 (5). 1121–1125.
- Clark, M., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E.W. (2014). Discrete choice experiments in health economics: a review of the literature. *PharmacoEconomics*. 32 (9). 883–902.
- Clarke, P. (2002). Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care. *Health Economics*. 11 (2). 117–127.
- Clarke, P., Wolstenholme, J. and Wordsworth, S. (2010). *Applied methods of cost-effectiveness analysis in healthcare*. Vol 3: OUP Oxford.
- Cleeland, C. and Ryan, K. (1992). Pain assessment: global use of the Brief Pain Inventory. *Annals of the Academy of Medicine*. 23 (2). 129–138.
- Coast, J. (1999). The appropriate uses of qualitative methods in health economics. *Health economics*. 8 (4). 345–353.
- Coast, J., Al-Janabi, H., Sutton, E., Horrocks, S.A., Vosper, J., Swancutt, D.R. and Flynn, T. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics*. 21 (6). 730–741.
- Coast, J. and Horrocks, S.A. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of health services research & policy*. 12 (1). 25–30.
- Coast, J., McDonald, R. and Baker, R. (2004). Issues arising from the use of qualitative methods in health economics. *Journal of health services research & policy*. 9 (3). 171–6.
- Committee on Safety of Medicines (1995). Combined oral contraceptives and thromboembolism. *Lancet*.
- Conrad, F.G. and Blair, J. (2009). Sources of Error in Cognitive Interviews. *Public Opinion Quarterly*. 73 (1). 32–55.
- Cookson, R. (2003). Willingness to pay methods in health care: a sceptical view. *Health economics*. 12 (11). 891–4.
- Corso, P., Hammitt, J.K. and Graham, J.D. (2001). Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. *The Journal of Risk and Uncertainty*. 23 (2). 165–184.

- CRD (2008). Systematic Reviews. CRD's (Centre for Reviews and Dissemination) guidance for undertaking reviews in health care. The University of York.
- Criswell, L., Saag, K.G., Mikuls, T.R., Cerhan, J.R., Merlino, L., Lum, R.F., Pfeiffer, K., Woehl, B. and Seldin, M.F. (2006). Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Annals of the rheumatic diseases*. 65 (9). 1163–7.
- Cunningham, C.E., Deal, K. and Chen, Y. (2010). Adaptive choice-based conjoint analysis of health service preferences. *The Patient*. 3 (4). 257–273.
- Czajkowski, M., Barczak, A., Budziński, W., Giergiczny, M. and Hanley, N. (2014). Within- and between- sample tests of preference stability and willingness to pay for forest management. *University of St Andrews Discussion Papers in Environmental Economics*. (6). 1–24.
- Czech, B., Krausman, P. and Borkhataria, R. (1998). Social construction, political power, and the allocation of benefits to endangered species. *Conservation Biology*. 12 (5). 1103–1112.
- Dancet, E., Van Empel, I., Rober, P., Nelen, W., Kremer, J. and D'Hooghe, T. (2011). Patient-centred infertility care: a qualitative study to listen to the patient's voice. *Human Reproduction*. 26 (4). 827–33.
- Debreu, G. (1960). Topological methods in cardinal utility theory. *Mathematical Methods in the Social Sciences*. 16–26.
- Denes-Raj, V. and Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of personality and social psychology*. 66 (5). 819– 829.
- Dennis, J. (2001). Are internet panels creating professional respondents? *Marketing Research*. 13 (2). 34–38.
- Denzin, N.K. (2009). The elephant in the living room: or extending the conversation about the politics of evidence. *Qualitative Research*. 9 (2). 139–160.
- Denzin, N.K. and Lincoln, Y.S. (2005). The Sage handbook of qualitative research. Sage Publications.
- Devlin, N., Tsuchiya, A., Buckingham, K. and Tilling, C. (2011). A uniform Time Trade Off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Economics*. 20 (3). 348–361.
- Diener, A., O'Brien, B. and Gafni, A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health economics*. 7 (4). 313–26.
- Dixon-Woods, M. (2004). The problem of appraising qualitative research. *Quality and Safety in Health Care*. 13 (3). 223–225.
- Dolan, J.G. and Iadarola, S. (2008). Risk communication formats for low probability events: an exploratory study of patient preferences. *BMC medical informatics and decision making*. 8 (1). 14.
- Dolan, P., Torgerson, D.J. and Wolstenholme, J. (1999). Costs of breast cancer treatment in the United Kingdom. *The Breast*. 8 (4). 205–7.
- Drummond, M. and McGuire, A. (2001). Economic evaluation in health care. Oxford University Press.
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. 2nd Ed. Springer Science & Business Media.
- Duchowski, A.T. (2002). A breadth-first survey of eye-tracking applications. Behavior research methods, instruments, & computers. 34 (4). 455–470.

- Dziak, J., Coffman, D., Lanza, S. and Li, R. (2012). Sensitivity and specificity of information criteria. *The Methodology Center Technical Report Series Pennsylvania State University*.
- Easterbrook, P.J., Berlin, J. a., Gopalan, R. and Matthews, D.R. (1991). Publication bias in clinical research. *Lancet*. 337 (8746). 867–876.
- Edwards, A., Elwyn, G., Covey, J., Matthews, E. and Pill, R. (2001). Presenting risk information-a review of the effects of 'framing' and other manipulations on patient outcomes. *Journal of health communication*. 6 (1). 61–82.
- Edwards, A., Elwyn, G. and Mulley, A. (2002). Explaining risks: turning numerical data into meaningful pictures. *BMJ British Medical Journal*. 324. 827–830.
- Edwards, N. and Jones, D. (2000). Uptake of breast cancer screening in older women. *Age and ageing*. 29 (2). 131–135.
- Ericson, K. and Simon, H. (1993). Protocol analysis: Verbal reports as data (revised edition).
- Evans, G., Barwell, J., Eccles, D.M., Collins, A., Izatt, L., Jacobs, C., Donaldson, A., Brady, A.F., Cuthbert, A., Harrison, R., Thomas, S., Howell, a, Miedzybrodzka, Z. and Murray, A. (2014). The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. *Breast Cancer Research*. 16 (5). 442.
- Evans, G., Blair, V., Greenhalgh, R., Hopwood, P. and Howell, a (1994). The impact of genetic counselling on risk perception in women with a family history of breast cancer. *British journal of cancer*. 70 (5). 934–8.
- Evans, G., Burnell, L.D., Hopwood, P. and Howell, a (1993). Perception of risk in women with a family history of breast cancer. *British journal of cancer*. 67 (3). 612–4.
- Fagerlin, A., Ubel, P., Smith, D. and Zikmund-Fisher, B. (2007). Making numbers matter: present and future research in risk communication. *American Journal of Health Behvaiour*. 31 (Supp 1). S47– S56.
- Fargher, E.A., Eddy, C., Newman, W.G., Qasim, F., Tricker, K., Elliott, R.A. and Payne, K. (2007). Patients' and healthcare professionals' views on pharmacogenetic testing and its future delivery in the NHS. *Pharmacogenomics*. 8 (11). 1511–9.
- Ferlay, J., Parkin, D.M. and Steliarova-Foucher, E. (2010). Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer*. 46 (4). 765–81.
- Fiebig, D., Keane, M., Louviere, J. and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*. 29 (3). 393–421.
- Findlay, J.M. (2009). Saccadic eye movement programming: sensory and attentional factors. *Psychological research*. 73 (2). 127–35.
- Fishburn, P. and Kochenberger, G. (1979). Two-piece von Neumann-Morgenstern utility function. Decision Sciences. 10 (4). 503–518.
- Fitzpatrick, E., Angus, D., Durieux-Smith, A., Graham, I.D. and Coyle, D. (2008). Parents' needs following identification of childhood hearing loss. *American journal of audiology*. 17 (1). 38–49.
- Fitzpatrick, E., Graham, I.D., Durieux-Smith, A., Angus, D. and Coyle, D. (2007). Parents' perspectives on the impact of the early diagnosis of childhood hearing loss. *International Journal of Audiology*. 46 (2). 97–106.
- Flynn, T. (2010). Using conjoint analysis and choice experiments to estimate QALY values: issues to consider. *PharmacoEconomics*. 28 (9). 711–722.

- Flynn, T., Louviere, J., Peters, T. and Coast, J. (2010). Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social science & medicine*. 70 (12). 1957–1965.
- Font, A.R. (2000). Mass tourism and the demand for protected natural areas: a travel cost approach. *Journal of Environmental Economics and Management*. 39 (1). 97–116.
- Forbes, L. and Ramirez, A. (2014). Offering informed choice about breast screening. *Journal of medical screening*. 21 (4). 194–200.
- Fraenkel, L. (2010). Feasibility of using modified adaptive conjoint analysis. *The Patient*. 3 (4). 209–215.
- Fraenkel, L., Bodardus, S. and Wittink, D.R. (2001). Understanding patient preferences for the treatment of lupus nephritis with adaptive conjoint analysis. *Medical care*. 39 (11). 1203–16.
- Franco, A., Malhotra, N. and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*. 345 (6203). 1502–1505.
- Frederick, S., Loewenstein, G. and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*. 40 (2). 351–401.
- Furedi, A. (1999). The public health implications of the 1995 'pill scare'. *Human Reproduction Update*. 5 (6). 621–626.
- Gafni, A. and Birch, S. (2003). NICE methodological guidelines and decision making in the National Health Service in England and Wales. *PharmacoEconomics*. 21 (3). 149–157.
- Gaissmaier, W. and Gigerenzer, G. (2008). Statistical illiteracy undermines informed shared decision making. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen: ZEFQ. 102 (7). 411–413.
- Gale, N., Heath, G. and Cameron, E. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology*. 1. 117.
- Galesic, M. and Garcia-Retamero, R. (2010). Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*. 170 (5). 462–468.
- Galesic, M. and Garcia-retamero, R. (2013). Using analogies to communicate information about health risks. *Applied Cognitive Psychology*. 27. 33–42.
- Garcia-Retamero, R. and Cokely, E.T. (2011). Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology*. 17 (3). 270–87.
- Gibson, J., Rigby, D., Polya, D. and Russell, N. (2015). Discrete choice experiments in developing countries: willingness to pay versus willingness to work. *Environmental and Resource Economics*. 1–25.
- Gidman, W., Elliott, R.A., Payne, K., Meakin, G.H. and Moore, J. (2007). A comparison of parents and pediatric anesthesiologists' preferences for attributes of child daycase surgery: a discrete choice experiment. *Paediatric Anaesthesia*. 17 (11). 1043–1052.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. and Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest*. 8 (2). 53–96.

- Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B. and Katsikopoulos, K. V (2005). 'A 30% chance of rain tomorrow': how does the public understand probabilistic weather forecasts? *Risk Analysis.* 25 (3). 623–9.
- Gold, M., Siegel, J., Russel, L. and Weinstein, M. (1996). *Cost-effectiveness in health and medicine*. Oxford University Press.
- Gøtzsche, P. and Nielsen, M. (2009). Screening for breast cancer with mammography (Review). *The Cochrane Library*. (4).
- Gøtzsche, P. and Olsen, O. (2000). Is screening for breast cancer with mammography justifiable? *The Lancet*. 355. 129–134.
- Green, J. and Thorogood, N. (2004). *Qualitative Methods for Health Research*. London: Sage Publications.
- Grindrod, K.A., Rosenthal, M., Lynd, L., Marra, C.A., Bougher, D., Wilgosh, C. and Tsuyuki, R.T. (2009). Pharmacists' perspectives on providing chronic disease management services in the community — Part I: Current practice environment. *Canadian Pharmacists Journal*. 142 (5). 234– 239.
- Grutters, J.P.C., Kessels, A.G., Dirksen, C.D., Van Helvoort-Postulart, D., Anteunis, L.J.C. and Joore, M.A. (2008). Willingness to accept versus willingness to pay in a discrete choice experiment. *Value in Health.* 11 (7). 1110–1119.
- Guan, Z., Lee, S., Cuddihy, E. and Ramey, J. (2006). The validity of the stimulated retrospective thinkaloud method as measured by eye tracking. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1253–1262.
- Gunther, O.H., Kurstein, B., Riedel-Heller, S.G. and Konig, H.-H. (2010). The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. *Health services research*. 45 (1). 212–229.
- Hahn, G. and Shapiro, S. (1966). A catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. *Report 66-C-165, General Electric Research and Development Centre, Schenectady, New York.*
- Hakes, J.K. and Viscusi, W.K. (2007). Automobile seatbelt usage and the value of statistical life. *Southern Economic Journal*. 73 (3). 659–676.
- Hall, A.E., Chowdhury, S., Hallowell, N., Pashayan, N., Dent, T., Pharoah, P. and Burton, H. (2014). Implementing risk-stratified screening for common cancers: A review of potential ethical, legal and social issues. *Journal of Public Health*. 36 (2). 285–291.
- Hall, J., Viney, R., Haas, M. and Louviere, J. (2004). Using stated preference discrete choice modeling to evaluate health care programs. *Journal of Business Research*. 57 (9). 1026–1032.
- Hammitt, J.K. and Graham, J.D. (1999). Willingness to pay for health protection: inadequate sensitivity to probability? *Journal of Risk and Uncertainty*. 18 (1). 33–62.
- Han, P., Klein, W., Killam, B., Lehman, T., Massett, H. and Freedman, A.N. (2012). Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. *Patient education and counseling*. 86 (1). 106–113.
- Hanley, N., Ryan, M. and Wright, R. (2003). Estimating the monetary value of health care: lessons from environmental economics. *Health economics*. 12 (1). 3–16.

- Hanley, N., Shogren, J.F. and White, B. (2001). *Introduction to Environmental Economics*. Oxford University Press.
- Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J. and Payne, K. (2014). Risk as an attribute in discrete choice experiments: a critical review. *The patient*. 7 (2). 151–70.
- Hartmann, M. and Fischer, M.H. (2014). Pupillometry: the eyes shed fresh light on the mind. *Current biology*. 24 (7). R281–2.
- Hauber, A., Fairchild, A.O. and Johnson, F. (2013). Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Applied health economics and health policy*. 11 (4). 319–29.
- Hauber, A.B., Johnson, F., Grotzinger, K.M. and Özdemir, S. (2010). Patients' benefit-risk preferences for chronic idiopathic thrombocytopenic purpura therapies. *The Annals of pharmacotherapy*. 44 (3). 479–488.
- Haughney, J., Barnes, G., Partridge, M. and Cleland, J. (2004). The Living & Breathing Study: a study of patients' views of asthma and its treatment. *Primary Care Respiratory Journal*. 13 (1). 28–35.
- Haward, M.F., Murphy, R.O. and Lorenz, J.M. (2008). Message framing and perinatal decisions. *Pediatrics*. 122 (1). 109–18.
- Henderson, J.M., Pollatsek, A. and Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & psychophysics*. 45 (3). 196–208.
- Henneman, L., Oosterwijk, J.C., van Asperen, C.J., Menko, F.H., Ockhuysen-Vermey, C.F., Kostense, P.J., Claassen, L. and Timmermans, D. (2013). The effectiveness of a graphical presentation in addition to a frequency format in the context of familial breast cancer risk communication: a multicenter controlled trial. *BMC medical informatics and decision making*. 13. 55.
- Herbild, L. (2007). Designing a DCE to outlay patients ' and the publics ' preferences for genetic screening in the treatment of depression. *Health Economics Working Papers, University of Southern Denmark*. 1. 1–31.
- Hersch, J., Jansen, J., Barratt, A., Irwig, L., Houssami, N., Howard, K., Dhillon, H. and McCaffery, K. (2013). Women's views on overdiagnosis in breast cancer screening: a qualitative study. *BMJ British Medical Journal*. 346 (f158). 1–13.
- Hess, R. and Siegrist, M. (2011). Risk communication with pictographs: the role of numeracy and graph processing. *Judgment and Decision Making*. 6 (3). 263–274.
- Hey, J.D. and Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*. 62 (6). 1291–1326.
- Hoare, T. (1996). Breast screening and ethnic minorities. *The British journal of cancer. Supplement.* 29. S38–S41.
- Hofvind, S., Thoresen, S. and Tretli, S. (2004). The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. *Cancer*. 101 (7). 1501–1507.
- Hogg, C.N.L. (2007). Patient and public involvement: what next for the NHS? *Health expectations*. 10 (2). 129–38.
- Hole, A.R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*. 27 (4). 1078–1094.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6 (2). 65–70.

- Holmes, T. and Boyle, K. (2004). Dynamic learning and context-dependence in sequential, attributebased contingent valuation. *American Agricultural Economics Association Conference*. 1–31.
- Holmes, T.P. and Adamowicz, W. (2003). Attribute-based methods. In: P. A. Champ, K. J. Boyle, & T. C. Brown (eds.). A Primer on Nonmarket Valuation. The Economics of non-market Goods and Resources. Springer Kluwer Academic Publishers, 171–219.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Hopwood, P., Shenton, A., Lalloo, F., Evans, G. and Howell, A. (2001). Risk perception and cancer worry: an exploratory study of the impact of genetic risk counselling in women with a family history of breast cancer. *Journal of Medical Genetics*. 38 (2). 139–139.
- Hougaard, J.L., Tjur, T. and Osterdal, L.P. (2012). On the meaningfulness of testing preference axioms in stated preference discrete choice experiments. *The European journal of health economics*. 13 (4). 409–17.
- Howard, K. and Salkeld, G. (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.
- Hsieh, Y.-H., Hogan, M.T., Barnes, M., Jett-Goheen, M., Huppert, J., Rompalo, A.M. and Gaydos, C. a (2010). Perceptions of an ideal point-of-care test for sexually transmitted infections--a qualitative study of focus group discussions with medical providers. *PloS one*. 5 (11).
- Huston, P. and Rowan, M. (1998). Qualitative studies: Their role in medical research. *Canadian Family Physician*. 44. 2453.
- Hutton, J.L. (2009). Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. *British journal of haematology*. 146 (1). 27–30.
- Hyman, H. (1944). Do they tell the truth? Public Opinion Quarterly. 557-559.
- Independent UK Panel on Breast Cancer Screening (2012). The benefits and harms of breast cancer screening: an independent review. *Lancet*. 380 (9855). 1778–86.
- Jensen, G.A. and Morrisey, M.A. (2011). Group health insurance: a hedonic price approach. *The Review* of *Economics and Statistics*. 72 (1). 38–44.
- Jepson, R., Clegg, A. and Forbes, C. (2000). The determinants of screening uptake and interventions for increasing uptake: a systematic review. *Health Technology Assessment*. 4 (14).
- Johnson, F., Hauber, A.B., Özdemir, S. and Lynd, L. (2010). Quantifying women's stated benefit-risk trade-off preferences for IBS treatment outcomes. *Value in Health.* 13 (4). 418–423.
- Johnson, F., Van Houtven, G., Özdemir, S., Hass, S., White, J., Francis, G., Miller, D.W. and Phillips, J.T. (2009). Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. *Journal of neurology*. 256 (4). 554–62.
- Johnson, F., Lancsar, E., Marshall, D., Kilambi, V., Mulbacher, A., Regier, D., Bresnahan, B., Kanninen, B. and Bridges, J.F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task. *Value in Health*. 16. 3–13.
- Johnson, F., Mohamed, A.F., Özdemir, S., Marshall, D. and Phillips, K. (2011). How does cost matter in health-care discrete-choice experiments? *Health economics*. 20 (3). 323–330.

Johnson, H. (1958). Demand theory further revised or goods are goods. *Economica*. 25. 148–149.

- Johnson, R.B. and Onwuegbuzie, A. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*. 33 (7). 14–26.
- Johnston, K., Brown, J., Gerard, K., O'Hanlon, M. and Morton, A. (1998). Valuing temporary and chronic health states associated with breast screening. *Social science & medicine*. 47 (2). 213–222.
- Jones, A. (2007). The elgar companion to health economics. 1st Ed. Edward Elgar Publishing Ltd.
- Jorgensen, B., Syme, G., Bishop, B. and Nancarrow, B. (1999). Protest responses in contingent valuation. *Environmental and Resource Economics*. 14 (1). 131–150.
- Just, M. and Carpenter, P. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*. 87 (4). 329–354.
- Kahneman, D. (2012). Thinking, fast and slow. Macmillan.
- Kahneman, D. and Knetsch, J. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal* of environmental economics and management. 22 (1). 57–70.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*. 47. 263–291.
- Kamenova, K., Reshef, A. and Caulfield, T. (2014). Angelina Jolie's faulty gene: newspaper coverage of a celebrity's preventive bilateral mastectomy in Canada, the United States, and the United Kingdom. *Genetics in Medicine*. 16 (7). 522–8.
- Von Karsa, L., Anttila, A., Ronco, G., Ponti, A., Arbyn, M., Segnan, N., Malila, N., Castillo-Beltran, Boniol, M., Ferlay, J. and Hery, C. (2008). Cancer screening in the European Union. *Report on the implementation of the Council Recommendation on cancer screening*.
- Katapodi, M.C., Lee, K., Facione, N.C. and Dodd, M.J. (2004). Predictors of perceived breast cancer risk and the relation between perceived risk and breast cancer screening: a meta-analytic review. *Preventive medicine*. 38 (4). 388–402.
- Keller, C., Siegrist, M. and Visschers, V. (2009). Effect of risk ladder format on risk perception in highand low-numerate individuals. *Risk Analysis*. 29 (9). 1255–64.
- Kennedy, A., Gately, C. and Rogers, A. (2005). Examination of the implementation of the Expert Patients Programme within the structures and locality contexts of the NHS in England (PREPP study) - Report II. *National Primary Care Research and Development Centre*.
- Kerlikowske, K., Grady, D. and Rubin, S. (1995). Efficacy of screening mammography: a metaanalysis. JAMA. 273 (2). 2–3.
- Khan, K., Kunz, R., Kleijnen, J. and Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*. 96. 118–121.
- Kirwan, C.C. (2013). Breast cancer screening: what does the future hold? *BMJ British Medical Journal*. 346 (jan23 1). f87–f87.
- Kjaer, T. and Gyrd-Hansen, D. (2008). Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. *Health policy*. 85 (1). 124–132.
- Kjaer, T., Gyrd-Hansen, D. and Willaing, I. (2006). Investigating patients' preferences for cardiac rehabilitation in Denmark. *International Journal of Technology Assessment in Health Care*. 22 (2). 211–218.
- Kløjgaard, M., Bech, M. and Søgaard, R. (2012). Designing a stated choice experiment: the value of a qualitative process. *Journal of Choice Modelling*. 5 (2). 1–18.

Klose, T. (1999). The contingent valuation method in health care. *Health policy*. 47. 97–123.

- Kowler, E., Anderson, E., Dosher, B. and Blaser, E. (1995). The role of attention in the programming of saccades. *Vision research*. 35 (13). 1897–916.
- Krinsky, I. and Robb, A. (1986). On approximating the statistical properties of elasticities. *The Review* of Economics and Statistics. 68. 715–719.
- Kruijshaar, M.E., Essink-Bot, M.L., Donkers, B., Looman, C.W.N., Siersema, P.D. and Steyerberg, E.W. (2009). A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. *BMC medical research methodology*. 9. 31.
- Laeng, B., Sirois, S. and Gredeback, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on Psychological Science*. 7 (1). 18–27.
- Lagarde, M. (2012). Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health Economics*. 22 (5). 554–567.
- Lancaster, K.J. (1966). A new approach to consumer theory. *The Journal of Political Economy*. 74 (2). 132–157.
- Lancsar, E. and Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *PharmacoEconomics*. 26 (8). 661–677.
- Lancsar, E. and Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health economics*. 15 (8). 797–811.
- Lane, D. a and Lip, G.Y. (2001). Ethnic differences in hypertension and blood pressure control in the UK. *Quarterly Journal of Medicine*. 94 (7). 391–6.
- LaRosa, J., He, J. and Vupputuri, S. (1999). Effect of statins on risk of coronary disease: a metaanalysis of randomized controlled trials. *JAMA*. 282 (24).
- Leal, J., Luengo-Fernandez, R. and Sullivan, R. (2012). The economic burden of malignant neoplasms in the European Union. 2012 Congress of the European Society for Medical Oncology.
- Leow, R.P. and Morgan-short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in second language acquisition*. 26 (1). 35–57.
- Lincoln, Y.S. and Guba, E.G. (1985). Naturalistic inquiry. Evaluation and Program Planning.
- Lindholm, L.H., Carlberg, B. and Samuelsson, O. (2005). Should beta blockers remain first choice in the treatment of primary hypertension? A meta-analysis. *Lancet*. 366 (9496). 1545–53.
- Lipkus, I. (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making*. 27 (5). 696–713.
- Lipkus, I. and Hollands, J. (1999). The visual communication of risk. JNCI monographs. 25. 149–163.
- Lipkus, I., Samsa, G. and Rimer, B. (2001). General performance on a numeracy scale among highly educated samples. *Medical decision making*. 21 (1). 37–44.
- Lloyd, A.J. (2003). Threats to the estimation of benefit: are preference elicitation methods accurate? *Health economics*. 12 (5). 393–402.
- Long, A. and Godfrey, M. (2004). An evaluation tool to assess the quality of qualitative research studies. *International Journal of Social Research Methodology*. 7 (2). 181–196.

- Louviere, J. (2006). What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. *Environmental and Resource Economics*. 34 (1). 173–188.
- Louviere, J. and Flynn, T. (2010). Using best-worst scaling choice experiments to measure public perceptions and preferences for healthcare reform in Australia. *The Patient*. 3 (4). 275–283.
- Louviere, J., Flynn, T. and Carson, R.T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*. 3 (3). 57–72.
- Louviere, J., Hensher, D. and Swait, J. (2000). *Stated choice methods: analysis and application*. Cambridge University Press.
- Louviere, J. and Islam, T. (2008). A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research*. 61 (9). 903–911.
- Luborsky, M.R. and Rubinstein, R.L. (1995). Sampling in qualitative research: rationale, issues, and methods. *Research on aging*. 17 (1). 89–113.
- Lundgrén-Laine, H. and Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. Qualitative health research. 20 (4). 565–75.
- Manor, B.R. and Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*. 128 (1-2). 85–93.
- Mansfield, A., Nathanson, V., Jayesinghe, N., Roycroft, G. and Panizzo, R. (2012). Risk: what's your perspective? a guide for healthcare professionals. *British Medical Association*.
- Mark, T.L., Kranzler, H.R., Poole, V.H., Hagen, C., McLeod, C. and Crosse, S. (2003). Barriers to the use of medications to treat alcoholism. *American Journal on Addictions*. 12 (4). 281–294.
- Marshall, M. (1996). Sampling for qualitative research. Family practice. 13 (6). 522–525.
- Marti, J. (2012). A best-worst scaling survey of adolescents' level of concern for health and non-health consequences of smoking. *Social science & medicine*. 75 (1). 87–97.
- Mason, J. (2002). Qualitative researching. Sage.
- McCaffery, K., Dixon, A., Hayen, A., Jansen, J., Smith, S. and Simpson, J.M. (2012). The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. *Medical Decision Making*. 32 (4). 532–44.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In: P. Zarembka (ed.). *Frontiers in Econometrics*. New York: Academic Press INC, 105–142.
- McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*. 5 (4). 275–297.
- McIntosh, E., Clarke, P., Frew, E. and Louviere, J. (2010). *Applied Methods of Cost-Benefit Analysis in Health Care*. Oxford University Press.
- McMenamin, M., Barry, H., Lennon, A.-M., Purcell, H., Baum, M., Keegan, D., McDermott, E., O'Donoghue, D., Daly, L. and Mulcahy, H. (2005). A survey of breast cancer awareness and knowledge in a Western population: lots of light but little illumination. *European Journal of Cancer*. 41 (3). 393–397.

McNeil, B., Pauker, S., Sox, H. and Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *The New England journal of medicine*. 306. 1259–1262.

Microsoft (2010). Excel.

- Midlands Therapeutics Review & Advisory Committee (2008). Colesevelam for the treatment of primary and hypercholesterolaemia. *MTRAC Verdict and Summary*. 1–2.
- Mishel, M. (1990). Reconceptualization of the uncertainty in illness theory. *Image: The Journal of Nursing Scholarship*. 22 (4). 256–262.
- Mockford, C., Staniszewska, S., Griffiths, F. and Herron-Marx, S. (2011). The impact of patient and public involvement on UK NHS health care: a systematic review. *International journal for quality in health care*. 1–11.
- Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 6 (7). e1000097.
- Montgomery, A. and Fahey, T. (2001). How do patients' treatment preferences compare with those of clinicians? *Quality in health care*. 10 (Suppl I). i39–43.
- Moore, R.A., Derry, S., McQuay, H.J. and Paling, J. (2008). What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. *Arthritis research & therapy*. 10 (1). 1–16.
- Morton, R., Devitt, J., Howard, K., Anderson, K., Snelling, P. and Cass, A. (2010). Patient views about treatment of stage 5 CKD: a qualitative analysis of semistructured interviews. *American Journal of Kidney Diseases*. 55 (3). 431–440.
- Moss, S.M., Cuckle, H., Evans, A., Johns, L., Waller, M. and Bobrow, L. (2006). Effect of mammographic screening from age 40 on breast cancer mortality at 10 years follow-up: a randomized controlled trial. *The Lancet*. 368 (9552). 2053–2206.
- Muangchana, C. and Warinsatian, P. (2011). Incorporation of private demand into cost-benefit analysis of a universal Hib vaccination program in Thailand. *Southeast Asian Journal of Tropical Medicineand Public Health*. 42 (2). 376–87.
- Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N. and Tsuchiya, A. (2015). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health.* 16 (1). 104–13.
- Nafees, B., Lloyd, A., Kennedy-Martin, T. and Hynd, S. (2006). How diabetes and insulin therapy affects the lives of people with type 1 diabetes. *European Diabetes Nursing*. 3 (2). 92–97.
- Naik Panvelkar, P., Armour, C. and Saini, B. (2010). Community pharmacy-based asthma serviceswhat do patients prefer? *The Journal of Asthma*. 47 (10). 1085–93.
- NCDDR (2004). What are the standards for quality research? *National Center for the Dissemination of Disability Research*. (9).
- Netten, A., Burge, P., Malley, J., Brazier, J., Flynn, T. and Forder, J. (2009). Outcomes of Social Care for Adults (OSCA) Interim Findings. *Personal Social Services Research Unit*. Discussion.
- Netten, A., Ryan, M., Smith, P., Skatun, D., Healey, A., Knapp, M. and Wykes, T. (2002). The development of a measure of social care outcome for older people. *Personal Social Services Research Unit*. Discussion.
- Von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behavior*. 2nd Ed. Princeton University Press.

- Nghi, N.Q., LaMontagne, S., Bingham, A., Rafiq, M., Phoung Mai, L.T. and Lien, N.T.P. (2010). Human papillomavirus vaccine introduction in Vietnam: formative research findings. *Sexual health*. 7 (3). 262–270.
- NICE (2013). Guide to the methods of technology appraisal. *Process and Methods Guides*. (April). 1–102.
- Nielsen, J.S., Gyrd-Hansen, D. and Kjaer, T. (2012). Valuation of morbidity and mortality risk reductions: does context matter? *Accident; analysis and prevention.* 48. 246–53.
- Novick, G. (2008). Is there a bias against telephone interviews in qualitative research? *Research in nursing & health*. 31 (4). 391–398.
- Nylund, K. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*. 14 (4). 535–569.
- Nyström, M., Andersson, R., Holmqvist, K. and van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*. 45 (1). 272–88.
- O'Brien, B. and Gafni, A. (1996). When do the 'dollars' make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical decision making*. 16 (3). 288–99.
- O'Cathain, A. and Thomas, K.J. (2004). 'Any other comments?' Open questions on questionnaires a bane or a bonus to research? *BMC medical research methodology*. 4. 25.
- O'Shea, J. and Harvey, R. (2003). Age related macular degeneration. A *Photo-Essay for Health Professionals*.
- Ohuchi, N., Yoshida, K. and Kimura, M. (1995). Comparison of false negative rates among breast cancer screening modalities with or without mammography: Miyagi trial. *Japanese journal of cancer research*. 86. 501–506.
- ONS (2012). What are the top causes of death by age and gender ? Mortality Statistics: Deaths Registered in England and Wales. *Office for National Statistics (ONS)*. 2–4.
- Orme, B. (1998). Sample size issues for conjoint analysis studies. *Sawthooth Software Research paper series*.
- Otero-Millan, J., Alba Castro, J., Macknik, S. and Martinez-Conde, S. (2014). Unsupervised clustering method to detect microsaccades. *Journal of Vision*. 14 (2, 18). 1–17.
- Paling, J. (2003). Strategies to help patients understand risks. BMJ British Medical Journal. 745–748.
- Parkin, D. and Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health economics*. 15 (7). 653–64.
- Payne, K., Fargher, E.A., Roberts, S.A., Tricker, K., Elliott, R. a, Ratcliffe, J. and Newman, W.G. (2011). Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. *Value in Health*. 14 (1). 121–134.
- Pearson, A. and Field, J. (2011). JBI critical appraisal checklist for qualitative research. *Joanna Briggs Institute*.
- Pennery, E.L. (2007). A personal viewpoint against the motion that breast cancer patient advocacy groups unfairly distort the cancer budget. *Breast Cancer Research*. 9 (Suppl 2). S5.
- Perman, R., Ma, Y., Common, M., Maddison, D. and McGilvary, J. (2003). *Natural resource and environmental economics*. Pearson Education.

- Peters, E., Hart, P.S. and Fraenkel, L. (2011). Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making*. 31. 432–6.
- Pharoah, P., Sewell, B. and Fitzsimmons, D. (2013). Cost effectiveness of the NHS breast screening programme: life table model. *BMJ British Medical Journal*. 346.
- Pitchforth, E., Watson, V., Tucker, J., Ryan, M., van Teijlingen, E., Farmer, J., Ireland, J., Thomson, E., Kiger, A. and Bryers, H. (2008). Models of intrapartum care and women's trade-offs in remote and rural Scotland: a mixed-methods study. *BJOG*. 115 (5). 560–569.
- Polak, J. and Liu, X. (2006). From random utility to random expected utility: theory and application to departure time choice. *European Transport Conference Advances in Discrete Choice Models*.
- Popay, J., Rogers, A. and Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*. 8 (3). 341–351.
- Pope, C. and Mays, N. (2008). Qualitative research in health care. 3rd editio. John Wiley & Sons.
- Preisendörfer, P. and Wolter, F. (2014). Who is telling the truth? A validation study on determinants of response behavior in surveys. *Public Opinion Quarterly*. 78 (1). 126–146.
- Public Health England (2013a). *NHS breast screening website for England: Costs.* [Online]. 2013. Available from: http://www.cancerscreening.nhs.uk/breastscreen/cost.html.
- Public Health England (2013b). *NHS breast screening website for England: What does the NHS Breast Screening Programme do?* [Online]. 2013. Available from: http://www.cancerscreening.nhs.uk/breastscreen/.
- Puri, V., Crabtree, T.D., Kymes, S., Gregory, M., Bell, J., Bradley, J.D., Robinson, C., Patterson, G.A., Kreisel, D., Krupnick, A.S. and Meyers, B.F. (2012). A comparison of surgical intervention and stereotactic body radiation therapy for stage I lung cancer in high-risk patients: a decision analysis. *The Journal of thoracic and cardiovascular surgery*. 143 (2). 428–36.
- Purves, D., Augustine, G. and Fitzpatrick, D. (2001). Types of eye movements and their functions. *Neuroscience*.
- QSR (2012). NVivo qualitative data analysis software: Version 10.0.
- Rabin, M. (2000). Risk aversion and expected-utility theory: a calibration theorem. *Econometrica*. 68 (5). 1281–1292.
- Randall, A. (1994). A difficulty with the travel cost method. Land Economics. 70 (1). 88–96.
- Raney, G.E., Campbell, S.J. and Bovee, J.C. (2014). Using eye movements to evaluate the cognitive processes involved in text comprehension. *Journal of visualized experiments*. 83 (e50780). 1–7.
- Ratcliffe, J., Couzner, L., Flynn, T., Sawyer, M., Stevens, K., Brazier, J. and Burgess, L. (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. *Applied health economics and health policy*. 9 (1). 15–27.
- Ratcliffe, J., Flynn, T., Terlich, F., Stevens, K., Brazier, J. and Sawyer, M. (2012). Developing adolescent-specific health state values for economic evaluation. *PharmacoEconomics*. 30 (8). 713–727.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*. 62 (8). 1457–506.

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin.* 124 (3). 372–422.
- Reeves, S., Kuper, A. and Hodges, B. (2008). Qualitative research methodologies: ethnography. BMJ British Medical Journal. 337. 512–514.
- Renn, O. (2006). Risk communication consumers between information and irritation. *Journal of Risk Research*. 9 (8). 833–849.
- Renshaw, C., Jack, R. and Dixon, S. (2010). Estimating attendance for breast cancer screening in ethnic groups in London. *BMC public health*. 10 (157). 1–8.
- Retzlaff-Roberts, D., Chang, C.F. and Rubin, R.M. (2004). Technical efficiency in the use of health care resources: a comparison of OECD countries. *Health Policy*. 69 (1). 55–72.
- Robinson, A., Spencer, A. and Moffatt, P. (2015). A Framework for Estimating Health State Utility Values within a Discrete Choice Experiment Modeling Risky Choices. *Medical Decision Making*. 35 (3). 341–350.
- Robst, J. (2006). Estimation of a hedonic pricing model for Medigap insurance. *Health services research*. 41 (6). 2097–113.
- Rohlfs, C., Sullivan, R. and Kniesner, T. (2015). New estimates of the value of a statistical life using air bag regulations as a quasi-experiment. *American Economic Journal*. 7 (1). 331–359.
- Rothman, A. and Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin*. 121 (1). 3–19.
- Russo, J., Johnson, E. and Stephens, D. (1989). The validity of verbal protocols. *Memory & cognition*. 17 (6). 759–769.
- Ryan, M. (2004). Discrete choice experiments in health care. *BMJ (Clinical research ed.)*. 328 (7436). 360–1.
- Ryan, M. and Gerard, K. (2003a). Using discrete choice experiments in health economics: moving forward. In: A. Scott, A. Maynard, & R. Elliott (eds.). *Advances in Health Economics*. 25–40.
- Ryan, M. and Gerard, K. (2003b). Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy*. 2 (1). 55–64.
- Ryan, M., Gerard, K. and Amaya-Amaya, M. (2008). Using discrete choice experiments to value health and health care (The Economics of Non-Market Goods and Resources). The Economics of Non-Market Goods and Resources. Springer.
- Ryan, M. and Watson, V. (2011). Using Discrete Choice Experiments in Health Economics: Theoretical and Practical Issues.
- Ryan, M., Watson, V. and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of a discrete choice experiment responses. *Health Economics*. 18. 321–336.
- Salisbury, C., Banks, J., Goodall, S., Baxter, H., Montgomery, A., Pope, C., Gerard, K., Simons, L., Lattimer, V., Sampson, F., Pickin, M., Edwards, S., Smith, H. and Boudioni, M. (2007). An evaluation of Advanced Access in general practice. *National Co-ordinating Centre for NHS Service Delivery and Organisation Research and Development*.
- Salvucci, D. and Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols. Proceedings of the 2000 symposium on Eye Tracking Research and Applications Symposium. 71– 78.

- Sapir, R., Patlas, M. and Strano, S. (2003). Does mammography hurt? *Journal of pain and symptom management*. 25 (1). 53–63.
- Sawtooth (2012). Sawtooth Software SSI Web 8.3.8.
- Scarpa, R., Gilbride, T., Campbell, D. and Hensher, D. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*. 36 (2). 151–174.
- Ščasný, M. and Alberini, A. (2012). Valuation of mortality risk attributable to climate change: investigating the effect of survey administration modes on a VSL. *International Journal of Environmental Research and Public Health*. 9 (12). 4760–4781.
- Schapira, M.M., Nattinger, A.B. and McHorney, C.A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*. 21 (6). 459– 467.
- Schmidt, M. (2004). Investigating risk perception: a short introduction. In: Loss of agro-biodiversity in Vavilov centers, with a special focus on the risks of genetically modified organisms (GMOs). Vienna: Schmidit PhD Thesis, 1–16.
- Schwartz, L., Woloshin, S., Black, W. and Welch, H.G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*. 127 (11). 1023–1028.
- Sharma, S., Brown, G.C., Brown, M., Hollands, H., Robins, R. and Shah, G.K. (2002). Validity of the time trade-off and standard gamble methods of utility assessment in retinal patients. *The British journal of ophthalmology*. 86 (5). 493–6.
- Shavers-Hornaday, V.L., Lynch, C.F., Burmeister, L.F. and Torner, J.C. (1997). Why are African Americans under-represented in medical research studies? Impediments to participation. *Ethnicity* & *Health*. 2 (1-2). 31–45.
- Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*. 41. 2915–2924.
- Shenton, A.K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education* for information. 22 (2). 63–75.
- Silverman, D. (2013). Doing qualitative research: A practical handbook.
- Silverman, D. (1993). Interpreting Qualitative Data. Sage Publications.
- Slaytor, E. and Ward, J. (1998). How risks of breast cancer and benefits of screening are communicated to women: analysis of 58 pamphlets. *BMJ British Medical Journal*. 8 (14). 263–265.
- Slovic, P., Peters, E., Finucane, M.L. and Macgregor, D.G. (2005). Affect, risk, and decision making. *Health psychology*. 24 (4 Suppl). S35–40.
- Smith, J. and Firth, J. (2011). Qualitative data analysis: the framework approach. *Nurse researcher*. 52–62.
- Smith Paintain, L., Antwi, G., Jones, C., Amoako, E., Adjei, R.O., Afrah, N. a, Greenwood, B., Chandramohan, D., Tagbor, H. and Webster, J. (2011). Intermittent screening and treatment versus intermittent preventive treatment of malaria in pregnancy: provider knowledge and acceptability. *PloS one*. 6 (8).
- Smith, R.D. (2000). The discrete-choice willingness-to-pay question format in health economics: should we adopt environmental guidelines? *Medical Decision Making*. 20. 194–204.

- Smith, R.D. and Sach, T.H. (2010). Contingent valuation: what needs to be done? *Health economics*, policy and law. 5 (Pt 1). 91–111.
- Sprague, D., Russo, J., Lavallie, D.L. and Buchwald, D.S. (2012). Influence of framing and graphic format on comprehension of risk information among American Indian tribal college students. *Journal of Cancer Education*. 27. 752–8.
- SR Research (2013). Data Viewer Software.
- SR Research (2012). Experiment Builder Software.
- StataCorp (2011). Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.
- Statistical Innovations (2013). Latent Gold.
- Stewart, D., Wilson, R., Selby, P. and Darbyshire, J. (2011). Patient and public involvement. Annals of oncology. 22 (Suppl 7). vii54–vii56.
- Stiggelbout, A.M., Wit, M.P.T. De, Frosch, D., Montori, V.M., Trevena, L. and Elwyn, G. (2012). Shared decision making: really putting patients at the centre of healthcare. *BMJ British Medical Journal*. 344 (e256). 1–6.
- Stolk, E.A., Oppe, M., Scalone, L. and Krabbe, P.F. (2010). Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value in health*. 13 (8). 1005–1013.
- Street, D. and Burgess, L. (2007). The Construction of Optimal Stated Choice Experiments: Theory and Methods. London: Wiley.
- Street, D., Burgess, L. and Louviere, J. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing*. 22 (4). 459–470.
- Sunstein, C.R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty*. 26 (2-3). 121–136.
- Tabár, L., Faberberg, G., Day, N.E. and Holmberg, L. (1987). What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *British journal of cancer*. 55 (5). 547–551.
- Taylor, K. and Dionne, J. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*. 92 (3). 413.
- Taylor, P. (2010). Modelling the impact of changes in sensitivity on the outcomes of the UK breast screening programme. *Journal of medical screening*. 17 (1). 31–36.
- Telser, H. and Zweifel, P. (2007). Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics*. 39 (1). 69–78.
- The NHS Information Centre Screening and Immunisations (2012). Breast Screening Programme 2010-2011. www.ic.nhs.uk.
- Thiene, M., Boeri, M. and Chorus, C. (2012). Random regret minimization: exploration of a new choice model for environmental and resource economics. *Environmental and resource economics*. 51 (3). 413–429.

Thurstone, L. (1927). A law of comparative judgment. Psychological review. 34 (4). 273-286.

Torgerson, D.J. and Raftery, J. (1999). Discounting. BMJ British Medical Journal. 319. 914–915.

- Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal: a review. *Journal* of health economics. 5. 1–30.
- Torres, C., Hanley, N. and Riera, A. (2011). How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management*. 62 (1). 111–121.
- Train, K. (2009). Discrete choice methods with simulation. 2nd Ed. Cambridge University Press.
- Tversky, A. (1972). Choice by elimination. Journal of mathematical psychology. 9. 341–367.
- Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*. 211. 453–458.
- US Department of Health and Human Services (2013). Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *National Institute of Health*.
- Vansteenkiste, P., Cardon, G., Philippaerts, R. and Lenoir, M. (2014). Measuring dwell time percentage from head-mounted eye-tracking data comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics*. 1–10.
- Varian, H.A.L.R. (1992). Microeconomic Analysis. 3rd Ed.
- Venkatachalam, L. (2004). The contingent valuation method: a review. *Environmental Impact* Assessment Review. 24 (1). 89–124.
- Viney, R., Lancsar, E. and Louviere, J. (2002). Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Review of Pharmacoeconomics and Outcomes Research*. 2 (4). 319–26.
- Viney, R., Savage, E. and Louviere, J. (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health economics*. 14 (4). 349–62.
- Walsh, D. and Downe, S. (2006). Appraising the quality of qualitative research. *Midwifery*. 22 (2). 108–19.
- Waltzman, J.T., Scholz, T. and Evans, G. (2011). What patients look for when choosing a plastic surgeon: an assessment of patient preference by conjoint analysis. *Annals of plastic surgery*. 66 (6). 643–647.
- Wang, J.T. (2010). Pupil Dilation and Eye-tracking. In: M. Schulte-Mecklenbeck, A. Kuhberger, & R. Ranyard (eds.). A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide. Psychology Press, 185–204.
- Watson, V., Ryan, M. and Watson, E. (2009). Valuing experience factors in the provision of Chlamydia screening: an application to women attending the family planning clinic. *Value in Health.* 12 (4). 621–623.
- Weinstein, M., Siegel, J. and Gold, M. (1996). Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA*. 276 (15). 1253–1258.
- Weinstein, N. (1998). Accuracy of smokers' risk perceptions. *Annals of behavioral medicine*. 20 (2). 135–140.
- Weinstein, N. (1984). Why it won't happen to me: perceptions of risk factors and susceptibility. *Health psychology*. 3 (5). 431–457.
- Welch, H.G. and Black, W.C. (2010). Overdiagnosis in cancer. *Journal of the National Cancer Institute*. 102 (9). 605–13.

- Welch, H.G. and Frankel, B. (2011). Likelihood that a woman with screen-detected breast cancer has had her 'life saved' by that screening. *Archives of Internal Medicine*. 171 (22). 2043–2046.
- Weller, D., Moss, S.M., Butler, P., Campbell, C., Coleman, D., Melia, J. and Robertson, R. (2006). English pilot of bowel cancer screening: evaluation of the second round. *NHS*. 1–226.
- Whitehead, S.J. and Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British medical bulletin.* 96. 5–21.
- Whitty, J., Walker, R., Golenko, X. and Ratcliffe, J. (2014). A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PloS one*. 9 (4). e90635.
- Willis, K. and Garrod, G. (1991). An individual travel cost method of evaluating forest recreation. *Journal of Agricultural Economics*. 42 (1). 33–42.
- De Wit, G., Busschbach, J.J. and De Charro, F.T. (2000). Sensitivity and perspective in the valuation of health status: whose values count? *Health economics*. 9 (2). 109–26.
- Witt, J., Scott, A. and Osborne, R.H. (2009). Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. *Health Economics*. 18 (6). 681–696.
- Witter, S., Thi Thu Ha, B., Shengalia, B. and Vujicic, M. (2011). Understanding the 'four directions of travel': qualitative research into the factors affecting recruitment and retention of doctors in rural Vietnam. *Human Resources for Health*. 9 (20).
- Women's Eye Health Task Force (2003). Age-Related Macular Degeneration. *Eye Research Institute Harvard*.
- Wooldridge, J. (2008). Introductory econometrics. 4th Ed. South Western College.
- Wu, C., Maurer, C., Wang, Y., Xue, S. and Davis, D. (1999). Water pollution and human health in China. *Environmental health perspectives*. 107 (4). 251–256.
- Xie, F., Pullenayegum, E., Gaebel, K., Oppe, M. and Krabbe, P.F.M. (2014). Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best-worst scaling? *European Journal of Health Economics*. 15 (3). 281–8.
- Yaffe, M. and Mainprize, J. (2011). Risk of radiation-induced breast cancer from mammographic screening. *Radiology*. 258 (1). 98–105.
- Yancey, A.K., Ortega, A.N. and Kumanyika, S.K. (2006). Effective recruitment and retention of minority research participants. *Annual review of public health*. 27. 1–28.
- Zeckhauser, R. and Sunstein, C.R. (2009). Dreadful Possibilities, Neglected Probabilities. *Harvard Environmental Economics Program: Discussion Papers*. 9 (3). 1–8.

Using discrete choice experiments to value benefits and risks in primary care

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

2015

Caroline Mary Vass School of Medicine

Volume II of II

Volume II

List of Appendices

Appendix 3.1: How to frame risk in a DCE: an exploration and structured review
Appendix 3.2: List of studies included in the structured review of risk communication
formats
Appendix 3.3: The published paper containing the results of the structured review274
Appendix 3.4: List of DCEs with a focus on breast cancer and screening from the
systematic review
Appendix 3.5: The design of the DCE used in the pilot study
Appendix 3.6: Review of questions used to elicit individual characteristics in commonly-
used surveys
Appendix 3.7: Confirmation letter of University Research Ethics approval
Appendix 3.8: The DCE survey used for the pilot study
Appendix 3.9: Online advertisement for study participants
Appendix 3.10: Preliminary interview schedule for pilot study
Appendix 3.11: Information sheet for interviewees for the qualitative interview study341
Appendix 3.12: Consent form for interviewees for the qualitative interview study
Appendix 3.13: NHS Choices video approval
Appendix 3.14: Pilot quantitative results of the DCE from the internet panel
Appendix 3.15: The design of the DCE used in the final empirical studies
Appendix 3.16: The DCE survey used for the final empirical studies
Appendix 3.17: Quantitative results of the DCE from the internet panel in the second pilot
study
Appendix 4.1: ResearchNow® response rate in the large internet panel DCE
Appendix 4.2: Sample characteristics of respondents who completed the DCE in the final
study
Appendix 4.3: Results of a simple logit regression to determine if any covariates were a
significant predictor of risk communication format
Appendix 4.4: Latent class analysis results from 6-preference-class (no scale-classes)
model for the large internet panel DCE
Appendix 5.1: Systematic review of qualitative research alongside healthcare DCEs395
Appendix 5.2: Survey to authors of published healthcare DCEs420
Appendix 5.3: Flow of papers through the systematic review

Appendix 5.4: Comparison of existing appraisal tools for assessing the quality of
qualitative research
Appendix 5.5: Final critical appraisal tool for assessing the quality of qualitative research
alongside DCEs
Appendix 5.6: Studies included in the systematic review of qualitative research conducted
alongside DCEs436
Appendix 5.7: Tabulated results from the systematic review of DCEs
Appendix 5.8: List of studies not included in the Clark et al. (2014) systematic review455
Appendix 5.9: The submitted paper describing the development of the appraisal tool456
Appendix 5.10: The questionnaire sent to authors of DCE studies to assess their experience
and views of using qualitative research methods
Appendix 5.11: The invitation to authors of DCE studies to complete the questionnaire to
assess their experience and views of qualitative research methods alongside DCEs478
Appendix 5.12: The reminder invitation to authors of DCE studies to complete the
questionnaire to assess their experience and views of qualitative research methods
alongside DCEs
Appendix 5.13: Author responses to the questionnaire of their experience and views of
conducting qualitative research alongside DCEs
Appendix 5.14: Paper advertisement to the general public for participation in the think-
aloud interviews
Appendix 5.15: Email advertisements to The University of Manchester staff for
participation in the think-aloud interviews
Appendix 5.16: Think-aloud interview schedule
Appendix 5.17: Flow of recruited study participants for the think-aloud interviews and the
eye-tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye-
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study
Appendix 6.1: Paper advertisement to the general public for participation in the eye- tracking study

Appendix 3.1: How to frame risk in a DCE: an exploration and structured review

Aims

The overall aim of this review was to systematically identify different formats which can be used to communicate risk information to DCE respondents. If risk is communicated effectively, then it will assist the respondents in making informed choices based on their preferences and thus improve the quality of the data collected. The review is described as 'structured' to distinguish it from the systematic review presented in Appendix 5.1, and because no second screening of abstracts was conducted.

Methods

The following section describes the steps taken to identify studies using different risk communication formats.

Search strategy

A systematic search of five electronic databases (Medline, Embase, Web of Science, PsychINFO and EconLit) was conducted in April 2013. The search checked the titles of articles in each database for the terms risk and either communication or format, and the whole abstract for topic terms (aid, presentation, display). The exact strategy included the following terms: Title=(risk*) AND Title=(communicat* or format*) AND Topic=(aid* or present* or display*).

The search strategy aimed for a low recall and very high precision and was developed based on extensive meetings with an expert librarian (personal communication with Mary Ingram). The databases were searched for studies published between January 2000 and April 2013. Any valuable communication formats reported in papers published prior to 2000 would likely be reproduced or compared in a later study which would be captured by the selected time horizon.

Screening

An initial title and abstract screen to select relevant studies was conducted. If studies could not be rejected with certainty, the whole article was retrieved and reviewed. Studies were eligible for inclusion if they: 1) tested or compared risk communication formats which aimed to aid people's understanding of quantitative information; 2) related to a healthcare risk (the probability of a health-related outcome); and 3) were published in peer-reviewed journals written in English. Any studies which were not peer-reviewed, were not written in English were excluded. Studies reviewing public health risks, such as nationwide vaccines,

were also excluded. This was because public health campaigns often looked at communicating risk to promote a particular behaviour or change preferences (for example, describing cancer risks to encourage smoking cessation) rather than objectively providing information. Studies not using written communication, such as interactive games, were also excluded because of programming challenges meaning they are not easily incorporated into a typical DCE.

After retrieving all the full papers, the studies were initially categorised into four categories: 1) empirical studies testing risk communication formats; 2) systematic reviews of risk communication formats; 3) papers providing an overview; and 4) studies specifically communicating uncertainty around point estimates. In this review no specific assessment for methodological quality was conducted; the studies were so different that no appropriate appraisal criteria existed.

Data extraction

Empirical studies comparing different risk communication formats were viewed to be of most interest. Data from these empirical studies were tabulated using the following headings: risk communication format; outcome measures; sample size; sample characteristics; and study conclusion. However, studies were also included if they provided a review or overview of different methods. Studies which looked at communicating uncertainty rather than risk were also acknowledged but did not undergo data extraction for the reasons explained in. The identified review pieces, overview papers and uncertainty studies are drawn upon in the discussion in section of this review.

Results

Figure A1 summarises the results of the search strategy and screening process (details about the search strategy and results can be found in Table A1). The search of all databases revealed 1,207 possible hits. After the removal of non-peer-reviewed material and duplicates, 390 titles and abstracts were reviewed. In total, 215 full articles were retrieved for further review and 99 papers were included in the final review.

Database	Search Terms	Hits
Web of Science	Title=(risk*) AND	482
	Title=(communicat* or format*)	
	AND Topic=(aid* or present* or	
	display*) [limited 2000-2013;	
	English Language]	
Medline	risk*(ti) AND (communicat* or	223
	format*)(ti) AND (aid* or	
	present* or display*)(ti) [limited	
	2000-2013; English Language]	
Embase	risk*(ti) AND (communicat* or	313
	format*)(ti) AND (aid* or	
	present* or display*)(ti) [limited	
	2000-2013; English Language]	
EconLit	risk*(ti) AND (communicat* or	7
	format*)(ti) AND (aid* or	
	present* or display*)(ti) [limited	
	2000-2013; English Language]	
PscyhINFO	risk*(ti) AND (communicat* or	182
	format*)(ti) AND (aid* or	
	present* or display*)(ti) [limited	
	2000-2013; English Language]	
Overall		1207
Removal of duplicates		470
Removal of books, patents, disser	390	

Table A1: Search results from all databases for the systematic review of risk communication methods, run on the 23rd April 2013

Of the studies excluded, most were either not exploring risk communication, were not looking at healthcare risks or were not using written risk communication formats. Appendix 3.2 list all 99 studies included in the review. Each included study was given an ID comprising a letter to indicate the type of study (R for risk; U for uncertainty; O for overview papers; and S for systematic reviews) and a number for identification. Of the included studies, 65 were empirical tests of risk communication formats, 21 were overviews of the risk literature with no reported search strategy, nine were systematic reviews with a structured search strategy, and four looked at presenting uncertainty (for example, the use of confidence intervals as opposed to point estimates).

Figure A1: Study selection process



Risk communication formats

The 65 empirical studies revealed a number of different ways of communicating risk information. Table A2 summarises the empirical studies identified by the review with a summary of the formats compared and which formats were found to be most and least effective as reported by the study authors.

The most commonly compared formats were numerical presentations of risk statistics such as percentages, natural frequencies (NF), decimals or ratios (n=55, 85%) either alone or in combination. When risk was presented as a frequency, multiple methods were tested including formatting the probability as X in 100 and 1 in X.

Other risk communication formats tested were visual images such as matrices of coloured shapes called 'icon arrays' (n=29, 45%) (these can also be referred to as 'risk grids' or 'pictographs'). Risk ladders were used in ten studies (15%) where the assigned risk was presented on a scale ranking it along with other probabilities (this was referred to as a risk continuum in R3; a risk thermometer in R47; and involved a magnifying graphic in R60). Two studies (R32, R36) used a risk ladder called the Paling Perspective Scale (Paling, 2003).

A number of studies also used graphical information (n=15, 23%), presenting the risk in bar charts (both horizontal and vertical), pie charts or line graphs. 11 studies (17%) used epidemiological measures to explain the risk. This included: number needed to treat (NNT) which is the inverse of the absolute risk reduction (ARR); number needed to harm (NNH) which is the inverse of the attributable risk when the NNT is negative; relative risk reduction (RRR); and survival curves (Hutton, 2009).

Thirty four studies (52 %) used qualitative descriptions of the risk, verbal analogies or vignettes. The verbal descriptors were sometimes used alone but were often combined with percentages (see R4, R8 and R52). R40, R44 and R59 used a traffic light to represent the risk information indicating low (green), medium (amber), and high (red) risk values.

ID (vear)	Type of risk of	communication format e	Overall conclusion					
(year)	Numerical	Frequency format	Graphical risk	Pictorial risk	Qualitative	Epidemiology	Most effective	Least effective
	format				format			
R1 (2002)					Vignette		Personal vignettes	Impersonal vignettes
R2	%				Verbal descriptors		%	Verbal
(2004)					r			descriptors
R3(201	% and	X in 100	Bar chart	Icon arrays, ladder*	Verbal descriptors,		Ladder*	Icon arrays
2)	frequency				Vignettes			,
R4	% and	X in 100	Pie chart		Verbal descriptors,		% combined with	Comparators
(2009)	frequency				Comparators		verbal descriptors	-
R5	Frequency	NF				NNT	Frequency	NNT
(2009)								
R6	% and	1 in XXX			Verbal descriptors		Frequency	No clear
(2010)	frequency							conclusion
R7	% and	1 in XXX and X in					%	Frequency
(2008)	frequency	XXX						
R8	%				Verbal descriptors		% combined with	No clear
(2007)							verbal descriptors	conclusion
R9			Bar chart	Icon arrays, flow chart			Bar chart	Icon arrays
(2008)								
R10				Icon arrays, ladder**	Comparators	NNT	Bar chart	Comparators
(2006)								
R11	Frequency	X in 1000		Ladder	Verbal descriptors,		Frequency	No clear
(2004)					Comparators			conclusion
R12	% and	1 in X			Vignettes		Frequency	%
(2008)	frequency							
R13			Bar chart, Line	Icon arrays, ladder		Survival curves	Bar chart	Icon arrays
(2001)			graph					
R14	%				Verbal descriptors		%	Verbal
(2008)								descriptors

Table A2: Characteristics of included empirical studies

ID (vear)	Type of risk communication format examined						Overall conclusion	
(jear)	Numerical format	Frequency format	Graphical risk	Pictorial risk	Qualitative format	Epidemiology	Most effective	Least effective
R15 (2011)	%					Prolongation of life	Prolongation of life	No clear conclusion
R16 (2013)					Comparators		Comparators	No clear conclusion
R17 (2009)	%			Icon arrays			%	Icon arrays
R18 (2011)	%		Bar chart		Verbal descriptors		Bar chart	No clear conclusion
R19 (2009)	Frequency	X in XXX and X in 100		Icon arrays			Frequency	Icon arrays
R20 (2011)	Frequency	X in 10,000, 1 in XXXX and XX in XXXX					Frequency (small denominator)	No clear conclusion
R21 (2011)	Frequency			Icon arrays			No clear conclusion	No clear conclusion
R22 (2009)	Frequency	1 in XXX			Verbal descriptors		Frequency combined with verbal descriptors	Verbal descriptors
R23 (2004)	% and frequency				Verbal descriptors		No clear conclusion	No clear conclusion
R24 (2004)	% and frequency	1 in X			Verbal descriptors		% and frequency	Verbal descriptors
R25 (2011)	% and frequency	X in 1000				NNT and prolongation of life	%	NNT
R26 (2004)	Frequency	1 in XXX and X in 10,000			Verbal descriptors		No clear conclusion	No clear conclusion
R27 (2000)	Frequency	X in 100		Icon arrays			No clear conclusion	No clear conclusion
R28 (2011)				Icon arrays			Icon arrays	No clear conclusion
R29 (2008)	%				Verbal descriptors		%	No clear conclusion
ID (vear)	Type of risk	communication format e	xamined				Overall conclusion	
---------------	---------------------	------------------------	---------------------------	--------------------------	-----------------------	-----------------	--------------------------	------------------------
(jeur)	Numerical format	Frequency format	Graphical risk	Pictorial risk	Qualitative format	Epidemiology	Most effective	Least effective
R30						ARR and RRR	ARR	RRR
(2011)								
R31	Frequency	X in 10 and X in					Frequency	No clear
(2012)		XXXX						conclusion
R32 (2009)	Ratios	1:X		Icon arrays, ladder***			No clear conclusion	No clear conclusion
R33	Frequency	X in 999		Ladder	Verbal descriptors,		Ladder	Frequency
(2009)					Comparators			
R34	Frequency	X in 10,000		Icon arrays	Vignette		Icon arrays, vignettes	Decimals
(2008)	and decimal			-	-			
R35	Frequency	1 in X			Verbal descriptors		Frequency	Verbal
(2009)								descriptors
R36	Frequency	1 in X		Ladder***			No clear conclusion	No clear
(2003)								conclusion
R37	% and	X in 100	Pie chart	Icon arrays			No clear conclusion	No clear
(2006)	frequency							conclusion
R38	Frequency	1 in X and X in XXX	Bar chart	Icon arrays			Icon arrays	Bar chart
(2012)			(horizontal and vertical)					
R39	Frequency	1 in X and 1.X	vertieal)	Icon arrays			Frequency	Ratios
(2009)	and ratios			icon arays			requency	Tutios
R40	% and	1 in X		Traffic light risk chart	Verbal descriptors	Survival curves	Traffic light risk chart	No clear
(2011)	frequency			6	1		6	conclusion
R41	% and	X in 100					%	No clear
(2011)	frequency							conclusion
R42	Frequency	1 in X					No clear conclusion	No clear
(2011)								conclusion
R43	Frequency	1 in X			Comparators		Frequency and	No clear
(2013)							comparator	conclusion
R44	%		Bar chart	Icon arrays, ladder***,			Traffic light risk chart	No clear
(2009)				traffic light risk chart				conclusion

ID (vear)	Type of risk o	communication format e	examined				Overall conclusion	
(yeur)	Numerical format	Frequency format	Graphical risk	Pictorial risk	Qualitative format	Epidemiology	Most effective	Least effective
R45	101 11110		Bar chart, pie	Icon arrays	10111111		Icon arrays	No clear
(2007)			chart				/	conclusion
R46	Frequency	X in 100	Bar chart	Icon arrays	Verbal descriptors,		Verbal descriptors	No clear
(2004)				-	vignettes			conclusion
R47	%		Bar chart, Line	Icon arrays, ladder**		Survival curves,	Bar chart	NNT
(2001)			graph			NNT and prolongation of life		
R48	% and		Bar chart	Icon arrays		protongation of me	No clear conclusion	No clear
(2006)	frequency			·				conclusion
R49	% and			Icon arrays		NNH	No clear conclusion	No clear
(2013)	frequency							conclusion
R50	% and	X in 10				NNT	Frequency	NNT
(2003)	frequency							
R51	% and	X in 100	Bar chart		Vignettes		Bar chart	No clear
(2011)	frequency				XX 1 1 1 1 .		0/ 11 1 11	conclusion
R52	%				Verbal descriptors		% combined with	%
(2003) D52	0/	V : 100		T	Varia 1 deceminterre			Ve als al
K55 (2010)	% and frequency	X in 100		Icon arrays	verbal descriptors		Icon arrays	descriptors
R54	% and	X in 100		Icon arrays	Verbal descriptors		Icon arrays	Verbal
(2010)	frequency	74 III 100		icon unuys	verbai desemptors		icon unuys	descriptors
R55	% and	X in 100		Icon arrays			No clear conclusion	No clear
(2008)	frequency			ý				conclusion
R56	%		Bar chart	Icon arrays			Icon arrays	Bar charts
(2004)				-			-	
R57	% and	1 in XX and 1:X	Pie chart		Verbal descriptors		Pie Charts	Verbal
(2009)	frequency							descriptors
	and ratio							
R58	% and	X in 100					%	Frequency
(2006)	trequency							NT 1
K59					Vignettes (traffic		No clear conclusion	No clear
(2002)					lights)			conclusion

ID (Type of risk o	communication format e	xamined				Overall conclusion	
(year)	Numerical format	Frequency format	Graphical risk	Pictorial risk	Qualitative format	Epidemiology	Most effective	Least effective
R60 (2012)				Icon arrays, ladder****			Icon arrays	Risk ladder
R61 (2008)	% and frequency	X in 1000		Icon arrays	Vignettes		Icon arrays	No clear conclusion
R62 (2009)	% and frequency	X in 1000		Icon arrays	Vignettes		No clear conclusion	No clear conclusion
R63 (2009)	%				Verbal descriptors		No clear conclusion	Verbal descriptors
R64 (2007)	Frequency	X in 100		Icon arrays	Vignettes	Survival curves	No clear conclusion	No clear conclusion
R65 (2008)	% and frequency	X in 100 and X in 1000		Icon arrays	Vignettes		Icon arrays	No clear conclusion

*Risk ladder here was referred to as a risk continuum in paper **Risk ladder here refers to a risk thermometer ***Risk ladder here refers to a Paling Perspective Scale ****Risk ladder here refers to the magnifying glass graphic

Effectiveness of risk communication format

To assess the effectiveness of the risk communication format, most studies (n= 53, 82%) asked the study participants to identify the tool they preferred by using a survey-based approach. Qualitative methods of enquiry such as interviews or focus groups were used either alone or in combination with quantitative surveys in 14 studies (22%). The surveys used to assess format effectiveness covered a variety of topics including: self-reported preferences; self-reported behaviour changes; and ability to recall risk information. Almost a third of studies measured respondents' performance on a numeracy test (n=21, 32%), often using the same three questions (see Figure A3). Details of the different measures of 'effectiveness' used by each study can be found in Table A3.

The difference in measures of effectiveness and the difference in risk communication formats compared in each study made it difficult to compare across studies or to analyse the evidence. Table A2 shows which formats the authors concluded were most effective and which were least. Many studies (n=32, 49%) did not achieve consensus on the best risk communication format. Of the numerical methods of presenting information, 11 (17%) studies found percentages to be the most effective method and 12 (18%) studies found frequency to be the most effective method.

Ten studies (15%) reported that qualitative descriptions of risk did not aid the understanding of the risk information, and this method appeared to be one of the least successful communication formats. Some studies did conclude that qualitative descriptors were effective when combined with percentages or frequencies (see R4, R8, R22, R52). In terms of graphical approaches to communicate risk, six (9%) studies reported that bar charts were the most effective communication format (R9, R10, R13, R18, R47, R51). In terms of the pictorial communications of risk, ten studies (15%) reported icon arrays to be the most effective (although R9 and R19 found them inferior). Only two studies reported risk ladders to be the most useful method (R3, R33).

ID (year)	Questionnaire	Interview	Focus Group
R1 (2002)	\checkmark		
R2 (2004)	\checkmark		
R3(2012)	\checkmark		
R4 (2009)	\checkmark		
R5 (2009)	\checkmark		
R6 (2010)	\checkmark		
R7 (2008)	\checkmark		
R8 (2007)	\checkmark		
R9 (2008)	\checkmark		
R10 (2006)	\checkmark		
R11 (2004)	\checkmark		
R12 (2008)	\checkmark		
R13 (2001)		\checkmark	
R14 (2008)	\checkmark		
R15 (2011)	\checkmark		
R16 (2013)	\checkmark		
R17 (2009)	\checkmark		
R18 (2011)	\checkmark		
R19 (2009)	\checkmark		
R20 (2011)	\checkmark		
R21 (2011)			
R22 (2009)	\checkmark		
R23 (2004)	\checkmark		
R24 (2004)	\checkmark		
R25 (2011)		\checkmark	
R26 (2004)		\checkmark	
R27 (2000)	\checkmark		
R28 (2011)	\checkmark		
R29 (2008)		\checkmark	
R30 (2011)			\checkmark
R31 (2012)	\checkmark	\checkmark	
R32 (2009)	\checkmark	\checkmark	
R33 (2009)			\checkmark
R34 (2008)	\checkmark		
R35 (2009)	\checkmark		
R36 (2003)	\checkmark		
R37 (2006)			
R38 (2012)	\checkmark		
R39 (2009)	\checkmark		
R40 (2011)	\checkmark		
R41 (2011)	\checkmark		
R42 (2011)	\checkmark		
R43 (2013)		\checkmark	
R44 (2009)	\checkmark		

Table A3: Methods for evaluating the effectiveness of risk communication formats used in studies included in the systematic review of risk communication formats

ID (year)	Questionnaire	Interview	Focus Group	
R45 (2007)			\checkmark	
R46 (2004)		\checkmark	\checkmark	
R47 (2001)			\checkmark	
R48 (2006)	\checkmark			
R49 (2013)	\checkmark			
R50 (2003)	\checkmark			
R51 (2011)	\checkmark			
R52 (2003)	\checkmark			
R53 (2010)	\checkmark			
R54 (2010)	\checkmark			
R55 (2008)	\checkmark			
R56 (2004)		\checkmark		
R57 (2009)	\checkmark			
R58 (2006)	\checkmark			
R59 (2002)	\checkmark			
R60 (2012)		\checkmark		
R61	\checkmark			
(2008) R62 (2009)	\checkmark			
R63 (2009)	\checkmark			
R64 (2007)	\checkmark			
R65 (2008)	\checkmark			

Uncertainty

The review identified four studies which compared the communication of risk using point estimates with an uncertain range of estimates (an ambiguous risk). Three studies concluded that presenting an ambiguous risk heightened the individual's concern about the risk in an affect known as 'ambiguity aversion' (U1, U2, U4). U4 also found that providing a set of confidence intervals decreased the risk information's perceived credibility, even in a highly educated sample of respondents. The studies investigating uncertainty matched results in the empirical studies which also found introducing confidence intervals resulted in groups believing the risk was less trustworthy and vague (R47). Conversely, one study (U3) concluded there was no difference between perceived credibility of the risk information presented as a point estimates or a confidence interval, but did report that using either of these methods were superior to the presentation of no risk information at all.

Non-empirical studies

Although studies which empirically tested and compared different risk communication formats were of most interest, the search also identified a number of review and overview pieces. The nine reviews which had a systematic search strategy and compared risk communication formats were often very limited in their approach, comparing only a few formats at a time. For example, one study (S1) looked specifically at frequencies, percentages and epidemiologically terms only, and another (S2) limited their study to graphical communication. The other systematic reviews were limited in their applicability to other health areas, for example one study (S9) looked specifically at the communication of cardiovascular risk.

This structured review also identified twenty-one overview or summary papers. These papers were not deemed sufficient for inclusions in a 'review of review' as most were based on opinion, rather than the results of empirical tests. A third of papers (n=7) had only one author (O3, O11, O13, O14, O15, O18, O21) which could indicate the studies had a restricted viewpoint, possibly confined to one individual's experience.

Discussion

This structured review demonstrated a lack of consensus in the risk communication literature on the best risk communication format to use. This finding may relate to the vast amount of possible risk presentation formats and the variety of measures of effectiveness used. A number of studies also concluded that the preferred risk communication format was the one which changed people's behaviour, but this inference is problematic and care needs to be taken to ensure respondents' understanding is affected but their preferences are not. Therefore, with such heterogeneity and a lack of consistency in the results it is difficult to draw any substantive conclusions.

The different measures of effectiveness, different comparators and different respondent samples make it challenging to create meaningful comparisons between studies. The discussion here is based on the findings of the empirical studies identified in the review, whilst drawing on some of the opinion pieces. Each risk communication format will be discussed in terms of its advantages and disadvantages, and its potential usefulness in a DCE context.

All different methods are shown in Figure A2 in which the square boxes represent broad categories and the circles represent individual communication formats. The solid lines indicate links and the dotted lines show possible overlaps between categories.



Figure A2: Identified risk communication formats

Pictorial risk

The structured review identified a number of different pictorial formats which could aid the communication of risk. One format, a risk ladder involves presenting respondents with the risk of interest next to other higher and lower risks which they may be more familiar with. The ladder can use a log scale or magnifying glasses to incorporate very small and very big risks. Some studies identified in the review used colour coded 'traffic light' ladders or thermometers to communicate the risk too. There are standardised risk ladders, such as the Paling Perspective Scale which uses odds that people can relate to (such as the lottery or being struck by lightning) (Moore et al., 2008). Whilst the ladder was found useful in some empirical studies, it was most effective in low-numerate individuals (Keller et al., 2009). There was also debate about whether the comparative risks should be voluntary or imposed, and whether the scales should be linear or logarithmic (Lipkus & Hollands, 1999). Furthermore, the problem of formatting a risk ladder for use in a DCE choice task may prevent successful implementation of the risk communication format.

Icon arrays appeared to be the most well received pictorial communication format which performed best in the empirical studies. Some studies also suggested that icon arrays made it easier for individuals to compare risk magnitudes; fundamental in the trade-off tasks in a DCE. Literature generally supported using blocked icons rather than a scattered approach, as scattering icons appeared to indicate randomness and uncertainty rather than risk (Han et al., 2012).

Graphs and charts

Graphical communication of risk is nothing new; Florence Nightingale popularised the use of pie charts in the mid-19th century (Paling, 2003). An advantage of pie charts is that they allow individuals to make a part-whole comparison of the situation. However, the method did not perform well in the review of empirical studies with only one study (R57) identifying it as the best method (and this study used a sample of children). Pie charts also suffered with their ability to communicate small risks where a risk of less than 0.5% can become a very small sector of the image.

Bar charts were the best performing graphical method identified in the review, with many trials reporting successful interpretation of the risk information. One study included in the review found risk was twice as likely to be correctly interpreted when risk was presented as a bar graph, rather than as text containing a percentage (Sprague et al., 2012). Other studies found the horizontal bars took longer to interpret and were better for

communicating uncertainty rather than risk (McCaffery et al., 2012; Schapira et al., 2001). However, the format performed poorly for small numerators such as risks less than 1 in 100 or 1 in 1,000 (McCaffery et al., 2012).

Whilst some empirical studies in the review tested the use of line graphs, these papers generally concluded that this approach was most effective for communicating uncertainty rather than risk. For example, Schapira et al. (2001) used a line graph to show confidence intervals around risks of breast cancer mortality. The review identified a number of studies which had looked at ranges (rather than point estimates) and found them to be inferior in terms of participants' ability to comprehend and trust the information. Line charts appear to be have limited use in a DCE communicating point estimates.

Numerical formats

The structured review identified multiple numerical formats for describing risk information. Percentages are a traditional risk communication format which describe the proportion per hundred. However, for small risks, percentages can be lost with decimal places. Natural frequencies or fractions were another format identified by the structured review. However, frequencies were open to criticisms; such as the numerator and denominator biases. Alternatively, decimal places or odds ratios could be used, but neither of these were any more effective.

Epidemiological terms

Some studies used epidemiological terms such as NNT, NNH, RRR and ARR. The epidemiological terms did not perform well in the empirical studies identified by the structured review (and no studies identified this format as 'best' when compared to other formats). In addition, it is not clear how well these would perform in a DCE context and whether this would make choices between alternative more confusing.

Qualitative descriptors

Qualitative descriptions of risk, where risk is presented as a statement, were met with mixed success. The British National Formulary (BNF) provides guidelines on the appropriate descriptions of probabilities for adverse reactions to drugs shown in Table A4 (BNF, 2015). In cases where the method was found to be effective, it was often combined with numerical information (such as frequencies or percentages).

Table A1: Qualitative descriptions of risk

Qualitative Description	Risk
Very common	Greater than 1 in 10
Common	1in 10 to 1 in 100
Less Common	1 in 100 to 1 in 1,000
Rare	1 in 1,000 to 1 in 10,000
Very Rare	Less than 1 in 10,000

Source: BNF (2015)

Other studies have found that qualitative descriptors may indicate uncertainty and cause individuals to assign their own values to each statements (Schmidt, 2004). For DCEs, encouraging simplifying heuristics and respondents to make their own interpretation could induce further ambiguity. Presenting information as illustrated in Table 4.4 could introduce additional information for respondents to remember and could further increase the cognitive burden of the experiment. It may also present additional challenges in the analysis of choice data, for example, choosing an appropriate coding for the quantitative levels of the risk attribute to estimate value trade-offs (such as risk tolerability).

Uncertainty

In DCEs, risk is frequently presented as point estimates with respondents rarely informed about the certainty of the statistic (Harrison et al., 2014). For people who understand risk information, uncertainty may provide them with more useful information. However, the studies in this structured review found individuals were less likely to trust ambiguous risk information. Given the cognitive burden of even a simple DCE, where respondents are required to imagine themselves in a hypothetical scenario considering a health intervention they are likely to be unfamiliar with risk and so the addition of intervals might be an unnecessary complication. Also, the introduction of uncertainty has implications for the coding of levels in the analysis of the DCE data and, possibly, the underlying economic theories which support discrete choice modelling.

Experience and Perceptions

A key finding from the review was that the risk communication format should be tailored to the specific respondent sample. Tailoring is required to take into account the respondents' experience and prior perceptions, and how this may affect their understanding. This could be because the communication format makes the risk more personable, for example, respondents identifying with the figures in icon arrays and therefore realising the magnitude presented. However, it could also be important in explaining risks objectively to people who have miscalculated their own susceptibility based on family experience. In the context of DCEs this is particularly important as respondents should base their choices on the risk information presented rather than their own inferred probabilities.

Numeracy and mathematical competency

The studies identified in the review used a range of samples and measured understanding of risk in various ways: some studies used interviews; a smaller number of studies used ability to recall information; and others tested the formats using mathematical questions. Of the studies using mathematical questions to assess the numeracy skills of their sample (regardless of the outcome measures), there were some questions commonly occurring. The original questions by Schwartz et al. (1997) are shown in Figure A3. Although there were some slight variations across other studies in the actual terminology used to frame these numeracy questions, they were qualitatively similar to those suggested by Schwartz et al. (1997).

Figure A3: Standardised numeracy questions

- 1) Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even?
- 2) In the BIG BUCKS LOTTERY, the chances of winning a \$10.00 prize is 1%. What is your best guess about how many people would win a \$10.00 prize if 1,000 people each buy a single ticket to BIG BUCKS?
- 3) In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?

Source: Schwartz et al. (1997)

Evaluating risk communication formats

The challenge of evaluating the most effective risk communication format is the balance between achieving what is attractive and appealing, what changes behaviour through manipulation and what actually improves an individual's ability to reason with risk. As Table A3 shows, multiple methods were used to measures format effectiveness. In the case of DCEs, the risk communication format should improve the individual's ability to make an informed decision but this outcome is difficult to measure quantitatively. In the original proposal of work, this PhD intended to use qualitative interviews to appraise the framing methods identified in the review. However, the usefulness of qualitative interviews alone was viewed to be minimal given the multiple outcome measures and mismatching between self-reported appeal and objective numerical performance. Therefore proceeding with a DCE employing risk communication formats identified by the structured review was thought to be the most appropriate approach.

Strengths and limitations

This review identified a number of ways of communicating risk information. The search strategy was broad in its range of databases but used specific terms, and this could be seen as a limitation. However, the specific search strategy and strict exclusion criteria facilitated a rapid review of a large research area. A lack of second reviewer and extractor may have resulted in missed papers or risk communication formats. However, the number of new framing formats decreased considerably towards the end of the extraction process suggesting data saturation was achieved. There were also many studies which could not conclude which method performed best. No studies included in this review compared all the identified formats which prevent a solid conclusion on the 'best' or the 'worst' risk communication formats.

Conclusion

The results of this review uncovered that there is a fine line between providing sufficient information for respondents to understand the risk and overwhelming them with too much. Training materials for respondents should be presented at a level that is accessible to all but should not be so elementary that respondents will ignore it. A prescriptive approach to the DCE, with extensive piloting and revisions, should therefore be conducted to ensure materials are appropriate for respondents.

The identified empirical studies also highlighted the importance of assessing numeracy skills on a respondent's ability to interpret probabilistic information. Preformed ideas about perceived risk based on experience were also identified as a challenge to effective communication of risk information. In understanding how respondents interpret risk in a DCE, numeracy, experience and perceptions could be useful covariates to collect in supplementary questions in a DCE study.

The review identified icon arrays as the risk communication format with the most empirical support which would make it an appropriate comparator to current practice in DCE literature which are percentages only (Harrison et al., 2014). However, no empirical study investigated the use of icon arrays in a trade-off task so it is unclear whether they would be superior in a DCE context.

Bibliography

BNF (2015). Adverse reactions to drugs. British National Formulary. 62.

Han, P., Klein, W., Killam, B., Lehman, T., Massett, H. and Freedman, A.N. (2012). Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. Patient education and counseling. 86 (1). 106–113.

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J. and Payne, K. (2014). Risk as an attribute in discrete choice experiments: a critical review. The patient. 7 (2). 151–70.

Hutton, J.L. (2009). Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. British journal of haematology. 146 (1). 27–30.

Keller, C., Siegrist, M. and Visschers, V. (2009). Effect of risk ladder format on risk perception in high- and low-numerate individuals. Risk Analysis. 29 (9). 1255–64.

Lipkus, I. and Hollands, J. (1999). The visual communication of risk. JNCI monographs. 25. 149–163.

McCaffery, K., Dixon, A., Hayen, A., Jansen, J., Smith, S. and Simpson, J.M. (2012). The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. Medical Decision Making. 32 (4). 532–44.

Moore, R.A., Derry, S., McQuay, H.J. and Paling, J. (2008). What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. Arthritis research & therapy. 10 (1). 1–16.

Paling, J. (2003). Strategies to help patients understand risks. BMJ British Medical Journal. 745–748.

Schapira, M.M., Nattinger, A.B. and McHorney, C.A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. Medical Decision Making. 21 (6). 459–467.

Schmidt, M. (2004). Investigating risk perception: a short introduction. In: Loss of agrobiodiversity in Vavilov centers, with a special focus on the risks of genetically modified organisms (GMOs). Vienna: Schmidit PhD Thesis, 1–16.

Schwartz, L., Woloshin, S., Black, W. and Welch, H.G. (1997). The role of numeracy in understanding the benefit of screening mammography. Annals of internal medicine. 127 (11).

1023-1028.

Sprague, D., Russo, J., Lavallie, D.L. and Buchwald, D.S. (2012). Influence of framing and graphic format on comprehension of risk information among American Indian tribal college students. Journal of Cancer Education. 27. 752–8.

Appendix 3.2: List of studies included in the structured review of risk communication formats

ID Full reference

- R1 Berry DC, Michas IC, and Bersellini E., (2003). Communicating information about medication: the benefits of making it personal. Psychology and Health,18(1), 127–39.
- R2 Berry D, Raynor T, Knapp P, and Bersellini E., (2004). Over the counter medicines and the need for immediate action, a further evaluation of European Commission recommended wordings for communicating risk. Patient Education and Counseling,53, 129–34.
- R3 Brewer NT, Richman AR, DeFrank JT, Reyna VF, and Carey LA., (2012). Improving communication of breast cancer recurrence risk. Breast Cancer Research and Treatment, 133, 553–61.
- R4 Brewer NT, Tzeng JP, Lillie SE, Edwards AS, Peppercorn JM, and Rimer BK., (2009). Health literacy and cancer risk perception: implications for genomic risk communication. Medical Decision Making, 29, 157–66.
- R5 Carling CLL, Kristoffersen DT, Montori VM, Herrin J, Schünemann HJ, Treweek S, et al., (2009). The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. PLoS Medicine, 6(8), 1–10.
- R6 Cheung YB, Wee HL, Thumboo J, Goh C, Pietrobon R, Toh HC, et al., (2010). Risk communication in clinical trials: a cognitive experiment and a survey. BMC Medical Informatics and Decision Making, 10(55), 1–9.
- R7 Cuite CL, Weinstein ND, Emmons K, and Colditz G., (2008). A test of numeric formats for communicating risk probabilities. Medical Decision Making, 28, 377–84.
- R8 Davis JJ., (2007). Consumers' preferences for the communication of risk information in drug advertising. Health Affairs, 26(3), 863–70.
- R9 Dolan J, and Iadarola S., (2008). Risk communication formats for low probability events, an exploratory study of patient preferences. BMC medical informatics and decision making, 8, 14.
- R10 Edwards A, Thomas R, Williams R, Ellner AL, Brown P, and Elwyn G., (2006). Presenting risk information to people with diabetes: evaluating effects and preferences for different formats by a web-based randomised controlled trial. Patient Education and Counseling, 63, 336–49.
- R11 Emmons KM, Wong MEI, Puleo E, Fletcher R, and Colditz G., (2010). Tailored computerbased cancer risk communication : correcting colorectal cancer risk perception. Journal of Health Communication, 9(2), 127–41.
- R12 Fair AKI, Murray PG, Thomas A, and Cobain MR., (2008). Using hypothetical data to assess the effect of numerical format and context on the perception of coronary heart disease risk. Am. Journal of Health Promotion, 22(4), 291–7.
- R13 Fortin JM, Hirota LK, Bond BE, O'Connor AM, and Col NF., (2001). Identifying patient preferences for communicating risk estimates: a descriptive pilot study. BMC medical informatics and decision making, 1, 2.
- R14 France J, Keen C, and Bowyer S., (2008). Communicating risk to emergency department patients with chest pain. Emergency Medicine Journal, 25(5), 276–8.
- R15 Galesic M, and Garcia-Retamero R., (2011). Communicating consequences of risky behaviors: life expectancy versus risk of disease. Patient Education and Counseling. Elsevier Ireland Ltd., 82, 30–5.
- R16 Galesic M, and Garcia-retamero R., (2013). Using analogies to communicate information about health risks. Applied Cognitive Psychology, 27, 33–42.

- R17 Galesic M, and Garcia-Retamero R, Gigerenzer G., (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. Health Psychology, 28(2), 210–6.
- R18 Garcia-Retamero R, and Cokely ET., (2011). Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. Journal of Experimental Psychology, 17(3), 270–87.
- R19 Garcia-Retamero R, and Galesic M., (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. American Journal of Public Health, 99(12), 2196–202.
- R20 Garcia-Retamero R, and Galesic M., (2011). Using plausible group sizes to communicate information about medical risks. Patient Education and Counseling. Elsevier Ireland Ltd,,84(2), 245–50.
- R21 Garcia-Retamero R, and Dhami MK., (2011). Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. Health Expectations, 14(Suppl 1), 46–57.
- R22 Graham PH, Martin RM, and Browne LH., (2009). Communicating breast cancer treatment complication risks: when words are likely to fail. Asia-Pacific Journal of Clinical Oncology, 5, 193–9.
- R23 Gurmankin AD, Baron J, and Armstrong K., (2004). Intended message versus message received in hypothetical physician risk communications: exploring the gap. Risk Analysis, 24(5), 1337–47.
- R24 Gurmankin AD, Baron J, and Armstrong K., (2004). The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. Medical Decision Making, 24, 265–71.
- R25 Gyrd-Hansen D, Halvorsen P, Nexøe J, Nielsen J, Støvring H, and Kristiansen I., (2011). Joint and separate evaluation of risk reduction: impact on sensitivity to risk reduction magnitude in the context of 4 different risk information formats. Medical Decision Making, 31, E1–10.
- R26 Hembroff L a, Holmes-Rovner M, and Wills CE., (2004). Treatment decision-making and the form of risk communication: results of a factorial survey. BMC Medical Informatics and Decision Making, 16,4(20), 1–9.
- R27 Henneman L, Oosterwijk JC, van Asperen CJ, Menko FH, Ockhuysen-Vermey CF, Kostense PJ, et al., (2013). The effectiveness of a graphical presentation in addition to a frequency format in the context of familial breast cancer risk communication: a multicenter controlled trial. BMC medical informatics and decision making, 13, 55.
- R28 Hess R, and Siegrist M., (2011). Risk communication with pictographs: the role of numeracy and graph processing. Judgment and Decision Making, 6(3), 263–74.
- R29 Hilton ZN, Carter AM, Harris GT, and Sharpe AJB., (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. Journal of Interpersonal Violence, 23, 171–88.
- R30 Hudson B, Toop L, Mangin D, and Pearson J., (2011). Risk communication methods in hip fracture prevention: a randomised trial in primary care. British Journal of General Practice, 469–76.
- R31 Ilic D, Murphy K, and Green S. Risk communication and prostate cancer: identifying which summary statistics are best understood by men. American Journal of Men's Health., (2012). ,6(6), 497–504.
- R32 Keller C, and Siegrist M., (2009). Effect of risk communication formats on risk perception depending on numeracy. Medical Decision Making, 29, 483–90.

- R33 Keller C, Siegrist M, and Visschers V. (2009). Effect of risk ladder format on risk perception in high- and low-numerate individuals. Risk Analysis, 29(9), 1255–64.
- R34 Kennedy A, Glasser J, Covello V, and Gust D., (2008). Development of vaccine risk communication messages using risk comparisons and mathematical modeling. Journal of Health Communication, 13(8), 793–807.
- R35 Knapp P, Raynor DK, Woolf E, Gardner PH, Carrigan N, and McMillan B., (2009). Communicating the risk of side effects to patients: an evaluation of UK regulatory recommendations. Drug Safety, 32(10), 837–49.
- R36 Lee DH, and Mehta MD., (2003). Evaluation of a visual risk communication tool, effects on knowledge and perception of blood transfusion risk. Transfusion,43, 779–87.
- R37 Lobb E, Butow PN, Moore A, Barratt A, Tucker K, Gaff C, et al., (2006). Development of a communication aid to facilitate risk communication in consultations with unaffected women from high risk breast cancer families, a pilot study. Journal of Genetic Counseling, 15(5), 393–405.
- R38 McCaffery KJ, Dixon A, Hayen A, Jansen J, Smith S, and Simpson JM., (2012). The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy, a randomized experimental study. Medical Decision Making, 32, 532–44.
- R39 Miron-Shatz T, Hanoch Y, Graef D, and Sagi M., (2009). Presentation format affects comprehension and risk assessment: the case of prenatal screening. Journal of Health Communication, 14(5), 439–50.
- R40 Neuner-Jehle S, Senn O, Wegwarth O, Rosemann T, and Steurer J., (2011). How do family physicians communicate about cardiovascular risk? Frequencies and determinants of different communication formats. BMC Family Practice. BioMed Central Ltd,,12(15), 1–9.
- R41 Peters E, Hart PS, and Fraenkel L., (2011). Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. Medical Decision Making, 31, 432–6.
- R42 Pighin S, and Savadori L., (2011). Overcoming number numbness in prenatal risk communication. Prenatal Diagnosis, 31, 809–13.
- R43 Pighin S, Savadori L, Barilli E, Rumiati R, Bonalumi S, Ferrari M, et al., (2013). Using comparison scenarios to improve prenatal risk communication. Medical Decision Making, 33, 48–58.
- R44 Price HC, Dudley C, Barrow B, Kennedy I, Griffin SJ, and Holman RR., (2009). Use of focus groups to develop methods to communicate cardiovascular disease risk and potential for risk reduction to people with type 2 diabetes. Family Practice, 26, 351–8.
- R45 Price M, Cameron R, and Butow P., (2007). Communicating risk information: the influence of graphical display format on quantitative information perception-Accuracy, comprehension and preferences. Patient Education and Counseling, 69, 121–8.
- R46 Royak-Schaler R, Blocker DE, Yali AM, Bynoe M, Briant KJ, and Smith S., (2004). Breast and colorectal cancer risk communication approaches with low-income African-American and Hispanic women, implications for healthcare providers. Journal of the National Medical Association, 96(5), 598–608.
- R47 Schapira MM, Nattinger AB, and McHorney CA., (2008). Frequency or probability? A qualitative study of risk communication formats used in health care. Medical Decision Making. 1,21(6), 459–67.
- R48 Schapira MM, Nattinger AB, and McAuliffe TL., (2006). The influence of graphic format on breast cancer risk communication. Journal of Health Communication, 11(6), 569–82.

- R49 Shepperd JA, Lipkus IM, Sanderson SC, Mcbride CM, Neill SCO, and Docherty S., (2013). Testing Different Communication Formats on Responses to Imagined Risk of Having Versus Missing the GSTM1 Gene. Journal of Health Communication, 18, 124–37.
- R50 Sheridan SL, Pignone MP, and Lewis CL., (2003). A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. Journal of General Internal Medicine, 18, 884–92.
- R51 Sprague D, Russo JE, Lavallie DL, and Buchwald DS., (2012). Influence of framing and graphic format on comprehension of risk information among American Indian tribal college students. Journal of Cancer Education, 27, 752–8.
- R52 Steiner M., (2003). Understanding risk, a randomized controlled trial of communicating contraceptive effectiveness. Obstetrics and Gynecology, 102(4), 709–17.
- R53 Tait AR, Voepel-Lewis T, Zikmund-Fisher BJ, and Fagerlin A., (2010). Presenting research risks and benefits to parents, does format matter? Anesthesia and Analgesia, 111(3), 718–23.
- R54 Tait AR, Voepel-Lewis T, Zikmund-Fisher B, and Fagerlin A., (2011). The effect of format on parents' understanding of the risks and benefits of clinical research: a comparison between text, tables, and graphics. Journal of Health Communication, 15(5), 487–501.
- R55 Timmermans DRM, Ockhuysen-Vermey CF, and Henneman L. 2008 Presenting health risk information in different formats, the effect on participants' cognitive and emotional evaluation and decisions. Patient Education and Counseling, 73, 443–7.
- R56 Timmermans D, Molewijk B, Stiggelbout A, and Kievit J., (2004). Different formats for communicating surgical risks to patients and the effect on choice of treatment. Patient Education and Counseling, 54, 255–63.
- R57 Ulph F, Townsend E, and Glazebrook C., (2009). How should risk be communicated to children, a cross-sectional study comparing different formats of probability information. BMC Medical Informatics and Decision Making, 9, 26.
- R58 Waters E, Weinstein ND, Colditz G, and Emmons K., (2006). Formats for improving risk communication in medical tradeoff decisions. Journal of Health Communication, 11(2), 167–82.
- R59 Wogalter MS, Smith-Jackson TL, Mills BJ, and Paine CS., (2002). The effects of print format in direct-to-consumer prescription drug advertisements on risk knowledge and preference. Therapeutic Innovation and Regulatory Science. 1,36, 693–705.
- R60 Wong ST, Pérez-Stable EJ, Kim SE, Gregorich SE, Sawaya GF, Walsh JME, et al., (2012). Using visual displays to communicate risk of cancer to women from diverse race/ethnic backgrounds. Patient Education and Counseling, 87, 327–35.
- R61 Wright a J, Takeichi C, Whitwell SCL, Hankins M, and Marteau TM., (2008). The impact of genetic testing for Crohn's disease, risk magnitude and graphical format on motivation to stop smoking, an experimental analogue study. Clinical Genetics, 73, 306–14.
- R62 Wright AJ, Whitwell SCL, Takeichi C, Hankins M, and Marteau TM., (2009). The impact of numeracy on reactions to different graphic risk presentation formats, An experimental analogue study. British Journal of Health Psychology, 14, 107–25.
- R63 Young S, and Oppenheimer D., (2009). Effect of communication strategy on personal risk perception and treatment adherence intentions. Psychology Health and Medicine, 14(4), 430–42.
- R64 Zikmund-Fisher BJ, Smith DM, Ubel P, and Fagerlin A., (2007). Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Medical Decision Making, 27, 663–71.

- R65 Zikmund-Fisher BJ, Ubel P, Smith DM, Derry H, McClure JB, Stark A, et al. 2008 Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. Patient Education and Counseling, 73, 209–14.
- U1 Han, P. K. J., Klein, W. M. P., Lehman, T., Killam, B., Massett, H., and Freedman, A. N. (2011). The communication of uncertainty regarding individualized cancer risk estimates: effects and influential factors. Medical Decision Making, 31(2), 354–366.
- U2 Han, P., and Klein, W. (2009). Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. Medical decision making 29(3), 391–403.
- U3 Lipkus, I., Klein, W., and Rimer, B. (2001). Communicating breast cancer risks to women using different formats. Cancer Epidemiology Biomarkers and Prevention, 895–898.
- U4 Longman, T., Turner, R. M., King, M., and McCaffery, K. J. (2012). The effects of communicating uncertainty in quantitative health risk estimates. Patient education and counseling, 89(2), 252–9.
- O1 Ahmed, H., Naik, G., Willoughby, H. and Edwards, A., (2012). Communicating risk. BMJ British Medical Journal, 344(e3996),1–7.
- O2 Anderson, E.E. and Iltis, A.S., (2008). Assessing and improving research participants' understanding of risk: potential lessons from the literature on physician-patient risk communication. Journal of empirical research on human research ethics, JERHRE, 3(3), 27–37.
- O3 Baglin, T., (2009). Communicating benefit and risk. British journal of haematology, 146, 31–3.
- O4 Barrett, B. and McKenna, P., (2011). Communicating benefits and risks of screening for prostate, colon, and breast cancer. Family Medicine, 43(4), 248–253.
- O5 Civan, A., Doctor, J. and Wolf, F., (2005). What makes a good format: Frameworks for evaluating the effect of graphic risk formats on consumers' risk-related behavior. AMIA Annual Symposium, p.927.
- O6 Conover, E. and Polifka, J.E., (2011). The art and science of teratogen risk communication. American journal of medical genetics. Part C, Seminars in medical genetics, 157, 227–233.
- O7 Crowson, C.S., Therneau, T.M., Matteson, E.L. and Gabriel, S.E., (2007). Primer, Demystifying risk--understanding and communicating medical risks. Nature clinical practice, 3(3), 181–187.
- O8 Edwards, A. and Elwyn, G., (2001). Understanding risk and lessons for clinical risk communication about treatment preferences. Quality in Health Care, 10(Suppl I), i9–i13.
- O9 Fagerlin, A., Zikmund-Fisher, B.J. and Ubel, P, (2011). Helping patients decide: ten steps to better risk communication. Journal of the National Cancer Institute, 103(19), 1436–43.
- O10 Fagerlin, A., Ubel, P., Smith, D. and Zikmund-Fisher, B.J., (2007). Making numbers matter: Present and future research in risk communication. American Journal of Health Behvaiour, 31(Supp 1), S47–S56.
- O11 Finucane, M., (2008). Emotion, affect, and risk communication with older adults, challenges and opportunities. Journal of Risk Research, 11(8), 983–997.
- O12 Garcia-Retamero, R., Okan, Y. and Cokely, E.T., (2012). Using visual aids to improve communication of risks about health: a review. The Scientific World Journal 2012, 1–10.
- O13 Gordon-Lubitz, R., (2003). Risk communication: problems of presentation and understanding. JAMA, 289(1), p.95.
- O14 Lewiecki, E.M., (2011). The role of risk communication in the care of osteoporosis. Current osteoporosis reports, 9, 141–8.

- O15 Picano, E., (2000). Informed consent and communication of risk from escape from a communication inferno. British medical journal, 329, 849–851.
- O16 Roach, P. and Marrero, D., (2005). A critical dialogue: communicating with type 2 diabetes patients about cardiovascular risk. Vascular health and risk management, 1(4), 301–307.
- O17 Slavin, D., Troy Tucker, W. and Ferson, S., (2008). A frequency/consequence-based technique for visualizing and communicating uncertainty and perception of risk. Annals of the New York Academy of Sciences, 1128, 63–77.
- O18 Thompson, K., (2002). Variability and uncertainty meet risk management and risk communication. Risk Analysis, (6), 1–11.
- O19 Thomson, R., Edwards, A. and Grey, J., (2005). Risk communication in the clinical consultation. Clinical Medicine, 5, 465–469.
- O20 Visschers, V., Wiedemann, P.M., Gutscher, H., Kurzenhauser, S., Seidl, R., Jardine, C. and Timmermans, D., (2012). Affect-inducing risk communication: current knowledge and future directions. Journal of Risk Research, 15(3), 257–271.
- O21 Zikmund-Fisher, B.J., (2013). The right tool is what they need, not what we have, a taxonomy of appropriate levels of precision in patient risk communication. Medical care research and review, MCRR, 70(1 Suppl), p.37S–49S.
- S1 Akl, E., Oxman, A., Herrin, J., Vist, G., Terrenato, I., Sperati, F., Costiniuk, C., Blank, D. and Schunemann, H., (2012). Using alternative statistical formats for presenting risks and risk reductions (Review). The Cochrane Library, (3), 1–87.
- S2 Ancker, J., Senathirajah, Y., Kukafka, R. and Starren, J., (2006). Design features of graphs in health risk communication, a systematic review. Journal of the American Medical Informatics Association, 13(9), 608–619.
- S3 Burkiewicz, J., Vesta, K. and Hume, A., (2008). Improving effectiveness in communicating risk to patients. The Consultant Pharmacist, 23(1), 37–43.
- S4 Ghosh, A.K. and Ghosh, K., (2005). Translating evidence-based information into effective risk communication: Current challenges and opportunities. Journal of Laboratory and Clinical Medicine, 145(4), 171–180.
- S5 Lipkus, I.M., (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Medical Decision Making, 27(5), 696– 713.
- S6 Moore, R.A., Derry, S., McQuay, H.J. and Paling, J., (2008). What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. Arthritis research and therapy, 10(1), 1–16.
- S7 Naik, G., Ahmed, H. and Edwards, A., (2012). Communicating risk to patients and the public. British Journal of General Practice, 213–216.
- S8 Visschers, V.H.M., Meertens, R.M., Passchier, W.W.F. and de Vries, N.N.K., (2009). Probability information in risk communication: a review of the research literature. Risk analysis, 29(2), 267–87.
- S9 Waldron, C.-A., van der Weijden, T., Ludt, S., Gallacher, J. and Elwyn, G., (2011). What are effective strategies to communicate cardiovascular risk information to patients? A systematic review. Patient education and counseling, 82, 169–81.

Appendix 3.3: The published paper containing the results of the structured review

Risk as an Attribute in Discrete Choice Experiments: A Systematic Review of the Literature

Mark Harrison, Dan Rigby, Caroline Vass, Terry Flynn, Jordan Louviere & Katherine Payne

The Patient - Patient-Centered Outcomes Research

ISSN 1178-1653 Volume 7 Number 2

Patient (2014) 7:151-170 DOI 10.1007/s40271-014-0048-1





SYSTEMATIC REVIEW

Risk as an Attribute in Discrete Choice Experiments: A Systematic Review of the Literature

Mark Harrison · Dan Rigby · Caroline Vass · Terry Flynn · Jordan Louviere · Katherine Payne

Published online: 25 February 2014 © Springer International Publishing Switzerland 2014

Abstract

Background Discrete choice experiments (DCEs) are used to elicit preferences of current and future patients and healthcare professionals about how they value different aspects of healthcare. Risk is an integral part of most healthcare decisions. Despite the use of risk attributes in DCEs consistently being highlighted as an area for further research, current methods of incorporating risk attributes in DCEs have not been reviewed explicitly.

Objectives This study aimed to systematically identify published healthcare DCEs that incorporated a risk attribute, summarise and appraise methods used to present and

Electronic supplementary material The online version of this article (doi:10.1007/s40271-014-0048-1) contains supplementary material, which is available to authorized users.

M. Harrison - C. Vass - K. Payne (🖂) Centre for Health Economics, Institute of Population Health, 4th floor, Jean McFarlane Building, University Place, The University of Manchester, Oxford Road, Manchester M13 9PL, UK e-mail: katherine.payne@manchester.ac.uk

M. Harrison

e-mail: mark.harrison@manchester.ac.uk

C. Vass

e-mail: caroline.vass@postgrad.manchester.ac.uk

D. Rigby

Economics, School of Social Sciences, Arthur Lewis Building, The University of Manchester, Manchester, UK e-mail: dan.rigby@manchester.ac.uk

T. Flynn · J. Louviere

Institute for Choice, UniSA Business School, Level 13, 140 Arthur Street, North Sydney, NSW 2061, Australia e-mail: terry.flynn@unisa.edu.au

J. Louviere

e-mail: jordan.louviere@unisa.edu.au

analyse risk attributes, and recommend best practice regarding including, analysing and transparently reporting the methodology supporting risk attributes in future DCEs. *Data Sources* The Web of Science, MEDLINE, EM-BASE, PsycINFO and Econlit databases were searched on 18 April 2013 for DCEs that included a risk attribute published since 1995, and on 23 April 2013 to identify studies assessing risk communication in the general (non-DCE) health literature.

Study Eligibility Criteria Healthcare-related DCEs with a risk attribute mentioned or suggested in the title/abstract were obtained and retained in the final review if a risk attribute meeting our definition was included.

Study Appraisal and Synthesis Methods Extracted data were tabulated and critically appraised to summarise the quality of reporting, and the format, presentation and interpretation of the risk attribute were summarised.

Results This review identified 117 healthcare DCEs that incorporated at least one risk attribute. Whilst there was some evidence of good practice incorporated into the presentation of risk attributes, little evidence was found that developing methods and recommendations from other disciplines about effective methods and validation of risk communication were systematically applied to DCEs. In general, the reviewed DCE studies did not thoroughly report the methodology supporting the explanation of risk in training materials, the impact of framing risk, or exploring the validity of risk communication.

Limitations The primary limitation of this review was that the methods underlying presentation, format and analysis of risk attributes could only be appraised to the extent that they were reported.

Conclusions Improvements in reporting and transparency of risk presentation from conception to the analysis of DCEs are needed. To define best practice, further research is needed to test how the process of communicating risk affects the way in which people value risk attributes in DCEs.

Key Points for Decision Makers

Discrete choice experiments (DCEs) in health often use risk attributes to value risk:benefit trade-offs and potentially the perceived value of risk reduction

People struggle to understand concepts of risk and probability, and process and distinguish between differences in risk

Improvements in reporting and transparency of risk presentation from conception to the analysis of DCEs are needed

Recommendations about risk presentation from other fields need to be adopted by researchers developing DCEs

Experimental studies to understand the impact of different methods of risk presentation/ communication are urgently needed

1 Background

Risk is a ubiquitous component of healthcare and decisions made by patients and healthcare professionals about commencing or continuing treatments that are shaped by perceived benefits and risks. Discrete choice experiments (DCEs) allow trade-offs between health and non-health outcomes and processes to be valued. Including a cost attribute allows indirect elicitation of willingness-to-pay (WTP) values [1]. In theory, including a risk attribute allows valuation of the balance between risk and benefits (risk:benefit trade-offs). In the decision theory literature, risks (prospects with known probabilities associated with possible outcomes) are distinguished from uncertainties (prospects with unknown probabilities of possible outcomes) [2]. However, significant potential difficulties in presenting risk information are recognised and if probability information is not well understood, not well presented or simply ignored then this limits the validity, usefulness and applicability of the results. A substantial, growing literature has explored the way humans perceive and interpret risk and probability information [3], and an extensive literature about improving risk communication is emerging [4-6].

Presentation of risk information has been studied in the contingent valuation (CVM) literature. Findings that WTP for risk reductions from CVM were inconsistent with economic theory prompted methodological research into risk communication, closer attention to risk communication methods, and rigorous checks of validity [3, 7]. The recommendations were that CVM could provide theoretically valid estimates of WTP for risk reduction if certain methods were applied and tested for validity against standard economic theory; WTP should be positively associated with a magnitude of risk reduction (or change), proportional to the size of risk reduction, and when the baseline risk is small the effect of the baseline risk (and changes in this risk) on WTP should be small [3, 7]. Methods of enhancing risk communication in CVM studies to achieve results consistent with economic theory centred on using visual aids (icon/dot arrays, risk ladders using a logarithmic scale, analogies) to set the risk in context [3, 7].

Detailed reviews of the application and methodological basis of DCEs in healthcare consistently highlight including risk as an attribute as an important area for further research, perhaps indicating a lack of progress in this area [1, 8-10]. To date, no review has explicitly summarised the methods used to incorporate risk as an attribute in DCEs. The aim of this paper was to supplement previous reviews and recommendations [11, 12] by identifying published DCE studies in healthcare that incorporated risk as an attribute and appraising the methods underlying presentation, format and analysis of risk attributes. The objective is not to criticise the authors of current healthcare DCEs, nor the methods used to date, but to summarise the current methodology, highlight potential areas that might be informed by developments in other fields, and make recommendations to improve risk communication in healthcare DCEs. The key objective is to provide a narrative review summarising the current state of incorporating risk in healthcare DCEs and to move towards recommendations on best practice for future DCEs or aspects of risk presentation in DCEs requiring further methodological exploration. This key objective was met by also completing a rapid review of the risk communication literature.

2 Methods

This study used two reviews: (1) a systematic review to identify relevant published work related to valuing a healthcare intervention using a DCE with risk included as an attribute; and (2) a rapid review to identify examples of good practice in risk communication in the general (non-DCE) health literature.

 Systematic Review of Risk in Discrete Choice Experiments (DCEs)

Risk attributes in DCEs in healthcare potentially aim to reflect the uncertain consequences of treatments. Typically, risk may be defined as being knowable (i.e. the different potential outcomes have a quantifiable probability of occurring), whereas uncertainty cannot be quantified [13]. This review takes an inclusive approach to the definition of risk, consistent with the decision theory literature, where risk attributes are defined as those presenting quantitative estimates of the probability associated with an uncertain outcome (either unwanted consequences or forgone benefits). However, we also include DCEs using qualitative descriptions of the likelihood of the unwanted prospect (or harms) to avoid potentially excluding relevant studies.

The electronic search strategy (Table 1) was derived from several sources, including two recent reviews in the area [1, 10], and a trained librarian (Ingram M, personal communication). Studies including a risk attribute were identified at the manual screening stage. The following databases were searched (on 18 April 2013): Web of Science, MEDLINE, EMBASE, PsycINFO and Econlit. The search strategy was validated by checking that references from the two recent reviews and other studies known to the team were captured. Previous systematic reviews of DCEs in healthcare identified two DCEs published before 1995, neither of which included a risk attribute [1]. Therefore, the time horizon was restricted to studies published since 1995.

All abstracts were screened independently to identify DCEs with risk attributes by three reviewers (MH and KP, as well as Stuart Wright), who met to discuss any disagreement. The inclusion criteria were healthcare related, DCEs (respondents make a clear choice between two or more alternative technologies/services comprising several attributes), risk—or an attribute consistent with our inclusive definition of a risk attribute—is mentioned in the title/ abstract. Health risks associated with non-healthcare products (e.g. food) were excluded. Conference abstracts and other examples of the non-peer reviewed literature, non-English articles and opinion/review/protocols/conference abstracts were excluded. Following the initial screening of

Table 1 Search terms

- 1. part-worth utility*.mp.
- paired comparison*.mp.
- 3. pairwise choice*.mp.
- 4. stated preference.mp.
- 5. exp Choice Behavior/
- (discrete adj choice\$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
- (conjoint adj (analys\$ or measurement\$ or study or studies or choice\$)).mp. [mp=title, original title, abstract, name of
- substance word, subject heading word, unique identifier]
- 8. 1 or 2 or 3 or 4 or 6 or 7
- 9. limit 8 to (humans and yr="1995 -Current")

abstracts, full copies of papers were obtained and retained in the final review if a risk attribute meeting our definition was included. Papers that framed the risk in certain terms (e.g. present/absent) were excluded.

Extracted data were tabulated. A published framework for critically appraising DCEs was used to summarise the quality of reporting in the DCEs [8]. In addition, the format, presentation and interpretation of the risk attribute were summarised using guidance from the CVM literature [7]. The review focussed on collating the following: background information/explanation of risk presented to respondents; framing of risk (description of risk attributes and format used to communicate changes in risk); stated assumptions about risk; analysis and interpretation of risk attributes (checks of linearity of risk preferences, sensitivity to scope, or risk attribute; calculating marginal rates of substitution [MRS] or WTP for risk reduction).

2.2 Rapid Review of Risk Communication Tools

Given the substantial and diverse literature that describes and/or evaluates different tools for the effective communication of risk in healthcare, a focussed rapid review was conducted to identify approaches to healthcare risk communication. The review was focussed by identifying studies that (1) tested any risk communication tool that aimed to aid people's understanding of quantitative information; (2) were related to a healthcare risk; and (3) were published in peer-reviewed journals from January 2000 onwards and written in English. Studies that looked at public health risks, such as nationwide vaccines, and studies that did not use written communication, such as videos or games, were excluded. The search strategy was devised with assistance from a trained librarian (Ingram M, personal communication) and was validated by checking that references from key studies known to the team were captured. The search strategy included the terms Title = (risk*) AND Title = (communicat* or format*) AND Topic = (aid* or present* or display*). The following databases were searched (on 23 April 2013): Web of Science, MEDLINE, EMBASE, PsycINFO and Econlit,

An initial title and abstract screen to select relevant studies was conducted. If studies could not be rejected with certainty, the whole article was retrieved and reviewed. After retrieving all the full papers, the studies were initially categorised into four categories: empirical studies testing risk communication tools; systematic review of risk communication tools; papers that provided an overview; and studies that looked specifically at communicating uncertainty. Data extraction focussed on describing the risk communication tool and whether a definitive conclusion was drawn regarding the relative effectiveness of each tool. These data were then summarised in a narrative review.

3 Results

3.1 Systematic Review of Risk in DCEs

Overall, 117 papers were identified and included. Figure 1 summarises the study identification and inclusion process. Few studies published before 2002 included a risk attribute, but thereafter the number increased to over 20 per year in 2011 and 2012 (Fig. 2). As a proportion of the total number of DCEs conducted [10], the growth in those with a risk component has been less dramatic. The included papers covered a range of healthcare settings, most commonly cancer (n = 27,23 %). Most DCEs investigated preferences for pharmacological treatments (n = 70, 60 %), followed by screening and diagnostics (n = 15, 13%). Fifty-nine (50 %) studies reviewed included a monetary attribute and, of these, most (n = 43, 73 %) calculated WTP (usually n = 40 [68 %] for risk reduction). Two (2 %) studies calculated the value of a statistical life [14, 15]. The mean number of attributes in the DCEs was six (range 3-12) and a mean of two (range 1-7) risk attributes were included. In three (3 %) DCEs, all attributes were risk attributes [16-18].

The results of the review are categorised into aspects of DCEs, presented in the order in which the reader would



Fig. 1 Flow of included/excluded studies. DCE discrete choice experiments

encounter each aspect when designing, framing, administering, analysing and interpreting DCEs. Each aspect is presented and reviewed in turn.

3.1.1 Explanatory and Background Training Materials

The reporting of background training materials presented was generally insufficient to assess the nature of explanatory material offered to respondents. Seventy-one (61 %) of the 117 identified studies reported providing background training information to participants, but only 27 (23 %) described this information in any detail. When described, background material included detail about the intervention or procedure for completing the DCE, although it was unclear whether any background on risk was provided. Three (3 %) studies provided full access to the entire survey materials in an online supplementary file [19-21], whilst others summarised information provided to respondents in a table [22] or written descriptions of multiple sections which helped respondents understand their risk attitude. A further two (2 %) studies reported providing a risk tutorial to enable respondents to interpret risk [23, 24].

3.1.2 Choice Question Format

Respondents were given an opt-out option with the choice sets in 41 (35 %) studies. Seventeen (15 %) studies offered a status quo opt-out, and 11 (9 %) defined the risk associated with this option to the respondent in the choice set, either as population risk values [19, 25–33] or the respondent's current choice and risk calculated using information collected during earlier sections of the questionnaire [23]. Five (4 %) studies allowed the respondents to consider their own current situation [14, 34–37], for example their child's current risk [36].

3.1.3 Communication of Risk Attributes

3.1.3.1 Presentation Risk attributes were mostly presented quantitatively (n = 72, 62%), using frequencies (n = 23, 32%), percentages (n = 27, 38%) or a combination of both (n = 22, 31%). In 13 (11\%) studies risks were presented with varying denominators, although none presented risks of equal magnitude using different denominators (ratio bias). Six (5\%) studies varied denominators for a risk within the same attribute [26, 35, 38–41]; for example comparing a 1 in 5 with a 1 in 10 chance [40]. However, these estimates were supplemented with either qualitative descriptions (e.g. 'minimal risk [1 in 150,000]' [39]), percentages [40], or both (e.g. 'moderate (1 in 5 or 20\%)' [38]). Five (4\%) studies used varying denominators between attributes [42–46], including one that presented two different side effects using numbers of Fig. 2 Increase in number of DCEs with a risk attribute (2013 excluded as literature search conducted on 18 April 2013). DCE discrete choice experiments



episodes but with a varying timeframe for occurrence (one in months, one in years) [44], and a further seven (6 %) used different denominators within and between attributes [26, 40, 47–51].

Nineteen (16 %) studies used a purely qualitative approach, suggesting that these attributes should be viewed as 'uncertain' instead of 'risky'; three (3 %) studies in haemophilia took an approach of describing risk in the context of existing (and assumed understood) products, e.g. risk of viral infection from human plasma-derived or recombinant products [52, 53], or current level of risk [54].

Twenty six (22 %) studies combined qualitative and quantitative methods to present risk. In some DCEs, qualitative and quantitative methods were used to describe different risk attributes, and in others within the same attribute, seemingly to aid communication (e.g. 'low, 1 in 20 (5 %)') [47]. Descriptions used for risks varied between studies; for example, risks considerably lower than 1 % were described as 'minimal' (1 in 150,000) [39] and 'very small' (1 in 10,000) [55], a risk of around 1 % was described as 'very low' and 10 % was described as 'low' [38], whereas others use 'low' for a risk <10 % [56]. Elsewhere, risks in the region of 20 % were described as 'moderate' [38] and 'medium' [56], and risks beyond 30 % were described as 'high' [56]. Where studies used purely qualitative terms such as 'small', the risk magnitudes supporting these descriptions were unclear [57, 58].

Absolute risks (n = 85, 73 %) were presented more often than relative risk (n = 8, 7 %), although five (4 %) studies included both [14, 27, 59–61]. Only two of eight studies using relative risk alone gave a baseline/underlying risk estimate to place the difference/change in overall risk in context [15, 33].

Risk attributes were virtually always presented as point estimates without variability or a range around each defined attribute level. One exception presented a range of values (e.g. <10, 10–30, >30 % [56]), although it is unclear whether this represents variability or an interval description. Others used risk levels ('<1 %' and 'about 5 %') [62]), and thresholds (≥ 2 , ≥ 5 or ≥ 10 in 100 patients) [63]. Other quantitative studies qualified the point estimates using words like 'almost' or 'about' to convey a lack of certainty [64, 65]. In one study, uncertainty related to two quantitative risk attributes (bone or kidney damage) was introduced using a qualitative attribute describing whether they could be successfully treated [66].

Visual aids were used to support communication of risk attribute information in 27 (23 %) studies. The most common visual risk presentation methods were risk icon/ dot arrays (n = 25, 21 %). Other methods included bar charts (n = 3, 3 %), pictograms, and risk ladders that represent the risk in the DCE alongside other, more familiar, risks (Table 2). Photographs were also used to help describe risk attributes [29, 49, 55].

3.1.3.2 Framing Risk was most often framed negatively as a consequence of treatment or an intervention (n = 102, 87%) rather than positively as something to be avoided (n = 26, 22%). Eleven (9\%) studies included a mix of positively and negatively framed risk attributes. Negative framing of risk most often concerned drug-related events or reactions, death, or adverse events/complications following treatment. Positively framed attributes concerned risk reductions and adverse events/complications avoided. Howard and Salkeld [45] explored the impact of positive or negative framing of three attributes related to the detection

Table 2	Summary of presentation and testing (of risk attribute						
Year	Author (ref.)	Sample	Risk	Absolute/relative risk	Graphical risk	Linearity in risk tested	Sensitivity to scope/scale	WTP for risk reduction
Output of	alises which considerates for							
CHURCH		1		•		;	;	;
2004	Ratcliffe et al. [50]	TA	Freq. and %	AR		Yes	ž	ź
2005	Lloyd et al. [40]	РТ	Freq. and %	AR		Yes	°Ž	Yes
2006	Ashcroft et al. [74]	Ħ	Freq. and %	AR		No	No No	No No
2007	Seston et al. [142]	PT	Freq. and %	AR		No	°Ž	°Ž
2007	Johnson et al. [78]	PT	Freq. and %	AR	Array and ladder	Yes	Yes	°Ž
2008	Lloyd et al. [47]	PT	Freq. and %	AR		No	°Ž	Yes
2009	Hauber et al. [63]	PT	Freq. and %	AR	Аптау	No	°Ž	°N N
2009	Johnson et al. [79]	PT	Freq. and %	AR	Array and ladder	Yes	Yes	No.
2010	Essens et al. [42]	PUB	Freq. and %	AR		No	°Ž	Yes
2010	Hauber et al. [143]	PT	Freq. and %	AR	Аптау	Yes	°Ž	°N N
2010	Swinburn et al. [46]	PT	Freq. and %	AR		No	°Ž	°N N
2010	Hauber et al. [144]	PT	Freq. and %	AR	Array	No	°Ž	°Ž
2011	van Houtven et al. [138]	PT	Freq. and %	AR	Array and ladder	Yes	Yes	°N N
2011	Mohamed et al. [145]	PT	Freq. and %	AR	Array	Yes	°Ž	°N N
2011	Pereira et al. [51]	DM	Freq. and %	AR		No	°Ž	Yes
2011	Hauber et al. [146]	ΡT	Freq. and %	AR	Array	No	Ŷ	°N N
2011	Guo et al. [25]	PT	Freq. and %	AR		No	°Ž	No.
2012	Kauf et al. [24]	PT	Freq. and %	AR	Array	Yes	°X	°N N
2012	Hodgkins et al. [147]	PT	Freq. and %	AR		Yes	°Ž	Yes
2012	Arden et al. [148]	Ηb	Freq. and %	AR	Array	No	°Ž	°N N
2013	Lathia et al. [149]	PT	Freq. and %	AR		Yes	°Ž	Yes
2013	Hauber et al. [150]	PT	Freq. and %	AR	Аптау	No	°Ž	No.
2001	Shackley et al. [71]	PT	Freq.	AR		No	°Ž	°N N
2002	Hall et al. [48]	PUB	Freq.	AR		No	Ŷ	°N N
2003	Salkeld et al. [43]	PUB	Freq.	AR		No	Ŷ	°N N
2004	Aristides et al. [151]	ΡT	Freq.	AR		No	°X	Yes
2008	McTaggart-Cowan et al. [44]	ΡT	Freq.	AR		Yes	^o X	Yes
2008	Tinetti et al. [16]	ΡT	Freq.	AR	Array	No	°N N	°N N
2008	Tinetti et al. [18]	ΡT	Freq.	AR	Array	No	^o N	°X
2009	Guimaraes et al. [152]	ΡT	Freq.	AR		No	Ŷ	Yes
2009	Howard et al. [45]	PT	Freq.	AR		No	No	Yes
2010	de Bekker-Grob et al. [49]	PT	Freq.	AR	Pictogram, bar chart	No	No	No No
2011	Oteng et al. [19]	PUB	Freq.	AR		No	°N No	Yes
2011	Sweeting et al. [20]	ΡT	Freq.	AR		No	Ŷ	Yes
2011	Scalone et al. [26]	PT	Freq.	AR		No	No	Yes

M. Harrison et al.

Table 2	continued							
Year	Author (ref.)	Sample	Risk	Absolute/relative risk	Graphical risk	Linearity in risk tested	Sensitivity to scope/scale	WTP for risk reduction
2011	Howard et al. [153]	PT	Freq.	AR		Yes	Ŷ	Ŷ
2011	Faggioli et al. [154]	PTAIP	Freq.	AR		No	No.	°X
2012	Laba et al. [22]	PUB	Freq.	AR		No	No.	Ŷ
2012	Bridges et al. [72]	PUB	Freq.	AR		No	No.	Ŷ
2012	Gidengil et al. [155]	Parent	Freq.	AR	Array	No	No.	Yes
2012	Pignone et al. [21]	PUB	Freq.	AR		No	No	^o Z
2013	Pignone et al. [17]	PUB	Freq.	AR		No	^o N	Ŷ
2013	de Bekker-Grob et al. [27]	PUB	Freq.	AR and RR ^{BL}		Yes	^o N	Yes
2013	Augustovski et al. [156]	PT	Freq.	AR	Array	Yes	No.	Yes
2013	Vlennix et al. [28]	PT	Freq.	AR		No	No.	Ŷ
1998	Bryan et al. [157]	PT	\$	AR		No	No.	Ŷ
2002	Bryan et al. [158]	PT	\$	AR		Yes	Yes	Ŷ
2002	Telser and Zweifel [81]	PUB	\$	RR		Yes	Yes	Yes
2004	Weston and Fitzgerald [62]	PUB	\$	AR		No	No.	Yes
2004	Bishop et al. [159]	PT	8	AR		No	^o N	Ŷ
2006	Lewis et al. [160]	PTAHP	\$	AR		No	^o N	Ŷ
2006	Lewis et al. [161]	đH	\$	AR		No	^o N	Ŷ
2007	Telser and Zweifel [82]	PUB	\$	RR		Yes	Yes	Yes
2007	Walzer [36]	PT	8	AR		No	^o N	Ŷ
2008	de Bekker-Grob et al. [162]	PT	8	RR		No	^o N	Yes
2009	Bunge et al. [163]	PT	8	AR		No	^o N	Ŷ
2009	Watson et al. [164]	PT	8	AR		No	No.	Ŷ
2009	de Bekker-Grob et al. [165]	PT	8	RR		No	^o N	Ŷ
2009	Kruijshaar et al. [166]	PT	8	AR		No	No.	Ŷ
2010	Hol et al. [167]	PT/PUB	8	AR	Аптау	No	^o N	Ŷ
2010	Torbica and Fattore [59]	đH	8	AR and RR ^{BL}		No	^o N	Ŷ
2010	Nayaradou et al. [60]	PUB	8	AR and RR		No	^o N	Ŷ
2010	Damen et al. [29]	Ы	8	AR		No	No No	Ŷ
2011	Wirotsko et al. [168]	PT	\$	AR		No	^o N	Ŷ
2011	Flood et al. [169]	Parent	\$	AR		No	^o N	Ŷ
2011	Schmeider et al. [65]	PT	8	AR		No	No.	°Ž
2012	Regier et al. [30]	Parent/HP	8	AR	Апау	No	No.	Yes
2012	Marti [31]	PT	8	AR		Yes	No.	Yes
2012	Damman et al. [170]	PUB	88	AR		Yes	No.	Ŷ
2012	Tong et al. [171]	PT	8	AR		No	No.	Ŷ
2012	Sung et al. [172]	PT	8	AR	Array	No	No.	Ŷ
2013	Borri et al. [23]	PUB	\$	AR		No	9 <mark>0</mark>	Yes

Table 2	continued							
Year	Author (ref.)	Sample	Risk	Absolute/relative risk	Graphical risk	Linearity in risk tested	Sensitivity to scope/scale	WTP for risk reduction
Qualitari	ive risk attribute/s							
2005	Mantovani et al. [54]	PTAHP	Description			No	No.	No.
2008	Lee et al. [52]	Ē	Description and Y/N			No	No.	No.
2008	Scalone et al. [53]	PTAIP	X/N			No	^o N	Yes
2008	Muhlbacher et al. [70]	ΡT	Description			No	No.	No.
2010	Marang-van de Mheen et al. [173]	PT	Description			No	No.	No.
2011	Watson et al. [174]	PUB	Description			Yes	No No	Ŷ
2011	Bijkenga et al. [175]	PUB	Description			No	No.	Yes
2011	Bijkenga et al. [176]	PUB	Description			No	No.	Yes
2011	Palumbo et al. [177]	ΡT	Description			No	No.	Yes
2011	Muhlbacher and Nubling [178]	PTAIP	Description			Yes	No.	Ŷ
2012	Tinelli et al. [32]	ΡT	Description			No	No.	Yes
2012	Manjunath et al. [179]	ΡT	Description			Yes	No.	Yes
2012	Kinsler et al. [180]	ΡT	Description			No	^o X	Ŷ
2012	King et al. [181]	ΡT	Description			Yes	No.	Ŷ
2012	Burnett et al. [182]	Parent	Description			Yes	^o X	⁰ ۲
2012	Bridges et al. [183]	ΡT	Description			Yes	No.	No.
2012	Lee et al. [184]	PUB	Description			No	No.	Ŷ
2013	Wittink et al. [185]	PUB	Description			No	No.	No.
2013	Zimmermann et al. [186]	ΡT	Description			Yes	^o X	No.
Quantita	tive and qualitative risk attributels							
2007	Lloyd et al. [38]	ΡT	Freq. and %	AR		No	No.	Yes
2009	Hauber et al. [66]	ΡT	Freq. and %	AR	Аптау	Yes	Yes	No.
2010	Johnson et al. [80]	ΡT	Freq. and %	AR	Аптау	Yes	No.	No.
2010	Cameron et al. [14]	PUB	Freq. and %	AR and RR ^{BL}		Yes	Yes	Yes
2011	Mohamed et al. [187]	£	Freq. and %	AR	Аптау	Yes	No.	No.
2005	Tsuge et al. [15]	PUB	Freq.	RR ^{BL}	Аптау	Yes	Yes	Yes
2006	Espelid et al. [35]	PTAIP	Freq.	AR	Photograph	Yes	^o X	⁰ ۲
2007	Ossa et al. [39]	PUB	Freq.	AR		Yes	Yes	Yes
2009	Petrou and McIntosh [188]	PT	Freq.	AR		No	No.	Yes
2011	Shafey et al. [41]	PTAIP	Freq.	AR		Yes	No.	No.
2011	Fiebig et al. [35]	PUB	Freq.	AR		Yes	No.	Yes
2013	Sadique et al. [189]	Parent	Freq.	AR		No	⁰ N	Yes
2002	Aristides et al. [190]	ΡT	8	AR		No	No.	Yes
2004	Watson et al. [191]	PUB	8	AR		No	^o X	Yes
2005	Lee et al. [56]	ΡT	88	AR		Yes	^o X	No.
2005	Fraenkel et al. [192]	PUB	¢ł.	AR	Bar chart	No	No.	No.

Year	Author (ref.)	Sample	Risk	Absolute/relative risk	Graphical risk	Linearity in risk tested	Sensitivity to scope/scale	WTP for risk reduction
2006	Fraenkel et al. [193]	PUB	æ	AR	Bar chart	No	9 <mark>X</mark>	No.
2006	Johnson et al. [33]	PUB	k	RR ^{BL}	Array	Yes	Yes	Yes
2007	Goto et al. [194]	Ы	k	RR		No	° <mark>X</mark>	^o X
2009	Griffith et al. [61]	PT.	Ŷ	AR and RR ^{BL}		No	°N N	No.
2009	Eberth et al. [57]	Ы	k	AR		No	No.	Yes
2010	van Dam et al. [58]	PT/PUB	Ŷ	AR	Array	No	° <mark>X</mark>	°Ž
2011	Goto et al. [73]	PUB	Ŷ	RR		No	° <mark>X</mark>	°Ž
2011	Schaarschmidt et al. [64]	Ы	Ŷ	AR		No	° <mark>X</mark>	Ŷ
2012	Doyle et al. [34]	Ы	₿ŝ	AR		No	° <mark>X</mark>	Yes
2012	Chancellor et al. [37]	PTAIP	Ŷ	AR		Yes	Ŷ	¢ <mark>N</mark>

and treatment of cancer. People seemed more willing to risk potential harms and had a higher WTP to detect a cancer than to reduce the chance of missing a cancer, suggesting evidence of framing effects. This is consistent with previous findings that people were more willing to pay for additional lives 'saved' rather than a reduction in lives 'lost' [67], and variance heterogeneity whereby people have different levels of certainty for losses and gains (see below) [68, 69].

The timeframe over which the risky 'event' should be considered by the respondent was stated explicitly in 63 (54 %) studies. Fifteen (13 %) studies provided information on risk latency, the delay between the decision and experiencing the 'risk', ranging from immediate to 10 years [15, 57, 70]. In 11 (9 %) studies the risk appeared to be immediate due to the nature of the intervention and risk (e.g. during/following surgery [56, 71, 72]. The remaining studies provided risk timeframes but did not specify any latency.

3.1.4 The Study Sample

Study samples most often comprised patients (n = 66, 56 %), parents or carers of patients (n = 4, 3 %), or healthcare professionals (n = 6, 5 %), suggesting most preferences were ex-post. Ex-ante preferences also were elicited from the public (n = 28, 24 %). Twelve (10 %) studies used multiple respondent types to compare patient preferences with other groups to better understand shared decision making or ex-post with ex-ante preferences.

3.1.5 Inclusion of Supplementary Questions

Most DCEs included supplementary sections seeking to identify respondent characteristics or knowledge of the intervention. None reported including questions to assess numeracy or literacy. Based on the reporting of the studies, respondent risk behaviour was not generally explored in the DCEs. Only Tsuge et al. [15] reported exploring risk attitudes, focusing on respondents' perceptions of risks (questions to gauge responsibility for the risks used in the DCE, and risk avoidance behaviour). Goto et al. [73] assessed the knowledge that respondents had about associations between smoking and a range of diseases. One study reported free text responses on the decision-making process made by clinicians whilst completing the DCE; clinicians felt that they focussed on longerterm risks whilst patients focussed on short-term benefits [74].

3.1.6 Analysis of Choice Data

Most studies (n = 53, 45%) assumed a linear additive indirect utility function, implying that only additive main effects matter and were estimated. However, 25 (21 %) studies stated that interaction terms were included to allow multiplicative effects. Interactions can be used to estimate two-way and higher order interactions between attributes. Ten (9%) studies interacted two risk attributes, e.g. the size of risk reduction with the type [15]. Two (2%) DCEs interacted two risk attributes (one mild, one serious AE) in a design allowing all two-way interactions between attributes to be estimated [48, 49]. Neither found a significant interaction but highlight the need to ensure face validity of risk attribute interactions so that significant interactions can be explained meaningfully.

Sensitivity to scope (or scale) of risk/risk reductions or testing for concave increases in WTP (diminishing marginal returns to scale) with increasing benefits (e.g. risk reduction) [75, 76] was reported to have been evaluated in 11 (9 %) studies. The terms 'scale' and 'scope' tend to be used interchangeably [77] but some studies draw distinctions between sensitivity to the size of the benefit (scale) [e.g. WTP for a 50 % compared with a 25 % risk reduction] with scope (e.g. comparing WTP for a 25 % risk reduction of heart disease with a 25 % risk reduction from heart disease and diabetes [75]). Several studies tested the rate at which preferences/WTP increase (i.e. proportional or concave), although a stronger test would be whether preferences are strictly proportional to the risk level. Thus, tests of sensitivity to scope overlap with tests of linearity in risk. We found evidence in 36 (31 %) studies of linearity in risk being explored, most commonly using dummy coding to account for risks/risk reductions of unequal size with interpretation of the proportionality of coefficients/MRS [50, 55] and/or graphical presentation of moving between the levels of the risk attribute. Several studies used piecewise-linear extrapolation to estimate preferences inbetween discrete risk attribute levels [63, 78-80]. Other approaches included coding attributes for differences between levels, using quadratic terms to identify non-linear relationships, polynomial approximations to test the proportionality of WTP for risk reductions with the magnitudes of risk reduction [15], and statistical tests of linearity of coefficients [81, 82].

3.2 Review of Risk Communication Tools

The search of all databases revealed 1,207 possible hits. After the removal of non-peer-reviewed material and duplicates, 390 titles and abstracts were reviewed. In total, 215 full articles were retrieved and 99 papers were included in the final review (electronic supplementary material [ESM] Appendix 1).

Of the 99 included studies, 65 (66 %) were empirical tests of risk communication tools, 21 (21 %) were overviews of the risk literature but did not report a search strategy, nine (9 %) were systematic reviews with a structured search strategy, and four (4 %) looked at presenting uncertainty (for example, the use of confidence intervals as opposed to point estimates). ESM Appendix 2 summarises the 65 studies that reported an empirical test of a risk communication tool. These 65 studies most commonly explored numerical presentation of risk information such as percentages, fractions, decimals or ratios (n = 55, 85 %). Qualitative descriptions and verbal analogies were investigated in 34 studies (52 %). Other methods included pictorial presentations, such as icon arrays (n = 29, 45 %), risk ladders (n = 10, 15 %) and graphical information (n = 15, 23 %). The criteria to quantify the 'effectiveness' of the tool as a risk communication strategy was measured using a numeracy test (n = 21, 32 %) or by asking the study participants which tool they preferred using qualitative methods of enquiry or a non-numerical questionnaire (n = 47, 72 %).

A number of studies stated clear conclusions that risk ladders, icon arrays and bar charts were the best way to communicate risk. Of the numerical methods of presenting information, 11 (17 %) studies found percentages to be the most effective method [6, 83–92] and 12 (18 %) studies found frequency to be the most effective method [93–104]. For graphical representations of risk, six (9 %) studies found bar charts to be the most effective communication method [105–110]. Ten studies reported that qualitative descriptions of risk did not aid people's understanding of the information, and this method was one of the least successful communication tools [83, 85, 98, 100, 111–116].

4 Discussion

The use of risk attributes in DCEs is common. Evidence from the identified reviews suggests that researchers conducting DCEs with risk attributes are often using the appropriate basic risk information (e.g. absolute risks with denominators), and visual aids such as icon/dot arrays. However, there was little available evidence that indicated that there was a consistent approach to communicating/framing risk information based on evidence from the risk communication literature in DCEs, to the extent as in related fields such as CVM [3, 7]. However, this may, in part, be due to a lack of reporting of methods used to support risk presentation because of journal word count restrictions.

Although there was a variety of (occasionally conflicting) conclusions to the studies included in the rapid review, some substantive conclusions could be drawn. The rapid review of risk communication methods found evidence supporting more sophisticated methods of presenting risk information through the use of graphical or pictorial images, the use of icon arrays and risk ladders. The use of icon arrays in DCEs with risk attributes is increasing over time.

The implications of the findings are now discussed under the headings under which elements of the inclusion

Table 3	Key	recommend	latione
1 apre 5	INC Y	recomment	auons

161

Aspect of DCE	Key recommendation(s)
Explanatory and background training materials	 Should be documented and available online (and maintained) in journals in which they appear. The quality of explanatory material should be subject to testing to the same extent as the statistical properties of the design and analysis
	Should define and place risk in context. The amount of explanatory/background information should be equivalent for all attributes to avoid overstating the importance of the risk attribute
	3. The numerical and risk literacy of respondents should be assessed during the piloting of the DCE. Strategies to overcome problems with numerical literacy should be used
	4. Appropriate visual aids, diagrams and other methods, such as information acceleration should be considered. The impact of the use of alternative means of risk communication should be tested formally
The choice question format	 Opt-out options may be required in DCEs to reflect real life situations. Researchers need to carefully define the research question and consider different ways of framing risk and uncertainty information in DCEs
	6. The meaning of status quo and no treatment opt-outs in terms of risk should be clear and conveyed to respondents where possible. If these are not known ex-ante, there is a need to elicit how respondents defined the underlying risk and incorporate that into the analysis
	Differences in error variance between researcher- and respondent-defined opt-out attribute levels should be tested
Framing and description of attributes	There is a need for greater interdisciplinary collaboration with experts from other fields such as psychology when framing and describing attributes to benefit from their expertise
	9. There is a need for empirical evidence to indicate whether good practice in presenting/communicating risk from the risk communication and CVM literatures should be adopted. There is a need for systematic testing that these methods of presenting risk information, which may change responses, improve the communication of risk
	10. The method of presenting risk should be validated using tests such as internal magnitude tests, tests of sensitivity to probability, and responses to changes in risk [3, 7]
	 A consistent format, including a consistent denominator, should be used when communicating risks and benefits [4, 5]
The study sample	12. Where possible, information on the respondents' knowledge of, and experiences with, the risk presented should be sought and included as potential determinants of preference heterogeneity in the analysis to aid interpretation of preferences
Data collection	13. Online data collection, which allows more flexible methods of risk presentation to be incorporated and developed, such as visual, audio and other technological aids, should be considered
Inclusion of supplementary questions	14. Supplementary questions should be used to better understand respondent characteristics such as attitude to risk, numeracy, and risk literacy that can affect trading behaviour and preferences for risk:benefit trade-offs
Analysis of choice data	15. Careful attention should be paid to the assumed functional form for the way that risk attributes enter into utility functions, and the design and analysis should be carefully planned to support the assumptions made
	16. Assumptions made about attitudes to risk and how risk has been interpreted should be tested formally at the analysis stage. It is important to consider whether there is evidence of preference heterogeneity, which may have been driven by attitudes to risk, numeracy and risk literacy

DCE discrete choice experiment

of risk within DCEs were characterised. Table 3 lists key recommendations relating to each section. Recommendations 3, 4 and 9 highlight areas for future research required before best practice can be defined; the remainder relate to the need for transparent reporting of study design and approaches to include risk.

4.1 Explanatory and Background Training Materials

Researchers designing DCEs need to ensure that respondents receive consistent, correctly communicated information which is then assumed to be used in making choices [117]. For respondents to make informed, reliable and incentive-compatible choices and trade-offs, particularly when trading off risk against non-risk attributes, they must be sufficiently interested to pay attention to details and understand the trade-offs presented [118, 119]. If risk is not effectively communicated to respondents, the result may be large variances around responses that preclude meaningful elicitation of preference information. The use of step-by-step descriptions or numerical data relating to risk may be useful as a training tool [4]; however, there was little evidence that this technique has been used. The potential for utilising this kind of approach is increasing as online or computerised administration allows more flexible presentation of information.

The evaluation of the quality of background or explanatory materials was restricted by the lack of availability of materials to readers of published DCEs, although encouragingly some recent studies published their entire surveys in online appendices [19–21].

4.2 Choice Question Format

The decision to include 'opt-outs' or 'no choice' options should be based upon the research question [11]. Expensive, risky, new alternatives or technologies are often associated with high proportions of non-choice, either because respondents are waiting to see if better options arise or because they have decided that they have no interest now or ever [120]. Where status quo options are allowed, they should be defined or respondents asked expost how they defined and used status quo in the DCE, otherwise it is difficult to know or understand what the 'opt-out' or 'no choice' actually means. Opt-out options were used in approximately one-third of studies reviewed, and status quo opt-outs were defined, particularly in studies published in the final 2 years of the review.

4.3 Communication of Risk Attributes

4.3.1 Presentation

The communication of risks and probabilities and the distinction between risks is problematic [121], more so when describing risks with low occurrence probabilities or small changes in risk, typical of healthcare interventions [3]. There was evidence of risks being presented with comparable numerators but varying denominators (denominator neglect), both within and between attributes, which may lead respondents to misinterpreting differences in magnitude between two levels or types of risk. Although there were no examples in the literature of ratio bias, it appears that the recommendations from the risk communication of using a consistent format, including a consistent denominator, for risks and benefits should be reiterated for researchers designing DCEs [4, 5].

A number of studies chose to describe risks qualitatively as an uncertainty, circumventing the problems associated with quantitative presentation of risk, but this may provide a different impression of the changes in risk presented. Qualitative description of risk provides scope for greater flexibility of description but may lead to variability, imprecision and a lack of consistency in how the magnitude of risk is interpreted. The review of risk communication tools suggested this was one of the least effective methods of risk communication. Furthermore, there was inconsistency in the terminology used to describe risks of a similar magnitude when qualitative descriptions were used to support quantitative estimates, and a lack of transparency relating to the likely magnitude of risk when a purely qualitative approach to risk (or uncertainty) presentation was used.

In line with recommendations from the risk communication literature, the identified DCEs were increasing using visual aids (such as icon/dot arrays) to support information communication. It is not clear whether different supplementary aids produce unintended consequences (i.e. demand artefacts). For example, dramatic, emotional photographs may dominate other information in the DCE [119], and the physical space required to incorporate photographs into attributes may draw the respondents' attention to that particular attribute.

4.3.2 Framing

One study investigated whether framing of risk as positive (e.g. a cancer detected) or negative (e.g. a cancer missed) found that positive framing can increase preferences for an attribute [45], replicating prior findings outside DCEs [122, 123]. However, as with the use of visual aids, it is unclear whether changes in preferences reflect improvements in the communication of risk. An alternative approach to framing risk is to present both positive and negative outcomes in the form 'out of 100 operations on people like you, we expect 95 to be successful and 5 to be unsuccessful' to avoid framing bias [124, 125]; no study reviewed reported utilising this approach. In addition, framing probabilities as relative or absolute risk reductions can affect respondent preferences and choices. Relative risk reductions may look larger than absolute risks, causing respondents to be more willing to choose treatments or avoid risks when information is presented using relative risks compared with absolute risks [4, 122, 126-128]. Reassuringly, risk information in DCEs was mainly described using absolute risk reductions, or both relative and absolute effects to give a complete picture of risk, the convention recommended by CONSORT [129].

Contextual factors, such as the timeframe, can influence the credibility of choices and the degree to which respondents believe the risks apply to them [130]. The timeframe for risks presented were only specified by around a third of studies in this review, although it is possible it was outlined in background materials. Recommendations from the risk communication literature emphasise the importance of specifying and then repeatedly reinforcing the time interval to ensure valid risk communication [131].

Testing the face and construct validity of results from DCEs is an important step in improving the communication of risk. DCE studies that checked the construct validity of risk preference results tended to be studies conducted in line with other recommendations of good practice from the risk communication and CVM literature, such as supplementing risk data with visual aids. The need for validation of and presentation of risk attributes suggests a role for greater interdisciplinary collaboration with experts from other fields such as psychology when framing and describing attributes. Tests of validity of risk preferences from DCEs associated with risk attributes, by comparing results with those expected on the basis of economic theory, is important for identifying the best ways to communicate risks to patients.

4.4 The Study Sample

Preferences from a range of study samples were included in the DCEs reviewed. It is important to distinguish ex-ante (prior to an event/anticipated) and ex-post (after the event/ experienced) preferences. Direct experience of a health condition may affect preferences, for example those who have not experienced a condition or event may over-estimate the dreadfulness of bad events (and underestimate the ability to adapt to changes in circumstances) [132], and levels of WTP for a life saved for a curative intervention (ex-post) may exceed those for prevention (ex-ante) [133]. Furthermore, if respondents do not treat a risk as applicable to them, they may use prior beliefs, experiences and the information to subjectively interpret risk information presented [134]. Risks in healthcare often have low occurrence probabilities, which pose a considerable problem, and although healthcare professionals should be experienced and knowledgeable about these risks, they may struggle to use probabilities to judge the level of risk [125]. There are also unique challenges to presenting risk which reflect patient decision making compared with other settings, such as the role of physicians in shared decision making.

Early recommendations of good practice in the CVM literature suggest separating out estimates of WTP by knowledge, relevance and immediacy, and understanding and belief [119]. Whilst the risk communication literature recommends that risk communication strategies may need to be targeted for each setting or audience [4, 124], Table 2 is indicative of a lack of variation in the way risk preferences are assessed in the DCE literature. Careful consideration of the context and consultation with methodologists from other disciplines may be beneficial when designing the optimal format for risk attributes.

4.5 Inclusion of Supplementary Questions

It would seem reasonable to understand the risk preferences of respondents when risk is included in a DCE. Current practice in DCEs seems to assume that respondents' risk attitudes are neutral, consistent with the cost-effectiveness analysis and decision analysis literature [135, 136]. Risk neutrality assumes people have linear monotonic utility functions in risk; however, it is possible that people may be risk averse or risk prone/seeking. Prospect theory predicts that people may be more risk averse when considering potential losses and risk prone/ seeking over gains [123].

Individual determinants of risk preferences, such as risk attitude, numeracy and health literacy, and experience were almost entirely overlooked in either the design or reporting of the identified DCEs. Although risk attitude is known to affect preferences [137], few studies reported attempting to assess respondent risk attitudes. However, there is little evidence that a respondent's attitude to one type of risk is consistent with their attitude toward bearing health risks, or between different kinds of health risks; for example, one study identified in this review showed that mortality risk tolerance varied across three kinds of mortality risks [138]. Most DCEs asked supplementary questions about respondent characteristics, but questions about knowledge of interventions, risk behaviours or numeracy were less common. The CVM literature recommends stratifying WTP analyses by respondent attitudes and ability and willingness to participate [119].

Numeracy affects how people interpret and understand quantitative risk information [6]. None of the reviewed DCEs reported measuring respondent numeracy and risk literacy. Only one study explicitly discussed methods to detect and overcome problems with numeracy, looking for signs that respondents appeared/expressed confusion or gave 'haphazard, inconsistent answers', and reported strategies to overcome problems with numerical literacy [18]. Although numeracy levels are not reported to influence negative/positive risk framing effects [6], step-by-step risk descriptions that include both positive and negative outcomes to reflect natural decisions and problem solving recommended to overcome framing effects [5] may also aid those with low numeracy [139]. Despite an apparent lack of focus on numerical literacy, the use of natural frequencies and icon arrays was prevalent and these formats are thought to aid low-numeracy groups [86, 124, 140].

4.6 Analysis of Choice Data

DCEs are based on Lancaster's economic theory of demand and random utility theory (RUT). Incorporating risk introduces another relevant theory, expected utility theory (EUT), concerning decision making under uncertainty. EUT views expected utility to equal the sum of the utilities of the various outcomes multiplied by the probability of each outcome occurring. RUT views the value a person places on a particular service or intervention as some function of the sum of its separate components or attributes, although RUT allows for incorporation of multiplicative effects through interactions and it is possible to test the assumptions of EUT. Thus, DCEs incorporating risk attributes require additional theoretical links, and DCE designs must support utility functions implied by the theoretical integration.

Most identified DCEs assumed that respondents evaluate risk linearly, incorporating risk attributes as simply as assumed for other attributes. If the indirect utility function is thought to be multiplicative in risk, this must be accounted for at both design and analysis stages; for example, using more general specifications that can capture functional forms for multiple risk attributes and/or implied interactions of these attributes. We found few DCEs that tried to do this (for example Johnson et al. [79, 138] and van Houtven et al. [79, 138]). One implication of more flexible design strategies is the need to consider larger designs, such as full factorials. Main effects designs are small because they deliberately confound the two-way and higher-order interactions with the main effects. If multiplicative functional forms are introduced, designs that properly separate the main effects from the interactions are needed (making the design larger than a main effects design alone). More general and flexible design strategies are discussed elsewhere [9].

We identified one DCE that reported that the weights respondents place on risks may not be linear but instead fluctuate with the range of probabilities [138], a departure from the assumptions of EUT, and expectations of proportionality of WTP or MRS to the size of risk reduction, but consistent with prospect theory. This work used nonlinear weighting techniques to explore the impact of risk:benefit trade-offs under a non-EUT framework.

5 Limitations

The primary limitation of this review is that the methods underlying presentation, format and analysis of risk attributes can only be appraised to the extent that they are reported. Guidelines for authors of peer-reviewed articles restrict manuscript length and may preclude detailed description of the development of surveys, wording and formatting, and other details such as preparing respondents to evaluate risk attributes. It may be that many studies included in this review have performed more rigorous and thorough development of risk attributes than is reported in publications. However, this limitation motivates the need for better reporting of methods, use of (and maintenance of) online appendices if necessary, supporting risk attributes as a key recommendation. More detailed reporting of this process would benefit researchers designing future DCEs and allow the quality of the background and explanatory material to be subject to appraisal to the same extent as the statistical properties of the design and analysis. This recommendation draws strong parallels with recommendations for improved reporting of qualitative methods used in developing and refining attributes for a DCE, which also cites lack of reporting conventions as a barrier to rigorous conduct, limiting the ability of researchers to draw on others' experience, with the consequence of 're-invention' and suppression of debate, which could inspire methodological development [141].

6 Conclusion

This study identified and reviewed a large body of literature that incorporated risk attributes in healthcare DCEs. There was some evidence that some examples of good practice in the presentation of risk, for example the use of absolute risks, and adoption of visual aids to support risk presentation were being adopted. However, little evidence was found that continuing methodological developments and recommendations from other fields about the need for effective risk communication were systematically being applied to DCEs. Specifically there was limited evidence found about use of training and background material to support risk communication and risk numeracy, and little consideration of the impact of information framing on preferences. Some evidence suggested that different formats were explored to try to help risk understanding and some examples where validity of risk communication was checked using tests of scope/scale and linearity. More generally, however, the reviewed DCE studies were not as methodologically thorough in risk presentation as the CVM field, where recommendations on minimal requirements for explaining risk in training materials, consideration of framing of risk in designs, and exploring scope/linearity in analyses exist. There seems a clear need for healthcare DCEs to adopt these recommendations. There also appears a real need for those conducting healthcare DCEs to engage with other disciplines to improve the design and presentation of risk attributes; there perhaps is not the amount of attention paid to the specific challenges in developing these attributes as needed. However, to define best practice there is also a need to move beyond statistical testing of the robustness of different methods of presentation towards testing the process of communicating risk in terms of whether people understand risks, which methods change the way in which people interpret risk information, and a framework for identifying which methods are best for the purpose of risk communication. This points to an urgent need for robust transdisciplinary methodological work supported by qualitative and quantitative analysis to inform
the best way of presenting and analysing risk attributes in DCEs.

Acknowledgments Professor Katherine Payne was part funded by a Research Councils UK (RCUK) Academic Fellowship between September 2007 and September 2012. No other sources of funding were used to conduct this study or prepare this manuscript. Caroline Vass is funded by a National Institute for Health Research (NIHR) School for Primary Care Research Studentship. Mark Harrison, Dan Rigby, Caroline Vass, Terry Flynn, Jordan Louviere, and Katherine Payne have no conflicts of interest that are directly relevant to the content of this article. The authors are grateful to Mary Ingram, Lawrence Library, Centre for Musculoskeletal Research, University of Manchester, for assistance in developing the electronic search strategies for both reviews, and to Stuart Wright, Manchester Centre for Health Economics, for double screening the abstract lists for inclusion in this review.

Contributions to authorship Mark Harrison, Katherine Payne, Terry Flynn, Jordan Louviere and Dan Rigby conceived the idea for this study and produced the protocol which informed the systematic search strategy. Mark Harrison and Katherine Payne applied the inclusion criteria and screening of abstracts, and extracted data from the included studies. Caroline Vass conducted the rapid review. All authors contributed to the writing of the manuscript. Katherine Payne acts as the overall guarantor.

References

- Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2003;2(1):55–64.
- Tversky A, Wakker P. Risk attitudes and decision weights. Econometrica. 1995;63(6):1255–80.
- Hammitt JK, Graham JD. Willingness to pay for health protection: inadequate sensitivity to probability? J Risk Uncertainty. 1999;18(1):33-62.
- Visschers VHM, Meertens RM, Passchier WWF, de Vries NNK. Probability information in risk communication: a review of the research literature. Risk Anal. 2009;29(2):267–87.
- Lipkus IM. Numeric, verbal, and visual formats of conveying health risk: suggested best practices and future recommendations. Med Decis Making. 2007;27(5):696–713.
- Peters E, Hart PS, Fraenkel L. Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. Med Decis Making. 2011;31(3):432–6.
- Corso PS, Hammitt JK, Graham JD. Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. J Risk Uncertainty. 2001;23(2):165–84.
- Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. Pharmacoeconomics. 2008;26(8):661–77.
- Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Econ Policy Law. 2009;4(4):527–46.
- de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.
- Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health. 2011;14(4): 403-13.

- Johnson FR, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health. 2013;16(1):3–13.
- Johansson P. Evaluating health risks: an economic approach. 1st ed. Cambridge: Cambridge University Press; 1995.
- Cameron TA, DeShazo JR, Johnson EH. The effect of children on adult demands for health-risk reductions. J Health Econ. 2010;29(3):364–76.
- Tsuge T, Kishimoto A, Takeuchi K. A choice experiment approach to the valuation of mortality. J Risk Uncertainty. 2005;31(1):73–95.
- Tinetti ME, McAvay GJ, Fried TR, Allore HG, Salmon JC, Foody JM, et al. Health outcome priorities among competing cardiovascular, fall injury, and medication-related symptom outcomes. J Am Geriatr Soc. 2008;56(8):1409–16.
- Pignone MP, Howard K, Brenner AT, Crutchfield TM, Hawley ST, Lewis CL, et al. Comparing 3 techniques for eliciting patient values for decision making about prostate-specific antigen screening: a randomized controlled trial. JAMA Int Med. 2013;173(5):11.
- Tinetti ME, McAvay GJ, Fried TR, Foody JM, Bianco L, Ginter S, et al. Development of a tool for eliciting patient priority from among competing cardiovascular disease, medication-symptoms, and fall injury outcomes. J Am Geriatr Soc. 2008;56 (4):730-6.
- Oteng B, Marra F, Lynd LD, Ogilvie G, Patrick D, Marra CA. Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme. Sex Transm Infect. 2011;87(1):52–7.
- Sweeting KR, Whitty JA, Scuffham PA, Yelland MJ. Patient preferences for treatment of achilles tendon pain: results from a discrete-choice experiment. Patient. 2011;4(1):45–54.
- Pignone MP, Brenner AT, Hawley S, Sheridan SL, Lewis CL, Jonas DE, et al. Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. J Gen Int Med. 2012;27(1):45–50.
- Laba TL, Brien JA, Jan S. Understanding rational non-adherence to medications: a discrete choice experiment in a community sample in Australia. BMC Family Practice. 2012;13:61.
- Boeri M, Longo A, Grisolia JM, Hutchinson WG, Kee F. The role of regret minimisation in lifestyle choices affecting the risk of coronary heart disease. J Health Econ. 2013; 32(1):253–60.
- Kauf TL, Mohamed AF, Hauber AB, Fetzer D, Ahmad A. Patients' willingness to accept the risks and benefits of new treatments for chronic hepatitis C virus infection. Patient. 2012;5(4):265-78.
- Guo N, Marra CA, FitzGerald JM, Elwood RK, Anis AH, Marra F. Patient preference for latent tuberculosis infection preventive treatment: a discrete choice experiment. Value Health. 2011;14 (6):937–43.
- Scalone L, Watson V, Ryan M, Kotsopoulos N, Patel R. Evaluation of patients' preferences for genital herpes treatment. Sex Transm Dis. 2011;38(9):802-7.
- de Bekker-Grob EW, Rose JM, Donkers B, Essink-Bot M-L, Bangma CH, Steyerberg EW. Men's preferences for prostate cancer screening: a discrete choice experiment. Br J Cancer. 2013;108(3):19.
- Vlemmix F, Kuitert M, Bais J, Opmeer B, van der Post J, Mol BW, et al. Patient's willingness to opt for external cephalic version. J Psychosom Obstet Gynecol. 2013;34(1):15–21.
- Damen TH, de Bekker-Grob EW, Mureau MA, Menke-Pluijmers MB, Seynaeve C, Hofer SO, et al. Patients' preferences for breast reconstruction: a discrete choice experiment. J Plastic Reconstruct Aesthet Surg. 2011;64(1):75-83.

289

- Regier DA, Diorio C, Ethier MC, Alli A, Alexander S, Boydell KM, et al. Discrete choice experiment to evaluate factors that influence preferences for antibiotic prophylaxis in pediatric oncology. PLoS One. 2012;7(10):e47470.
- Marti J. Assessing preferences for improved smoking cessation medications: a discrete choice experiment. Eur J Health Econ. 2012;13(5):533–48.
- Tinelli M, Ozolins M, Bath-Hextall F, Williams HC. What determines patient preferences for treating low risk basal cell carcinoma when comparing surgery vs imiquimod? A discrete choice experiment survey from the SINS trial. BMC Dermatol. 2012;12:19.
- Johnson FR, Manjunath R, Mansfield CA, Clayton LJ, Hoerger TJ, Zhang P. High-risk individuals' willingness to pay for diabetes risk-reduction programs. Diabetes Care. 2006;29(6):1351–6.
- Doyle S, Lloyd A, Birt J, Curtis B, Ali S, Godbey K, et al. Willingness to pay for obesity pharmacotherapy. Obesity. 2012;20(10):2019–26.
- Fiebig DG, Knox S, Viney R, Haas M, Street DJ. Preferences for new and existing contraceptive products. Health Econ. 2011;20 Suppl 1:35–52.
- Walzer S. What do parents want from their child's asthma treatment? Ther Clin Risk Manage. 2007;3(1):167–75.
- Chancellor J, Martin M, Liedgens H, Baker MG, Muller-Schwefe GH. Stated preferences of physicians and chronic pain sufferers in the use of classic strong opioids. Value Health. 2012;15(1):106–17.
- Lloyd A, McIntosh E, Rabe KF, Williams A. Patient preferences for asthma therapy: a discrete choice experiment. Prim Care Respir J. 2007;16(4):241–8.
- Ossa DF, Briggs A, McIntosh E, Cowell W, Littlewood T, Sculpher M. Recombinant erythropoietin for chemotherapyrelated anaemia: economic value and health-related quality-oflife assessment using direct utility elicitation and discrete choice experiment methods. Pharmacoeconomics. 2007;25(3):223–37.
- Lloyd A, McIntosh E, Price M. The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics. 2005;23(11): 1167–81.
- Shafey M, Lupichuk SM, Do T, Owen C, Stewart DA. Preferences of patients and physicians concerning treatment options for relapsed follicular lymphoma: a discrete choice experiment. Bone Marrow Transplant. 2011;46(7):962–9.
- Essers BA, van Helvoort-Postulart D, Prins MH, Neumann M, Dirksen CD. 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. 2010;28(6):507–20.
- Salkeld G, Solomon M, Short L, Ryan M, Ward JE. Evidencebased consumer choice: a case study in colorectal cancer screening. Aust N Z J Public Health. 2003;27(4):449–55.
- McTaggart-Cowan HM, Shi P, Fitzgerald JM, Anis AH, Kopec JA, Bai TR, et al. An evaluation of patients' willingness to trade symptom-free days for asthma-related treatment risks: a discrete choice experiment. J Asthma. 2008;45(8):630–8.
- Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. Value Health. 2009;12(2):354–63.
- Swinburn P, Lloyd A, Ali S, Hashmi N, Newal D, Najib H. Preferences for antimuscarinic therapy for overactive bladder. BJU Int. 2011;108(6):868–73.
- Lloyd A, Penson D, Dewilde S, Kleinman L. Eliciting patient preferences for hormonal therapy options in the treatment of metastatic prostate cancer. Prostate Cancer Prostatic Dis. 2008;11(2):153-9.

- Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ. 2002;11 (5):457–65.
- de Bekker-Grob EW, Hofman R, Donkers B, van Ballegooijen M, Helmerhorst TJM, Raat H, et al. Girls' preferences for HPV vaccination: a discrete choice experiment. Vaccine. 2010;28 (41):6692–7.
- Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J. Patients' preferences for characteristics associated with treatments for osteoarthritis. Rheumatology. 2004;43(3):337-45.
- Pereira CC, Mulligan M, Bridges JF, Bishai D. Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis. Vaccine. 2011;29(7):1443–7.
- Lee WC, Joshi AV, Woolford S, Sumner M, Brown M, Hadker N, et al. Physicians' preferences towards coagulation factor concentrates in the treatment of Haemophilia with inhibitors: a discrete choice experiment. Haemophilia. 2008;14(3):454–65.
- Scalone L, Mantovani LG, Borghetti F, Von MS, Gringeri A. Patients', physicians', and pharmacists' preferences towards coagulation factor concentrates to treat haemophilia with inhibitors: results from the COHIBA Study. Haemophilia. 2009;15(2):473–86.
- Mantovani LG, Monzini MS, Mannucci PM, Scalone L, Villa M, Gringeri A, et al. Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. Haemophilia. 2005;11(6): 589–97.
- Espelid I, Cairns J, Askildsen JE, Qvist V, Gaarden T, Tveit AB. Preferences over dental restorative materials among young patients and dental professionals. Eur J Oral Sci. 2006;114(1): 15–21.
- 56. Lee A, Gin T, Lau AS, Ng FF. A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg. 2005;100 (1):87–93.
- Eberth B, Watson V, Ryan M, Hughes J, Barnett G. Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis. Med Decis Making. 2009;29(6):707–15.
- van Dam L, Hol L, de Bekker-Grob EW, Steyerberg EW, Kuipers EJ, Habbema JD, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. Eur J Cancer. 2010;46(1):150–9.
- Torbica A, Fattore G. Understanding the impact of economic evidence on clinical decision making: a discrete choice experiment in cardiology. Soc Sci Med. 2010;70(10):1536–43.
- Nayaradou M, Berchi C, Dejardin O, Launoy G. Eliciting population preferences for mass colorectal cancer screening organization. Med Decis Making. 2010;30(2):224–33.
- Griffith JM, Lewis CL, Hawley S, Sheridan SL, Pignone MP. Randomized trial of presenting absolute v. relative risk reduction in the elicitation of patient values for heart disease prevention with conjoint analysis. Med Decis Making. 2009;29 (2):167–74.
- Weston A, Fitzgerald P. Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics. 2004;22(18):1195–208.
- Hauber AB, Mohamed AF, Johnson FR, Falvey H. Treatment preferences and medication adherence of people with type 2 diabetes using oral glucose-lowering agents. Diabet Med. 2009;26(4):416–24.
- Schaarschmidt ML, Schmieder A, Umar N, Terris D, Goebeler M, Goerdt S, et al. Patient preferences for psoriasis treatments:

process characteristics can outweigh outcome attributes. Arch Dermatol. 2011;147(11):1285-94.

- Schmieder A, Schaarschmidt ML, Umar N, Terris DD, Goebeler M, Goerdt S, et al. Comorbidities significantly impact patients' preferences for psoriasis treatments. J Am Acad Dermatol. 2012;67(3):363–72.
- Hauber AB, Mohamed AF, Watson ME, Johnson FR, Hernandez JE. Benefits, risk, and uncertainty: preferences of antiretroviralnaive African Americans for HIV treatments. Aids Patient Care STDS. 2009;23(1):29–34.
- Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science. 1981;211(4481):453–8.
- Salisbury LC, Feinberg FM. Alleviating the constant stochastic variance assumption in decision research: theory, measurement, and experimental test. Mark Sci. 2010;29(1):1–17.
- Fiebig DG, Keane MP, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Mark Sci. 2010;29(3):393–421.
- Muhlbacher AC, Lincke HJ, Nubling M. Evaluating patients' preferences for multiple myeloma therapy, a Discrete-Choice-Experiment. Psychosoc Med. 2008; 5:Doc10.
- Shackley P, Slack R, Michaels J. Vascular patients' preferences for local treatment: an application of conjoint analysis. J Health Services Res Policy. 2001;6(3):151–7.
- Bridges JF, Searle SC, Selck FW, Martinson NA. Designing family-centered male circumcision services: a conjoint analysis approach. Patient. 2012;5(2):101–11.
- Goto R, Takahashi Y, Ida T. Changes in smokers' attitudes toward intended cessation attempts in Japan. Value Health. 2011;14(5):785-91.
- Ashcroft DM, Seston E, Griffiths CE. Trade-offs between the benefits and risks of drug treatment for psoriasis: a discrete choice experiment with U.K. dermatologists. Br J Dermatol. 2006;155(6):1236–41.
- Yeung RYT, Smith RD, Mcghee SM. Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale'. Health Econ. 2003;12(9):791–6.
- Heberlein TA, Wilson MA, Bishop RC, Schaeffer NC. Rethinking the scope test as a criterion for validity in contingent valuation. J Environ Econ Manage. 2005;50(1):1–22.
- Carson RT, Flores NE, Meade NF. Contingent valuation: controversies and evidence. Environ Resour Econ. 2001;19(2): 173-210.
- Johnson FR, Ozdemir S, Mansfield C, Hass S, Miller DW, Siegel CA, et al. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. Gastroenterology. 2007;133(3):769–79.
- Johnson FR, Van Houtven G, Ozdemir S, Hass S, White J, Francis G, et al. Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. J Neurol. 2009;256(4):554–62.
- Johnson FR, Hauber AB, Ozdemir S, Lynd L. Quantifying women's stated benefit-risk trade-off preferences for IBS treatment outcomes. Value Health. 2010;13(4):418–23.
- Telser H, Zweifel P. Measuring willingness-to-pay for risk reduction: an application of conjoint analysis. Health Econ. 2002;11(2):129–39.
- Telser H, Zweifel P. Validity of discrete-choice experiments evidence for health risk reduction. Appl Econ. 2007;39(1): 68–78.
- Berry D, Raynor T, Knapp P, Bersellini E. Over the counter medicines and the need for immediate action: a further evaluation of European Commission recommended wordings for communicating risk. Patient Educ Couns. 2004;53(2):129–34.

- Cuite CL, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Med Decis Making. 2008;28(3):377–84.
- France J, Keen C, Bowyer S. Communicating risk to emergency department patients with chest pain. Emerg Med J. 2008;25(5):276–8.
- Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. Health Psychol. 2009;28(2):210–6.
- Gyrd-Hansen D, Halvorsen P, Nexoe J, Nielsen J, Stovring H, Kristiansen I. Joint and separate evaluation of risk reduction: impact on sensitivity to risk reduction magnitude in the context of 4 different risk information formats. Med Decis Making. 2011;31(1):E1-10.
- Hilton NZ, Carter AM, Harris GT, Sharpe AJB. Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. J Interpers Violence. 2008;23(2):171–88.
- Waters EA, Weinstein ND, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. J Health Commun. 2006;11(2):167–82.
- Brewer NT, Tzeng JP, Lillie SE, Edwards AS, Peppercorn JM, Rimer BK. Health literacy and cancer risk perception: implications for genomic risk communication. Med Decis Making. 2009;29(2):157–66.
- Davis JJ. Consumers' preferences for the communication of risk information in drug advertising: most consumers want drug sideeffect information to be rich in detail and easily accessible. Health Aff (Millwood). 2007;26(3):863–70.
- Steiner MJ, Dalebout S, Condon S, Dominik R, Trussell J. Understanding risk: a randomized controlled trial of communicating contraceptive effectiveness. Obstet Gynecol. 2003;102(4):709–17.
- Carling CLL, Kristoffersen DT, Montori VM, Herrin J, Schunemann HJ, Treweek S, et al. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. PLoS Med. 2009;6(8):e1000134.
- Cheung YB, Wee HL, Thumboo J, Goh C, Pietrobon R, Toh HC, et al. Risk communication in clinical trials: a cognitive experiment and a survey. BMC Med Inform Decis Making. 2010;10:55.
- Emmons KM, Wong M, Puleo E, Weinstein N, Fletcher R, Colditz G. Tailored computer-based cancer risk communication: correcting colorectal cancer risk perception. J Health Commun. 2004;9(2):127–41.
- Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. Am J Public Health. 2009;99(12):2196–202.
- Garcia-Retamero R, Galesic M. Using plausible group sizes to communicate information about medical risks. Patient Educ Couns. 2011;84(2):245–50.
- Graham PH, Martin RM, Browne LH. Communicating breast cancer treatment complication risks: when words are likely to fail. Asia Pac J Clin Oncol. 2009;5(3):193–9.
- Ilic D, Murphy K, Green S. Risk communication and prostate cancer: identifying which summary statistics are best understood by men. Am J Mens Health. 2012;6(6):497–504.
- Knapp P, Raynor DK, Woolf E, Gardner PH, Carrigan N, McMillan B. Communicating the risk of side effects to patients an evaluation of UK regulatory recommendations. Drug Saf. 2009;32(10):837–49.
- Miron-Shatz T, Hanoch Y, Graef D, Sagi M. Presentation format affects comprehension and risk assessment: the case of prenatal screening. J Health Commun. 2009;14(5):439–50.

- Pighin S, Savadori L, Barilli E, Rumiati R, Bonalumi S, Ferrari M, et al. Using comparison scenarios to improve prenatal risk communication. Med Decis Making. 2013;33(1):48–58.
- Sheridan SL, Pignone MP, Lewis CL. A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. J Gen Intern Med. 2003;18(11): 884–92.
- 104. Fair AKI, Murray PG, Thomas A, Cobain MR. Using hypothetical data to assess the effect of numerical format and context on the perception of coronary heart disease risk. Am J Health Promot. 2008;22(4):291-6.
- Dolan JG, Iadarola S. Risk communication formats for low probability events: an exploratory study of patient preferences. BMC Med Inform Decision Making. 2008;8:14.
- Edwards A, Thomas R, Williams R, Ellner AL, Brown P, Elwyn G. Presenting risk information to people with diabetes: evaluating effects and preferences for different formats by a webbased randomised controlled trial. Patient Educ Couns. 2006;63(3):336–49.
- Fortin JM, Hirota LK, Bond BE, O'Connor AM, Col NF. Identifying patient preferences for communicating risk estimates: a descriptive pilot study. BMC Med Inform Decis Making. 2001;1:2.
- Garcia-Retamero R, Cokely ET. Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. J Exp Psychol Appl. 2011;17(3):270–87.
- Schapira MM, Nattinger AB, McHorney CA. Frequency or probability? A qualitative study of risk communication formats used in health care. Med Decis Making. 2001;21(6):459–67.
- Sprague D, LaVallie DL, Wolf FM, Jacobsen C, Sayson K, Buchwald D. Influence of graphic format on comprehension of risk information among american indians. Med Decis Making. 2011;31(3):437–43.
- Berry DC, Michas IC, Bersellini E. Communicating information about medication: the benefits of making it personal. Psychol Health. 2003;18(1):127–39.
- Gurmankin AD, Baron J, Annstrong K. The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. Med Decis Making. 2004;24(3):265–71.
- Tait AR, Voepel-Lewis T, Zikmund-Fisher BJ, Fagerlin A. Presenting research risks and benefits to parents: does format matter? Anesth Analg. 2010;111(3):718-23.
- 114. Tait AR, Voepel-Lewis T, Zikmund-Fisher BJ, Fagerlin A. The effect of format on parents' understanding of the risks and benefits of clinical research: a comparison between text, tables, and graphics. J Health Commun. 2010;15(5):487–501.
- Ulph F, Townsend E, Glazebrook C. How should risk be communicated to children: a cross-sectional study comparing different formats of probability information. BMC Med Inform Decis Making. 2009;9:26.
- Young S, Oppenheimer DM. Effect of communication strategy on personal risk perception and treatment adherence intentions. Psychol Health Med. 2009;14(4):430–42.
- Fraenkel L, Wittink DR, Concato J, Fried T. Are preferences for cyclooxygenase-2 inhibitors influenced by the certainty effect? J Rheumatol. 2004;31(3):591–3.
- Whitington D. Improving the performance of contingent valuation studies in developing countries. Environ Resour Econ. 2002;22(1-2):323-67.
- Arrow K, Solow R, Portney PR, Learner EE, Radner R, Schuman H. Report of the NOAA panel on contingent valuation. US Department of Commerce; 1993.
- Adamowicz W, Louviere J, Swait J. Introduction to attributebased stated choice methods. US Department of Commerce; 1998.

- Kahneman D, Tversky A. Psychology of prediction. Psychol Rev. 1973;80(4):237–51.
- Edwards A, Elwyn G. Understanding risk and lessons for clinical risk communication about treatment preferences. Qual Health Care. 2001;10:I9–13.
- Kahneman D, Tversky A. Prospect theory: analysis of decision under risk. Econometrica. 1979;47(2):263–91.
- Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. Science. 2011;333(6048):1393–400.
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2007;8:53–96.
- Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness. Ann Intern Med. 1992;117(11):916–21.
- Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet. 1994;343(8907):1209–11.
- Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexoe J, Nielsen JB. Laypersons' understanding of relative risk reductions: randomised cross-sectional study. BMC Med Inform Decis Making. 2008;8:31.
- CONSORT Group. The CONSORT Statement. http://www. consort-statement.org/consort-statement/. Accessed 8 Nov 2011.
- 130. Krupnick A, Alberini A, Cropper M, Simon N, O'Brien B, Goeree R, et al. Age, health and the willingness to pay for mortality risk reductions: a contingent valuation survey of Ontario residents. J Risk Uncertainty. 2002;24(2):161–86.
- Fagerlin A, Zikmund-Fisher BJ, Übel PA. Helping patients decide: ten steps to better risk communication. J Natl Cancer Inst. 2011;103(19):1436–43.
- Kahneman D, Sugden R. Experienced utility as a standard of policy evaluation. Environ Resour Econ. 2005;32(1):161–81.
- Weinstein MC, Shepard DS, Pliskin JS. The economic value of changing mortality probabilities: a decision-theoretic approach. Q J Econ. 1980;94(2):373–96.
- Viscusi WK. A Bayesian perspective on biases in risk perception. Econ Lett. 1985;17(1-2):59-62.
- Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Med Decis Making. 2005;25(4):460–7.
- Prosser LA, Wittenberg E. Do risk attitudes differ across domains and respondent types? Med Decis Making. 2007;27 (3):281-7.
- Prosser LA, Kuntz KM, Bar-Or A, Weinstein MC. The relationship between risk attitude and treatment choice in patients with relapsing-remitting multiple sclerosis. Med Decis Making. 2002;22(6):506-13.
- Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit–risk preferences and probability-weighted utility using choice-format conjoint analysis. Med Decis Making. 2011;31 (3):469–80.
- Spiegelhalter D. Quantifying uncertainty. In: Skinns L, Scott M, Cox T, editors. Risk. Cambridge: Cambridge University Press; 2011. p. 17–33.
- Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? Med Decis Making. 2010;30(6):672-84.
- 141. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ. 2012;21(6):730–41.
- 142. Seston EM, Ashcroft DM, Griffiths CE. Balancing the benefits and risks of drug treatment: a stated-preference, discrete choice experiment with patients with psoriasis. Arch Dermatol. 2007;143(9):1175–9.

- Hauber AB, Johnson FR, Grotzinger KM, Ozdemir S. Patients' benefit-risk preferences for chronic idiopathic thrombocytopenic purpura therapies. Ann Pharmacother. 2010;44(3):479–88.
- Hauber AB, Mohamed AF, Beam C, Medjedovic J, Mauskopf J. Patient preferences and assessment of likely adherence to hepatitis C virus treatment. J Viral Hepat. 2011;18(9):619–27.
- 145. Mohamed AF, Hauber AB, Neary MP. Patient benefit-risk preferences for targeted agents in the treatment of renal cell carcinoma. Pharmacoeconomics. 2011;29(11):2011.
- Hauber AB, Gonzalez JM, Schenkel B, Lofland JH, Martin S. The value to patients of reducing lesion severity in plaque psoriasis. J Dermatol Treat. 2011;22(5):266–75.
- Hodgkins P, Swinburn P, Solomon D, Yen L, Dewilde S, Lloyd A. Patient preferences for first-line oral treatment for mild-tomoderate ulcerative colitis: a discrete-choice experiment. Patient. 2012;5(1):33–44.
- 148. Arden NK, Hauber AB, Mohamed AF, Johnson FR, Peloso PM, Watson DJ, et al. How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? J Rheumatol. 2012;39(5):1056-63.
- Lathia N, Isogai PK, Walker SE, De AC, Cheung MC, Hoch JS, et al. Eliciting patients' preferences for outpatient treatment of febrile neutropenia: a discrete choice experiment. Support Care Cancer. 2013;21(1):245–51.
- Hauber AB, Arden NK, Mohamed AF, Johnson FR, Peloso PM, Watson DJ, et al. A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. Osteoarthritis Cartilage. 2013;21(2):289-97.
- Aristides M, Weston AR, Fitzgerald P, Le RC, Maniadakis N. Patient preference and willingness-to-pay for Humalog Mix25 relative to Humulin 30/70: a multicountry application of a discrete choice experiment. Value Health. 2004;7(4):442–54.
- Guimaraes C, Marra CA, Colley L, Gill S, Simpson SH, Meneilly GS, et al. A valuation of patients' willingness-to-pay for insulin delivery in diabetes. Int J Technol Assess Health Care. 2009;3:359–66.
- Howard K, Salkeld G, Pignone M, Hewett P, Cheung P, Olsen J, et al. Preferences for CT colonography and colonoscopy as diagnostic tests for colorectal cancer: a discrete choice experiment. Value Health. 2011;14(8):1146–52.
- 154. Faggioli G, Scalone L, Mantovani LG, Borghetti F, Stella A. PREFER study group. 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. 2011;42(1):26-34.
- Gidengil C, Lieu TA, Payne K, Rusinak D, Messonnier M, Prosser LA. Parental and societal values for the risks and benefits of childhood combination vaccines. Vaccine. 2012;30(23): 3445–52.
- Augustovski F, Beratarrechea A, Irazola V, Rubinstein F, Tesolin P, Gonzalez J, et al. Patient preferences for biologic agents in rheumatoid arthritis: a discrete-choice experiment. Value Health. 2013;16(2):385–93.
- Bryan S, Buxton M, Sheldon R, Grant A. Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. Health Econ. 1998;7(7):595–603.
- Bryan S, Roberts T, Heginbotham C, McCallum A. QALYmaximisation and public preferences: results from a general population survey. Health Econ. 2002;11(8):679–93.
- Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, et al. Women and health care professionals' preferences for Down's syndrome screening tests: a conjoint analysis study. BJOG. 2004;111(8):775–9.
- 160. Lewis SM, Cullinane FM, Carlin JB, Halliday JL. Women's and health professionals' preferences for prenatal testing for Down

- Lewis SM, Cullinane FN, Bishop AJ, Chitty LS, Marteau TM, Halliday JL. A comparison of Australian and UK obstetricians' and midwives' preferences for screening tests for Down syndrome. Prenat Diagn. 2006;26(1):60–6.
- 162. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HA, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos Int. 2008;7:1029–37.
- Bunge EM, de Bekker-Grob EW, van Biezen FC, Essink-Bot ML, de Koning HJ. Patients' preferences for scoliosis brace treatment: a discrete choice experiment. Spine. 2010;35(1): 57-63.
- Watson V, Ryan M, Watson E. Valuing experience factors in the provision of Chlamydia screening: an application to women attending the family planning clinic. Value Health. 2009;12(4): 621–3.
- 165. de Bekker-Grob EW, Essink-Bot MI, Meerding WJ, Koes BW, Steyerberg EW. Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete-choice experiment. Pharmacoeconomics. 2009;27(3):211–9.
- 166. Kruijshaar ME, Essink-Bot ML, Donkers B, Looman CW, Siersema PD, Steyerberg EW. A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. BMC Med Res Methodol. 2009;9:31.
- 167. Hol L, de Bekker-Grob EW, van Dam L, Donkers B, Kuipers EJ, Habbema JD, et al. Preferences for colorectal cancer screening strategies: a discrete choice experiment. Br J Cancer. 2010;102(6):972-80.
- Wirostko B, Beusterien K, Grinspan J, Ciulla T, Gonder J, Barsdorf A, et al. Patient preferences in the treatment of diabetic retinopathy. Patient Prefer Adherence. 2011;5:229–37.
- Flood EM, Ryan KJ, Rousculp MD, Beusterien KM, Divino VM, Block SL, et al. Parent preferences for pediatric influenza vaccine attributes. Clin Pediatr. 2011;50(4):338–47.
- Damman OC, Spreeuwenberg P, Rademakers J, Hendriks M. Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery? Med Decis Making. 2012;32(2): 287–300.
- 171. Tong BC, Huber JC, Ascheim DD, Puskas JD, Ferguson TB Jr, Blackstone EH, et al. Weighting composite endpoints in clinical trials: essential evidence for the heart team. Ann Thorac Surg. 2012;94(6):1908–13.
- 172. Sung L, Alibhai SM, Ethier MC, Teuffel O, Cheng S, Fisman D, et al. Discrete choice experiment produced estimates of acceptable risks of therapeutic options in cancer patients with febrile neutropenia. J Clin Epidemiol. 2012;65(6):627–34.
- 173. Marang-vandeMheen PJ, Elsinga J, Otten W, Versluijs M, Smeets HJ, Vree R, et al. The relative importance of quality of care information when choosing a hospital for surgical treatment: a hospital choice experiment. Med Decis Making. 2011; 31(6):816-27.
- Watson V, Carnon A, Ryan M, Cox D. Involving the public in priority setting: a case study using discrete choice experiments. J Public Health. 2012;34(2):253–60.
- Bijlenga D, Bonsel GJ, Birnie E. Eliciting willingness to pay in obstetrics: comparing a direct and an indirect valuation method for complex health outcomes. Health Econ. 2011;20(11): 1392–406.
- Bijlenga D, Birnie E, Mol BW, Bonsel GJ. Obstetrical outcome valuations by patients, professionals, and laypersons: differences within and between groups using three valuation methods. BMC Pregnancy Childbirth. 2011;11:93.

- 177. Palumbo A, De La FP, Rodriguez M, Sanchez F, Martinez-Salazar J, Munoz M, et al. Willingness to pay and conjoint analysis to determine women's preferences for ovarian stimulating hormones in the treatment of infertility in Spain. Hum Reprod. 2011;26(7):1790-8.
- Muhlbacher AC, Nubling M. Analysis of physicians' perspectives versus patients' preferences: direct assessment and discrete choice experiments in the therapy of multiple myeloma. Eur J Health Econ. 2011;12(3):193–203.
- Manjunath R, Yang JC, Ettinger AB. Patients' preferences for treatment outcomes of add-on antiepileptic drugs: a conjoint analysis. Epilepsy Behav. 2012;24(4):474-9.
- 180. Kinsler JJ, Cunningham WE, Nurena CR, Nadjat-Haiem C, Grinsztejn B, Casapia M, et al. Using conjoint analysis to measure the acceptability of rectal microbicides among men who have sex with men in four South American cities. AIDS Behav. 2012;16(6):1436-47.
- King MT, Viney R, Smith DP, Hossain I, Street D, Savage E, et al. Survival gains needed to offset persistent adverse treatment effects in localised prostate cancer. Br J Cancer. 2012;106(4): 638–45.
- Burnett HF, Regier DA, Feldman BM, Miller FA, Ungar WJ. Parents' preferences for drug treatments in juvenile idiopathic arthritis: a discrete choice experiment. Arthritis Care Res. 2012;64(9):1382–91.
- Bridges JF, Mohamed AF, Finnern HW, Woehl A, Hauber AB. Patients' preferences for treatment outcomes for advanced nonsmall cell lung cancer: a conjoint analysis. Lung Cancer. 2012;77(1):224–31.
- Lee SJ, Newman PA, Comulada WS, Cunningham WE, Duan N. Use of conjoint analysis to assess HIV vaccine acceptability: feasibility of an innovation in the assessment of consumer health-care preferences. Int J STD AIDS. 2012;23(4):235–41.
- Wittink MN, Morales KH, Cary M, Gallo JJ, Bartels SJ. Towards personalizing treatment for depression: developing treatment values markers. Patient. 2013;6(1):35–43.

- Zimmermann TM, Clouth J, Elosge M, Heurich M, Schneider E, Wilhelm S, et al. Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. J Affect Disord. 2013;148:210–9.
- Mohamed AF, Johnson FR, Hauber AB, Lescrauwaet B, Masterson A. Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey, and Italy. Eur J Gastroenterol Hepatol. 2012;24(4): 419–26.
- Petrou S, McIntosh E. Women's preferences for attributes of first-trimester miscarriage management: a stated preference discrete-choice experiment. Value Health. 2009;12(4):551-9.
- Sadique MZ, Devlin N, Edmunds WJ, Parkin D. The effect of perceived risks on the demand for vaccination: results from a discrete choice experiment. PLoS ONE. 2013;8(2):e54149.
- Aristides M, Chen J, Schulz M, Williamson E, Clarke S, Grant K. Conjoint analysis of a new chemotherapy: willingness to pay and preference for the features of raltitrexed versus standard therapy in advanced colorectal cancer. Pharmacoeconomics. 2002;11:775–84.
- Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M. Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol. 2004;172(6 Pt 1):2321–5.
- Fraenkel L, Constantinescu F, Oberto-Medina M, Wittink DR. Women's preferences for prevention of bone loss. J Rheumatol. 2005;32(6):1086–92.
- Fraenkel L, Gulanski B, Wittink DR. Preference for hip protectors among older adults at high risk for osteoporotic fractures. J Rheumatol. 2006;33(10):2064–8.
- Goto R, Nishimura S, Ida T. Discrete choice experiment of smoking cessation behaviour in Japan. Tobacco Control. 2007;16(5):336–43.

Appendix 3.4: List of DCEs with a focus on breast cancer and screening from the systematic review

- X1 Taguchi R, Yamazaki Y, Nakayama K. (2010). Eliciting preferences for mammography: using a discrete choice experiment approach Nihon Koshu Eisei Zasshi.;57(2):83-9 [Article in Japanese]
- Kimman, M. L., Dellaert, B. G. C., Boersma, L. J., Lambin, P., & Dirksen, C. D. (2010).
 Follow-up after treatment for breast cancer: one strategy fits all? An investigation of patient preferences using a discrete choice experiment. Acta oncologica. 49(3), 328-337
- Peacock, S., Apicella, C., Andrews, L., Tucker, K., Bankier, A., Daly, M. B., & Hopper, J. L.
 (2006). A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. British Journal of Cancer, 95(10), 1448–1453
- Damen, T. H. C., De Bekker-Grob, E. W., Mureau, M. A. M., Menke-Pluijmers, M. B.,
 Seynaeve, C., Hofer, S. O. P., & Essink-Bot, M.-L. (2011). Patients' preferences for breast
 reconstruction: A discrete choice experiment. Journal of Plastic, Reconstructive and Aesthetic
 Surgery, 64(1), 75–83
- X5 Kruijshaar, M. E., Essink-Bot, M.-L., Donkers, B., Looman, C. W. N., Siersema, P. D., & Steyerberg, E. W. (2009). A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. BMC medical research methodology, 9(1), 31
- Marshall, D. A., Johnson, F. R., Kulin, N. A., Ozdemir, S., Walsh, J. M. E., Marshall, J. K.,
 Van Bebber, S., et al. (2009). How do physician assessments of patient preferences for
 colorectal cancer screening tests differ from actual preferences? A comparison in Canada and
 the United States using a stated-choice survey. Health economics, 18(12), 1420-1439
- X7 Salkeld, G., Solomon, M., Short, L., Ryan, M., & Ward, J. E. (2003). Evidence-based consumer choice: a case study in colorectal cancer screening. Australian and New Zealand journal of public health, 27(4), 449–55.
- X8 Wordsworth, S., Ryan, M., Skatun, D., & Waugh, N. (2006). Women's preferences for cervical cancer screening: a study using a discrete choice experiment. International Journal of Technology Assessment in Health Care, 22(3), 344–350.
- X9 Oteng, B., Marra, F., Lynd, L. D., Ogilvie, G., Patrick, D., & Marra, C. A. (2011). Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme. Sexually transmitted infections, 87(1), 52-57.
- X10 De Bekker-Grob, E.W. et al., (2013). Men's preferences for prostate cancer screening: a discrete choice experiment. British journal of cancer, 108(3), 533–41.
- Walters, S. J., Winslow, M., Collins, K., Robinson, T., Green, T., Madan, J., Wyld, L. (2011).
 Health care professionals' preferences for extending mammographic breast screening to the over 70s. Journal of Geriatric Oncology, 2(1), 1–10
- X12 Berchi, C., Dupuis, J. M., & Launoy, G. (2006). The reasons of general practitioners for promoting colorectal cancer mass screening in France. European Journal of Health Economics, 7(2), 91–98.
- X13 De Bekker-Grob, E. W., Hol, L., Donkers, B., van Dam, L., Habbema, J. D. F., van Leerdam,
 M. E., Steyerberg, E. W. (2010). Labeled versus unlabeled discrete choice experiments in
 health economics: an application to colorectal cancer screening. Value in health, 13(2), 315-323

- X14 Fiebig, D. G., Haas, M., Hossain, I., Street, D. J., & Viney, R. (2009). Decisions about Pap tests: what influences women and providers? Social science & medicine, 68(10), 1766-1774
- X15 Hol, L., de Bekker-Grob, E. W., van Dam, L., Donkers, B., Kuipers, E. J., Habbema, J. D. F., Essink-Bot, M. L. (2010). Preferences for colorectal cancer screening strategies: a discrete choice experiment. British journal of cancer
- X16 Howard, K., & Salkeld, G. (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, 102(6), 972-980
- Marshall, D., Johnson, F. R., Phillips, K. a, Marshall, J. K., Thabane, L., & Kulin, N. a. (2007).
 Measuring patient preferences for colorectal cancer screening using a choice-format survey.
 Value in Health, 10(5), 415–30.
- X18 Nayaradou, M., Berchi, C., Dejardin, O., & Launoy, G. (2010). Eliciting population preferences for mass colorectal cancer screening organization. Medical decision making : an international journal of the Society for Medical Decision Making, 30(2), 224–33.
- X19 Van Dam, L., Hol, L., de Bekker-Grob, E. W., Steyerberg, E. W., Kuipers, E. J., Habbema, J. D. F., van Leerdam, M. E. (2010). What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. European journal of cancer, 46(1), 150-159.

Appendix 3.5: The design of the DCE used in the pilot study

Figure A1: The Ngene $^{\textcircled{B}}$ syntax used for the design of the DCE in the pilot study

Design ;alts = alt1, alt2, alt3 ;rows = 40 ;block=4

;eff=(mnl,d)

;model:

$$\label{eq:U(alt1)} \begin{split} &U(alt1) = s[(n,0.4,0.2)] * saved[3,7,10,14] + r[(n,-0.05,0.02)] * risk[1,5,10,20] + p[(n,-0.5,0.2)] * price[1,2.5,7.5,10] + sr[-0.001] * saved * risk + ps[-0.001] * price * saved + pr[0.001] * price * risk / \end{split}$$

U(alt2) = s*saved + r*risk + p*price + sr*saved*risk + ps*price*saved + pr*price*risk \$

	Fixed	Mean	Standard	Median	Minimum	Maximum
			deviation			
D error	0.00043	0.0009	0.00076	0.000649	0.000352	0.0057
	7	14				
A error	0.00906	0.0225	0.027613	0.013891	0.004997	0.222763
	5	77				
B estimate	41.4266	27.946	14.3455	25.48795	3.449105	78.86291
	6	01				
S estimate	984.542	3508.0	6327.503	1493.657	633.2704	55365.97
	9	36				
		Main Ef	fects		Interactions	
Prior	Detecti	Risk	Cost	Detection*	Cost*	Detection*
	on			Risk	Detection	Cost
Fixed prior value	on 0.4	-0.05	-0.5	Risk	Detection -0.001	Cost 0.001
Fixed prior value Sp estimates	on 0.4 0.40695 8	-0.05 3.9496 58	-0.5 0.530248	Risk -0.001 126.6861	Detection -0.001 984.5429	Cost 0.001 291.0116
Fixed prior value Sp estimates Sp t-ratios	on 0.4 0.40695 8 3.07242	-0.05 3.9496 58 0.9862	-0.5 0.530248 2.691636	Risk -0.001 126.6861 0.174137	Detection -0.001 984.5429 0.062465	Cost 0.001 291.0116 0.114895
Fixed prior value Sp estimates Sp t-ratios	on 0.4 0.40695 8 3.07242 6	-0.05 3.9496 58 0.9862 26	-0.5 0.530248 2.691636	Risk -0.001 126.6861 0.174137	Detection -0.001 984.5429 0.062465	Cost 0.001 291.0116 0.114895
Fixed prior value Sp estimates Sp t-ratios Sb mean	on 0.4 0.40695 8 3.07242 6 55.5733	-0.05 3.9496 58 0.9862 26 28.097	-0.5 0.530248 2.691636 3.126129	Risk -0.001 126.6861 0.174137 176.8636	Detection -0.001 984.5429 0.062465 3470.463	Cost 0.001 291.0116 0.114895 2255.73
Fixed prior value Sp estimates Sp t-ratios Sb mean estimates	on 0.4 0.40695 8 3.07242 6 55.5733 9	-0.05 3.9496 58 0.9862 26 28.097 36	-0.5 0.530248 2.691636 3.126129	Risk -0.001 126.6861 0.174137 176.8636	Detection -0.001 984.5429 0.062465 3470.463	Cost 0.001 291.0116 0.114895 2255.73
Fixed prior value Sp estimates Sp t-ratios Sb mean estimates Sb mean t-	on 0.4 0.40695 8 3.07242 6 55.5733 9 2.55596	-0.05 3.9496 58 0.9862 26 28.097 36 0.8359	-0.5 0.530248 2.691636 3.126129 1.998019	Risk -0.001 126.6861 0.174137 176.8636 0.154099	Detection -0.001 984.5429 0.062465 3470.463 0.047967	Cost 0.001 291.0116 0.114895 2255.73 0.083003

Table A1: The Ngene® efficiency measures for the multinomial logit in the pilot study

		Blo	ck 1	Blo	ck 2	Blo	ck 3	Blo	ck 4
(Choice Set	Alter-							
		native							
		1	2	1	2	1	2	1	2
1	Detection	3	7	10	10	14	3	3	10
	risk	20	1	1	20	20	5	10	5
	Cost	1	8	1	1	10	8	1	3
2	Detection	10	7	10	3	7	14	7	14
	risk	1	10	1	20	10	1	10	1
	Cost	10	8	10	1	8	10	3	10
3	Detection	10	10	14	7	7	14	10	10
	risk	20	1	20	5	5	1	5	10
	Cost	1	3	10	8	8	10	8	3
4	Detection	14	14	7	10	7	7	10	3
	risk	1	20	5	5	10	5	5	10
	Cost	3	1	3	10	8	10	3	1
5	Detection	10	14	14	3	7	3	3	7
	risk	1	20	20	20	1	20	5	5
	Cost	1	1	10	1	8	1	8	8
6	Detection	3	10	14	14	10	3	10	3
	risk	20	5	20	1	10	5	10	10
	Cost	1	10	1	3	3	3	3	3
7	Detection	3	14	14	3	14	7	14	7
	risk	10	1	1	10	1	10	1	10
	Cost	3	10	10	3	10	3	10	8
8	Detection	7	7	3	7	14	10	3	3
	risk	5	10	20	5	20	1	5	20
	Cost	10	8	1	8	1	1	8	1
9	Detection	3	14	7	14	7	3	10	10
	risk	5	20	5	1	5	20	10	5
	Cost	8	10	3	10	8	1	3	8
10	Detection	3	10	7	14	14	7	3	10
	risk	10	10	1	20	20	1	10	10
	Cost	3	3	8	10	10	8	1	3

Table A2:	The	Ngene®.	generated	design	of the	DCE in	n the	pilot s	tudy

Detection: 3=3%; 7=7%; 10=10%; 14=14%

Risk: 1=1%; 5=5%; 10=10%; 20=20%

Cost: 1=£100 (£20/screen); 3=£250 (£50/screen); 8=£750 (£150/screen); 10=£1,000 (£200/screen)

		Census	BHPS				
Age	Question	What is your date of birth?	Can I just check your age?				
	Answer	Free-text/comment	Free-text/comment				
Education	Question	Which of these qualifications do you	Have you attended any education				
		have? Tick every box that applies if you	institution full-time				
		have any of the qualifications listed. If	since September?				
		your qualification is not listed, tick the					
		box that contains its nearest equivalent.					
	Answer	• 1-4 O levels	• Yes				
		• NVQ level 1	• No				
		• 5+ O levels					
Country	Question	What is your country of birth?	Where were you born?				
of birth							
	Answer	• England	Free-text/comment				
		• Wales					
		• Scotland					
		• Wales					
		• Northern Ireland					
		• Republic of Ireland					
Ethnicity	Question	What is your ethnic group?	To which of these ethnic groups do				
			you consider you belong?				
	Answer	White	White				
		• British	• British				
		• Irish	Mixed				
		• Gypsy/traveller	• White and Black Caribbean				
		Mixed	• White and Black African				
		• White and Black Caribbean	• White and Asian				
		• White and Black African	Asian or Asian British				
		• White and Asian	• Indian				
		• Other	• Pakistani				
		Asian	Bangladeshi				
		• Indian	Black or Black British				
		• Pakistan	Caribbean				
		• Bangladeshi	• African				
		• Chinese	Chinese				
		• Other					
		Black					
		• African					
		Caribbean					
		• Other					
		Other					
		• Arab					
		• Other					

Appendix 3.6: Review of questions used to elicit individual characteristics in commonly-used surveys

		Census	BHPS
Religion	Question	What is your religion?	What is your religion?
	Answer	No religion	Catholic
		• Christian	• Presbyterian
		• Buddhist	Church of Ireland
		• Hindu	• Methodist
		• Jewish	• Baptist
		• Muslim	• Presbyterian
		• Sikh	• Brethren
		• Other	• Protestant
			• Buddhist
			• Hindu
			• Jewish
			• Muslim
			• Sikh
			No religion
Employment status	<i>Question</i> Last week, were you: tick all that apply, include any paid work.		Last week were you in paid employment at all, including being
			away temporarily from a job you would normally have been doing?
	Answer	• Working as an employee?	
		• On a government sponsored training scheme?	
		• Self-employed or freelance?	
		• Working paid or unpaid or your family's business?	
		 Away from work ill, on maternity leave, on holiday or temporarily laid off? 	
		• Doing any other kind off paid work?	
		• None of the above	
Income	Question	N/A	On average, what was your
			MONTHLY income from this
			Job/business over the last 12
	A		months?
	Answer		Fiee-text/comment

Appendix 3.7: Confirmation letter of University Research Ethics approval



The University of Manchester

Faculty of Engineering and Physical Sciences (B42) The University of Manchester Sackville Street Building Manchester M13 9PL

+44(0)161 306 3084 Adrian.Jarvis@manchester.ac.uk

Ref: AJ/ethics/1809/13

Miss. Vass Centre for Health Economics Institute for Population Health 3rd October 2013

Dear Miss. Vass

Research Ethics Committee 3

Vass, Payne, Rigby, Campbell: Using discrete choice experiments to value risks and benefits in primary care (ref 13178)

I write to confirm that the Chair is now satisfied that you have addressed the concerns of the Ethics Committee of the 18th of September 2013 and has therefore given the above research project a favourable ethical opinion

This approval is effective for a period of five years and if the project continues beyond that period it must be submitted for review. It is the Committee's practice to warn investigators that they should not depart from the agreed protocol without seeking the approval of the Committee, as any significant deviation could invalidate the insurance arrangements and constitute research misconduct. We also ask that any information sheet should carry a University logo or other indication of where it came from, and that, in accordance with University policy, any data carrying personal identifiers must be encrypted when not held on a university computer or kept as a hard copy in a location which is accessible only to those involved with the research.

Finally, I would be grateful if you could complete and return the attached form at the end of the project or by September 2014.

We hope the research goes well.

Yours sincerely

Adran Jerris.

Adrian Jarvis Ethics Committee 3 Secretary

Appendix 3.8: The DCE survey used for the pilot study

MANCHESTER CENTRE FOR HEALTH ECONOMICS

Start

Γ

Which breast screening programme would you prefer?

A survey to find out how you balance the risks and benefits in breast screening.

Click the button below to enter the survey.

Manchester Centre for Health Economics





Consent

We are interested in how women understand the different risks and benefits of a national breast screening programme.

You have been invited to take part in this survey because, as a woman aged between 18 and 70, you may have considered, or may consider in the future, taking part in a breast screening programme.

The survey will take around 20 minutes to complete.



If you have any questions, please contact:

Caroline Vass Manchester Centre for Health Economics Jean McFarlane Building The University of Manchester Oxford Road Manchester M13 9PL <u>Caroline, vass@postgrad.manchester.ac.uk</u> Tel. 077 1675 1262

	$\langle $		
0%			100%

NHS tute fo National Institu Health Res





National Institute for Health Research School for Primary Care Research



Intro

What is breast cancer and what is breast screening?

Breast cancer is a group of cells which can grow in an abnormal way and form into a lump (known as a cancer or tumour). As the lump grows, cancerous cells can spread to other parts of the body and this can be very serious.

Breast screening is a method of detecting breast cancer at a very early stage. Screening involves an xray of each breast (a mammogram) which is taken while carefully compressing the breast. Screening for breast cancer has been shown to be safe although most women find it a bit uncomfortable, and a few find it painful.

The x-ray can detect small changes in breast tissue which may indicate a very early stage cancer. These cancers are generally too small to be detected by touch. If there is something unusual on the x-ray then the woman will be asked to return for further tests to check whether this is a cancer.



NHS National Institute for Health Research chool for Primary Care Research



INTRO2

As mentioned previously, the x-ray images are not always clear and sometimes further tests are needed to confirm if cancerous cells are present. These tests usually require a small sample of tissue (a biopsy) to be taken from the breast by a doctor, using a long needle, under local anaesthetic. Even when there is a tumour, sometimes these small lumps are treated with surgery and chemotherapy when really they would not have grown into something dangerous. It is impossible for doctors to tell which tumours will become life-threatening so everything identified in screening is treated.

This means that screening for breast cancer can lead to further tests and procedures which may have been unnecessary. There will be a delay between the tests and hearing the results which might be stressful for you. Also, the procedures, such as the biopsy, can be uncomfortable or painful.







INTRO3pic

This figure shows the different options that could happen in screening.





intro4

In the UK, women over the age of 50 are offered screening up until the age of 70. Although the service is free, women who participate in screening incur some cost. For example, they may have to take some time off work, travel to the centre for breast screening, pay for parking and transport, or maybe pay for childcare.

There are **benefits** to breast screening for cancer which means tumours are picked up earlier and the long-term health outcomes of women are improved. However, there are **risks** because it is inconvenient and can mean women undergo procedures unnecessarily.

In this survey we are interested in understanding how women balance these benefits and risks. To do that, we would now like you to complete this survey where you will be shown different screening programmes. You will be asked to choose your preferred programme from a set of possible screening programmes.

There are no wrong or right answers; we are just interested in your views





IntroRISK

Risk is a term used to explain the chance that something bad might happen. In breast cancer screening, there is a chance that women might have unnecessary treatment after their mammogram. This will not happen to all women, but there is a risk it could happen to some women.

The women shaped figures represent the total number of women having screening. The purple coloured figure shows the one woman who went for screening and had unnecessary treatment. This means 1 out of every 100 women screened was treated when she didn't need to be, but 99 women were treated appropriately.





NHS National Institute for Health Research

OR



IntroNORISK

Risk is a term used to explain the chance that something bad might happen. In breast cancer screening, there is a chance that women might have unnecessary treatment after their mammogram. This will not happen to all women, but there is a risk it could happen to some women.



A risk of 1% means that 1 out of every 100 women screened was treated when she didn't need to be, but 99 women were treated appropriately.





IntroRISK1



					*								*				*			
٨		*	Å		Å			*	٨	-								Å	豪	
ii m	山山	山山	II m	iii m	重点	iii m	П. "m.	重点	il.	一点	山山	11 (m)	iii iii	₩ ▲	曲	前	H A	iii.	重点	
T	T	Ţ	11	T	T	Ť	T	II.	T	Ţ	T	T	Ţ	T	T	Ĩ	T	T	T	
	#					#	骨	#	#	#	骨	#		#	*	#	#		#	
*					南	*	*	禴	*				*					*	南	



OR



IntroNORISK2



Similarly, a risk of 5% means that 5 out of every 100 women screened was treated when she didn't need to be, but 95 women were treated appropriately.



National Institute for Health Research

NHS



attributesandlevs

You are going to be asked to choose between screening programmes which differ in terms of:

Women who will have cancers detected by screening

This describes the number of women who will have a cancer identified by a mammogram (x-ray of the breast). This describes the number of women, who go for screening over a period of 20 years, that are identified as having a cancer.

Women who will have unnecessary treatment

This shows the number of women requiring a procedure when the lump was not cancerous or would not have caused any harmful symptoms in their lifetime. The unnecessary treatment includes biopsies and possibly further surgery and chemotherapy.

Out-of-pocket cost

This describes the cost to you of participating in a screening programme. There are no plans to introduce charges for screening with the NHS. This cost reflects your out-of-pocket expenses incurred when participating in the screening programme. For example, transport costs, car parking, child care, time off work. In the UK, women will be screened 5 times over a 20 year period (a lifetime programme). We'll show the cost of the programmes over your life time and what that works amounts to per screen.

Each screening programme we show you will be described in terms of these three characteristics.

Each characteristic, however, has four different levels that will change.



National Institute for Health Research for Primary Care



detectedbyscreen

Women who will have cancers detected by screening





detectedbyscreenNORISK

Women who will have cancers detected by screening



means that if 100 women are screened for 20 years, then 3 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 7 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 10 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 14 women will be identified as having cancer.





unnecessary



unnecessaryNORISK

Women who will have an unnecessary treatment



means that if 100 women are screened for 20 years, then 1 woman will be treated unnecessarily.

5% means that if 100 women are screened for 20 years, then 5 women will be treated unnecessarily. 10%

means that if 100 women are screened for 20 years, then 10 women will be treated unnecessarily.







costs

Out-of-pocket cost to you of screening programme per year



£20 (or £100 over your lifetime)



£50 (or £250 over your lifetime)



£150 (or £750 over your lifetime)



£200 (or £1,000 over your lifetime)





Exampleprog

Below is an example of one possible breast screening programme. Programme A



0%

National Institute for Health Research School for Primary Care Research





100%

ExampleprogNORISK

Below is an example of one possible breast screening programme. Programme A



0%

NHS
National Institute for
Health Research
School for Primary Care Research

100%



Examplechoice

Below is an example of the type of question we will ask you. If these were your only options, which would you choose?

	Programme A	Programme B	
Women who will have cancers detected by screening	10%	10%	
Women who will have an unnecessary follow-up	20%	1%	
Cost to you	E20 per screen (£100 over your lifetime)	E50 per screen (£250 over your lifetime) E20 E20 E20 E20 E20 E20 E20 E20 E20 E20	NONE: I wouldn't choose any of these.
	Θ	0	Θ

You simply select the programme which you prefer.



ExamplechoiceNORISK

Below is an example of the type of question we will ask you. If these were your only options, which would you choose? Programme A



You simply select the programme which you prefer.





FeedbackB3

Thinking back over what you just read, is there anything about the breast screening information you've been given that seems unclear? (e.g words that are confusing or things which could be phrased better)



FeedbackBexpl2

Is there any more information or further explanation you would like?





STARTDCE

Please press the "next" arrow below to begin the survey.



NHS National Institute for Health Research School for Primary Care Research



DCE_Random1

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

	Programme A	Programme B	
Women who will have cancers detected by screening	10%	10%	
Women who will have an unnecessary follow-up	20%	1%	NONE: I wouldn't choose any of these
Cost to you	£20 per screen (£100 over your lifetime)	(£250 over your lifetime)	
	DCE_Random1=1	DCE_Random1=2	DCE_Random1=3

National Institute for Health Research



MANCHESTER CENTRE FOR HEALTH ECONOMICS

DCEnorisk_Random1

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

(1 of 10 choice questions)



National Institute for Health Research School for Primary Care Research



Endofchoicetask

Thank you for providing your choices. You will now be asked some questions about the survey and a few questions about yourself.



National Institute for Health Research School for Primary Care Research



FeedbackA

How confident are you that you would make the same choices if faced with the situations in real-life?

Feedbacka=1 Very confident I would make the same choices Development of the same choices Development of the same choices Feedbacka=3 Not confident I would make the same choices Feedbacka=4 Confident I would make different choices

FeedbackAexpl

Please explain your answer

	$\langle \neg \neg \rangle$	
0%		100%



FeedbackB

On a scale of 1-5, how easy or difficult did you find making choices between the alternatives?

	1 (Very easy)	2	3 (Neither difficult nor easy)	4	5 (Very difficult)
	FeedbackB_r1=1	FeedbackB_r1=2	FeedbackB_r1=3	FeedbackB_r1=4	FeedbackB_r1=5
Fee	dbackBexpl				
	Please explain you	ranswer			
			$\langle \neg \neg \rangle$		
		0%		100%	



FeedbackB1

On a scale of 1-5, how easy or difficult did you find the survey to understand?

(Ve	1 ry easy)	2	3 (Neither difficult nor easy)	4	5 (Very difficult)
Feedba	ackB1_r1=1	FeedbackB1_r1=2	FeedbackB1_r1=3	FeedbackB1_r1=4	FeedbackB1_r1=5

FeedbackB2

Was there anything in the survey that was particularly unclear?

FeedbackB2=1 No, I understood everything

\sim		
FeedbackB2=2	FeedbackB2_2_other There were some	things that weren't clear or well explained
\sim	(please comment in the box below)	

FeedbackBexpl1

Is there anything else about the survey we could change to make it easier to understand (e.g phrasing, explanations)?






FeedbackANA

Did you find yourself making choices based on one or two characteristics of the screening programme?

FeedbackANA=1 Yes, I based choices on one or two characteristics FeedbackANA=2 No, I considered ALL characteristics

FeedbackANAdrop

Would you say you considered the characteristic women who have a cancer detected?
Please select

FeedbackANAdrop1

Would you say you considered the characteristic <u>risk of uncessary treatment</u>?

 Please select
 Image: Select Se

FeedbackANAdrop2

١	Would you say you considered the characteristic <u>cost</u> ?	
	Please select	

FeedbackANAexpl

If you based your choices on one or two characteristics, why was this? (Please tick all that apply):

FeedbackANAexp[_1] There was just too many characteristics on which to make a decision

FeedbackANAexpl_2 The other characteristic(s) were too complicated or it was unclear what it meant

FeedbackANAexpl_3 FeedbackANAexpl_3 other Other reason (please explain)



National Institute for Health Research

T.



BackgroundB







NHS National Institute for Health Research



BackgroundD

What is the highest level of education you have obtained?



0%

NHS National Institute for Health Research tool for Primary Care Research

100%





NHS National Institute for Health Research School for Primary Care Research



BackgroundF



0%

NHS National Institute for Health Research ichool for Primary Care Research

100%

	MANCHESTER CENTRE FOR HEALTH ECONOMICS	
BackgroundG Do you have any children?		
BackgroundG=2		
0%	100%	
		NHS National Institute for Health Research School for Primary Care Research

MCHE	MANCHESTER CENTRE FOR HEALTH ECONOMICS
------	---

BackgroundH

Do you have any daughters?

	$\langle $	
0%		100%

National Institute for Health Research chool for Primary Care Research



ExperienceA

Have you ever had breast screening (a mammogram)?



ExperienceB

Have you received follow-up for breast cancer?

ExperienceB=1	Yes
ExperienceB=2	No
ExperienceB=3	Don't know

	$\langle \neg \neg \rangle$	
0%		100%





ExperienceC

Has anyone in your family (blood relative or otherwise) had breast cancer?

\cup	
ExperienceC=3	Don't know
\bigcirc	Don't know

ExperienceC1

Do you know of anybody else (e.g friends or colleagues) who has had breast cancer?

ExperienceC1=1	Yes
ExperienceC1=2	No
ExperienceC1=3	Don't know

	$\langle \neg \neg \rangle$	
0%		100%

NHS National Institute for Health Research



RiskD

How would you describe your current risk of breast cancer?



RiskD1

How would you describe your future risk of breast cancer?

RiskE

To what extent are you concerned about your own risk of breast cancer?



RiskF

Is there anything else you would like to tell us about your risk of breast cancer?



Notional Institute for Health desparth



FeelRiskA

On a scale of 1-5, how much do you agree with the following statements?

"I feel life is very much out of our hands, and fate determines everything"

3 1 2 (Neither agree nor 4 5					5
	(Strongly disagree)	(Disagree)	disagree)	(Agree)	(Strongly agree)
	FeelRiskA_r1=1	FeelRiskA_r1=2	FeelRiskA_r1=3	FeelRiskA_r1=4	FeeiRiskA_r1=5

FeelRiskB

"I get a thrill when I gamble, even if it's just the lottery"

1 (Strongly disagree)	2 (Disagree)	3 (Neither agree nor disagree)	4 (Agree)	5 (Strongly agree)
FeelRiskB_r1=1	FeelRiskB_r1=2	FeelRiskB_r1=3	FeelRiskB_r1=4	FeelRiskB_r1=5

FeelRiskC

"Even if something is unlikely, I feel it will probably happen to me"

3					
	1	2	(Neither agree nor	4	5
	(Strongly disagree)	(Disagree)	disagree)	(Agree)	(Strongly agree)
	FeelRiskC_r1=1	FeelRiskC_r1=2	FeelRiskC_r1=3	FeelRiskC_r1=4	FeelRiskC_r1=5

FeelRiskD

"I have an idea about whether or not I'll develop breast cancer"



Wational Institute for Houlth Resourch choul for Primary Care Research



NumeracyA

This next section will ask you some questions to understand how familiar you are with probabilities.

If you are unsure of the answer, you can skip the question.

Imagine I flip a coin 1,000 times. What is your best guess about how many times the coin would fall heads up in 1,000 flips?

NumeracyB

If there was a lottery where the chance of winning £10 is 1%, if 1,000 people bought a ticket, how many people would you expect to win?

NumeracyC

In another lottery, the chance of winning a car is 1 in 1,000. What percentage of tickets in the lottery will win a car?



National Institute for Health Research



Comments

Did any of the questions in the second half of the survey (about you, risk or breast cancer) seem confusing or unclear? (You can use the arrows to look back)

Comments2

Can you think of any other questions which might help us understand how you made your choices in the choice task?

Comments1

If you have any further comments or suggestions about the survey or the issues it raised then please let us know using the space here.

0%

National Institute for Health Research



help

Thank you for completing this survey.

If you have been affected by any of the issues raised in this survey, the following organisations may be of assistance:

- National Hereditary Breast Cancer Helpline on 01629 813000 - Breast Cancer Care on 0808 800 6000

If you have any questions or concerns please contact Caroline Vass on 077 1675 1262 or caroline.vass@postgrad.manchester.ac.uk.

Manchester Centre for Health Economics, Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL



NHS National Institute for Health Research

Appendix 3.9: Online advertisement for study participants



You are here: My Manchester > Research volunteering > Opportunity

Research volunteering	UNDERSTANDING RISK IN HEALTHCARE: AN INTERVIEW STUDY				
Volunteering Opportunity	DESCRIPTION				
Submission	We make choices that involve risk all the time. How we travel, what we choose to eat, all involves balancing the risks and benefits. We are interested in how you balance risks				
Guidelines for recruiting volunteers	We are looking for female volunteers aged 18-70 to take part in an interview. The interview will involve the use of colourful images so volunteers with red-green colour blindness				
Removing your request	are not eligible to partake in this study. The interview should take no more than 45 minutes and can take place on the university campus or at a location convenient for you.				
Contact us	If you would like to find out more information please contact Caroline Vass.				
Join mailing list	Name of Research Ethics Committee: University Research Ethics Committee 3				
Share this page:	University Ethics Committee number and/or NHS Reference number (if applicable): AJ/ethics/1809/13				
	CONTACT DETAILS				
	- corpling visce@postgrad mansheeter on uk or 0771 675 1262				

• caroline.vass@postgrad.manchester.ac.uk or 0771 675 1262

Appendix 3.10: Preliminary interview schedule for pilot study

Introduction:

My name is Caroline Vass and I'm a PhD student at the Manchester Centre for Health Economics. My PhD is about how people make a balance between risks and benefits when making choices in health care. I am particularly interested in how we can use surveys to understand their views on risks and benefits. This interview is about how you balance risks and benefits in health care. I will ask you to complete a questionnaire and to say out-loud how you have come to your answers.

The interview should take about 45 minutes, and there will be imaginary examples looking at breast cancer screening programmes.

Have you read the information sheet? Please complete this consent form. So, are you ok to continue?

Topic 1: The DCE

Here is a survey made up of questions about imaginary breast cancer screening programmes. Take a few minutes to read through the information and definitions at the front of the survey. Do you have any questions at this point?

Instruction 1: As you complete the survey I want you to talk aloud by trying to say out-loud what you're thinking as you're completing this questionnaire.

Don't worry about making sense or talking to me, I am merely going to listen to what you have to say. If you go a bit quiet, I might ask some questions to get you talking again. Is that ok? Probes:

- What are you thinking now?
- Why did you choose that one?
- Would you choose that if I wasn't here?
- Are you considering all of the information presented?

[If respondent is choosing A's or B's, probe why...] [If respondent fails internal tests [QUESTION 6], probe why...]

Topic 2: Ease of Completion

Question 1: What did you think about the questions?

Question 2: Did you find yourself concentrating on a particular characteristic or do you think you weighed them up evenly?

Question 3: Would you change any of your answers?

Question 4: Do you think you are satisfied with your choices?

Topic 3: Further questions

These questions are just so I can understand if different people see things differently.

Question 1: What is your current employment status? What kind of work do / did you do? Was/Is that full or part-time? Do you have to use numbers in this job?

Question 2: Do you play games that involve a gamble? If not clear, suggest betting on horses or bingo. Where do you play <insert game>? Is that online?

<u>Topic 4: Final thoughts</u> Do you have any other feedback or thoughts? Thank you for completing this interview.

Appendix 3.11: Information sheet for interviewees for the qualitative interview study



NHS National Institute for Health Research National School for Primary Care

Understanding risk in healthcare: an interview study

You are being invited to take part in a research study that aims to understand how people complete surveys. The example we will use is choosing a screening programme for breast cancer. Before you decide whether to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the aim of the research?

This study wants to understand how people make choices about healthcare. The aim is to explore how people balance the risks and benefits when making a decision about breast cancer screening.

Why have I been chosen?

This study aims to explore the views of women. Any female member of the public who is able to complete an interview in English is able to take part in this study.

What would I be asked to do if I took part?

You will be invited to complete an interview at a location convenient to you or on University of Manchester premises if you prefer. The interview should take no more than 45 minutes. During the interview I will ask you to complete a questionnaire with different imaginary breast cancer screening programmes and you will be asked which you prefer. It is important to remember that there are no right or wrong answers; we are just trying to explore your views. I will also ask your permission to record the interview so that I do not have take notes to remember what was said.

What happens to the data collected?

The interview will be recorded and then typed up into a written record of what was said .The answers to the survey will be used to help us understand how breast cancer screening programmes can be better provided. Your name will not be recorded and each interview will be given an anonymous identification number. We might want to use some of the direct quotes from the interview in a document and/or research paper summarising the findings but your name will not be recorded with these quotes.

How is confidentiality maintained?

The law called the Data Protection Act (1998) tells us how to keep your information secure. If you provide personal information (name, address, contact details) it will remain confidential and we will not give your details to anyone else.

Will I be paid for participating in the research?

We are not able to pay you for taking part in this study. If you decide to travel to the University for the interview then we can reimburse your travel expenses in line with university policy. To thank you for your time, you will receive an Amazon voucher for participating in this research.

Where will the research be conducted?

The interview will take place at a time and location convenient for you where it is possible to record what was said privately and clearly. If you prefer, then you are free to travel to the University for the interview. The location of the interview is your choice.

Will the outcomes of the research be published?

The main outcome of this research will be a PhD thesis for Caroline Vass, a student based in the Manchester Centre for Health Economics at The University of Manchester. In addition, we may want to report the findings at conferences or in a published journal article.

My question is not covered here, who do I contact for more information ?

Please contact Caroline Vass (PhD Student in the Centre for Health Economics) on 077 1675 1262 or email caroline.vass@postgrad.manchester.ac.uk.

<u>Thank you for reading this information sheet</u> <u>Please keep this information sheet for your records</u>

Appendix 3.12: Consent form for interviewees for the qualitative interview study



NHS National Institute for Health Research

School for Primary Care Research

Consent form Understanding risk in healthcare

		Please tick			
 I confirm that I have read the attached info project and have had the opportunity to consi questions and had these answered satisfactor 	rmation sheet on the abov ider the information and as prily.	e k			
2. I understand that my participation in the stu free to withdraw at any time without giving a	udy is voluntary and that I a reason	am			
3. I understand that the interviews will be digitally voice-recorded and notes may be taken					
4. I agree to the use of anonymous quotes in	academic articles.				
I agree to take part in the above project					
Name of participant	Date	Signature			
Name of person taking consent	Date	Signature			

Appendix 3.13: NHS Choices video approval

RE: Feedback - Content - 177133 JA

NHS Choices Service Desk [servicedesk@nhschoices.nhs.uk] Sent:08 May 2014 11:52 To: Caroline Vass

Dear Caroline,

Thank you for contacting the NHS Choices Service Desk. Please see the response below from our Video Content team.

" Hi Caroline

We would be happy for you to use the video. You can either link to the video, or embed the video in your webpage via the embed code on the video player."

Please get back to us if you have any other queries. Kind Regards, James The NHS Choices Service Desk

-----Original Message-----From: NHS Choices Service Desk Sent: 04 April 2014 13:03 To: 'Caroline Vass' Subject: RE: Feedback - Content - 177133 JA

Dear Caroline,

Thank you for contacting the NHS Choices Service Desk.

We have now assigned your query to the Subject Matter Expert (SME) within NHS Choices. A member of our service desk will be in touch as soon as we receive a response.

If you require an update at any time, please feel free to respond to this email quoting the reference 177133 in the subject line.

Kind Regards,

James

The NHS Choices Service Desk

Appendix 3.14: Pilot quantitative results of the DCE from the internet panel

	Overall	Icon arrays	Percentages only
Completed to termination	56	29	27
Completed choice questions	62	31	31

Table A1: Response statistics

Table A2: Effect of attributes on choice probabilities and interactions (conditional logit)

	No Interactions		Two-way Interactions	
Detect	0.142***	(0.02)	0.145***	(0.02)
Risk	-0.035***	(0.01)	-0.011	(0.01)
Cost	-0.001***	(0.00)	-0.001***	(0.00)
Detect*risk			-0.003	(0.00)
Detect*cost			0.000	(0.00)
Risk*cost			0.000	(0.00)
Observations	1860		1860	

* p<0.05 ** p<0.01 *** p<0.001; Standard errors in parentheses

Table A3: Effect of attributes on choice probabilities by risk communication format (conditional logit)

	Percentage partici	ipants	Icon array parti	cipants
Detect	0.107***	(0.03)	0.179***	(0.03)
Risk	-0.013	(0.01)	-0.060***	(0.01)
Cost	-0.001*	(0.00)	-0.001**	(0.00)
Observations	930		930	

* p < 0.05 ** p < 0.01 *** p < 0.001; Standard errors in parentheses

Table A4: Effect of attributes on choice probabilities with interaction of risk communication format
(conditional logit)

Effect of attributes on choice probabilities: pooled model		
Detect	0.107***	(0.03)
Risk	-0.013	(0.01)
Cost	-0.001*	(0.00)
IAP*detect	0.073	(0.04)
IAP*risk	-0.047**	(0.01)
IAP*cost	-0.000	(0.00)
Observations	1860	

* p<0.05 ** p<0.01 *** p<0.001; Standard errors in parentheses

 Table A5: Effect of attributes on choice probabilities and testing for significance of the scale parameter (heteroskedastic conditional logit)

	Pooled model	
Detect	0.107***	(0.03)
Risk	-0.013	(0.01)
Cost	-0.070*	(0.03)
IAP*detect	0.016	(0.05)
IAP*risk	-0.028	(0.02)
Observations	1860	
Scale term: IAP	0.380	(0.573)

* p<0.05 ** p<0.01 *** p<0.001; Standard errors in parentheses

Table A6: Summary of internal validity tests

Risk version	Mean	Std. Dev	Freq	
Percentages only	. 19354839	. 39529157	930	
Icon arrays	. 03225806	. 17677976	930	
Total	. 11290323	. 31655957	1860	

Table A7: Willingness-to-Pay estimates (conditional logit)

	Overall		Icon Arrays		Percentages Only	
	Risk	Detect	Risk	Detect	Risk	Detect
WTP	-40.95747	165.60727	-58.776433	174.87881	-18.5265	152.36964
LL	-63.490184	108.93042	-95.267242	101.76807	-45.162919	65.777144
UL	-18.424909	222.28411	-22.285624	247.98955	8.1099195	238.96214

Table A8:	Effect of attributes o	n choice probabili	ties by risk c	communication method	(mixed logit)
-----------	------------------------	--------------------	----------------	----------------------	---------------

	Percentage parti	cipants	Icon array participants	
Cost	-0.001*	(0.00)	-0.002***	(0.00)
Risk	-0.052	(0.04)	-0.126**	(0.04)
Detect	0.255*	(0.10)	0.561***	(0.13)
Standard deviation				
Risk	0.193***	(0.04)	0.173***	(0.05)
Detect	0.457***	(0.10)	0.564***	(0.12)
Observations	930		930	

* p<0.05 ** p<0.01 *** p<0.001; Standard errors in parentheses

Table A9: Willingness-to-Pay estimates (mixed logit)

	Overall		Icon Arrays		Percentages Only	
	Risk	Detect	Risk	Detect	Risk	Detect
WTP	-67.52141	252.98664	-68.411058	305.1174	-44.149608	215.4771
LL	-114.29974	128.84089	-119.31961	135.38281	-116.65542	28.015402
UL	20743079	377.13238	-17.502507	474.852	28.356208	402.9388

Appendix 3.15: The design of the DCE used in the final empirical studies

	Fixed	Mean	Standard	Median	Minimum	Maximum
			deviation			
D error	0.000934	0.016483	0.02654	0.007882	0.001087	0.27508
A error	0.012077	0.609606	2.254365	0.12509	0.011562	29.740286
B estimate	77.21236 1	27.702787	14.063811	24.75517	2.212302	84.387436
S estimate	173.1685 89	48301.844 74	358251.9342	557.376104	14.588552	4479484.918
Main Effects			Interactions			
Prior	Detect-	Risk	Cost	Detection*	Cost*	Detection*
	1011			K15K	Detection	Cost
Fixed prior value	0.1	-0.06	-0.14	-0.004	-0.015	-0.002
Sp estimates	7.557505	12.976973	7.845484	20.841688	5.689372	173.168589
Sp t-ratios	0.712963	0.544088	0.699755	0.429328	0.82172	0.148943
Sb mean estimates	826.0207 16	482.41346 3	20722.56166	2727.330143	23176.52967	1439.73165
Sb mean t- ratios	0.484684	0.22619	0.307684	0.669557	0.486714	0.60736

Table A1: Ngene® efficiency measure for the multinomial logit in the final empirical studies

		Blo	ck 1	Blo	Block 2	
	Choice Set	Alternative	Alternative	Alternative	Alternative	
		1	2	1	2	
	Detection	3	14	3	14	
1	risk	1	0	0	10	
	Cost	1	3	10	8	
	Detection	10	7	7	10	
2	risk	20	1	1	10	
	Cost	1	3	8	10	
	Detection	10	7	7	10	
3	risk	5	5	5	10	
	Cost	8	1	3	8	
	Detection	3	3	14	10	
4	risk	0	20	20	0	
	Cost	10	1	1	8	
	Detection	14	3	10	7	
5	risk	10	1	5	1	
	Cost	8	10	3	10	
	Detection	7	3	7	3	
6*	risk	1	20	1	20	
	Cost	1	8	1	8	
	Detection	10	3	3	14	
7	risk	1	20	20	10	
	Cost	10	8	1	3	
	Detection	7	7	7	10	
8	risk	10	5	0	5	
	Cost	3	10	10	3	
	Detection	3	14	10	7	
9	risk	20	1	10	20	
	Cost	8	10	8	1	
	Detection	14	3	7	10	
10	risk	1	0	5	5	
	Cost	3	1	3	8	
	Detection	14	14	14	3	
11	risk	0	20	10	0	
	Cost	10	3	1	1	

Detection: 3=3%; 7=7%; 10=10%; 14=14%Risk: 0=0%; 1=1%; 5=5%; 10=10%; 20=20%Cost: $1=\pounds100$ ($\pounds20$ /screen); $3=\pounds250$ ($\pounds50$ /screen); $8=\pounds750$ ($\pounds150$ /screen); $10=\pounds1,000$

(£200/screen)

*Internal validity test

Appendix 3.16: The DCE survey used for the final empirical studies

MANCHESTER 1824 The University of Manchester

Start

Which breast screening programme would you prefer?

A survey to find out how you balance the risks and benefits in breast screening.

Click the button below to enter the survey.

Manchester Centre for Health Economics



National Institute for Health Research hool for Primary Care Research



Consent

We are interested in how women understand the different risks and benefits of a national breast screening programme.

You have been invited to take part in this survey because, as a woman aged between 18 and 70, you may have considered, or may consider in the future, taking part in a breast screening programme.

The law called the Data Protection Act (1998) tells us how to keep your information secure. If you provide personal information (name, address, contact details) it will remain confidential and we will not give your details to anyone else.

The main outcome of this research will be a PhD thesis for Caroline Vass, a student based in the Manchester Centre for Health Economics at The University of Manchester. In addition, we may want to report the findings at conferences or in a published journal article.

The survey will take around 20 minutes to complete.

I have read the participant information above and wish to participate in this study.

Your answers to this survey will remain strictly confidential. Unlike other surveys you may have answered, the results will only be used by The University of Manchester for a project supervised by the NHS.

No results or data will ever be passed to a private company or organisation.

If you have any questions, please contact:

Caroline Vass Manchester Centre for Health Economics Jean McFarlane Building The University of Manchester Oxford Road Manchester M13 9PL <u>Caroline.vass@postgrad.manchester.ac.uk</u> Tel. 077 1675 1262

100% 0%

NHS National Institute for Health Research ichool for Primary Care Research



BackgroundA What age category do you fall into? BackgroundA BackgroundA BackgroundA BackgroundA BackgroundA 50+ Composition Co

National Institute for Health Research chool for Primary Care Research



Intro

What is breast cancer and what is breast screening?

Breast cancer is a group of cells which can grow in an abnormal way and form into a lump (known as a cancer or tumour). As the lump grows, cancerous cells can spread to other parts of the body and this can be very serious.

Breast screening is a method of detecting breast cancer at a very early stage. Screening involves an xray of each breast (a mammogram) which is taken while carefully compressing the breast. Screening for breast cancer has been shown to be safe although most women find it a bit uncomfortable, and a few find it painful.

The x-ray can detect small changes in breast tissue which may indicate a very early stage cancer. These cancers are generally too small to be detected by touch. If there is something unusual on the x-ray then the woman will be asked to return for further tests to check whether this is a cancer.



National Institute for Health Research chool for Primary Care Research



INTRO2

As mentioned previously, the x-ray images are not always clear and sometimes further tests are needed to confirm if cancerous cells are present. These tests usually require a small sample of tissue (a biopsy) to be taken from the breast by a doctor, using a long needle, under local anaesthetic. Even when there is a tumour, sometimes these small lumps are treated with surgery and chemotherapy when really they would not have grown into something dangerous. It is impossible for doctors to tell which tumours will become life-threatening so everything identified in screening is treated.

This means that screening for breast cancer can lead to further tests and procedures which may have been unnecessary. There will be a delay between the tests and hearing the results which might be stressful for you. Also, the procedures, such as the biopsy, can be uncomfortable or painful.



NHS National Institute for Health Research



INTRO3pic

This figure shows the different options that could happen in screening.



National Institute for Health Research



VIDEO

Here is a short video explaining what happens in a mammography. This video is available on NHS England's website should you require any more information or wish to watch it again later.



National Institute for Health Research



intro4

In the UK, women over the age of 50 are offered screening up until the age of 70. Although the service is free, women who participate in screening incur some cost. For example, they may have to take some time off work, travel to the centre for breast screening, pay for parking and transport, or maybe pay for childcare.

There are **benefits** to breast screening for cancer which means tumours are picked up earlier and the long-term health outcomes of women are improved. However, there are **risks** because it is inconvenient and can mean women undergo procedures unnecessarily.

In this survey we are interested in understanding how women balance these benefits and risks. To do that, we would now like you to complete this survey where you will be shown different screening programmes. You will be asked to choose your preferred programme from a set of possible screening programmes.

There are no wrong or right answers; we are just interested in your views



National Institute for Health Research School for Primary Care Research



IntroRISK

Risk is a term used to explain the chance that something bad might happen. In breast cancer screening, there is a chance that women might have an unnecessary follow-up after their mammogram. This will not happen to all women, but there is a risk it could happen to some women.

The women shaped figures represent the total number of women having screening. The purple coloured figure shows the one woman who went for screening and had an unnecessary follow-up. This means 1 out of every 100 women screened was followed-up when she didn't need to be, but 99 women were treated appropriately.





A risk of 1% means that 1 out of every 100 women screened was followed-up when she didn't need to be, but 99 women were treated appropriately.



NHS National Institute for Health Research



IntroRISK1

This grid shows that 5 out of every 100 women screened was followed-up when she didn't need to be, but 95 women were treated appropriately.



NHS National Institute for Health Research

OR



IntroNORISK2



Similarly, a risk of 5% means that 5 out of every 100 women screened was followed-up when she didn't need to be, but 95 women were treated appropriately.







The University of Manchester

attributesandlevs

You are going to be asked to choose between screening programmes which differ in terms of:

Women who will have cancers detected by screening This describes the number of women who will have a cancer identified by a mammogram (x-ray of the breast). This describes the number of women, who go for screening over a period of 20 years, that are identified as having a cancer.

Women who will have unnecessary follow-up

This shows the number of women requiring a procedure when the lump was not cancerous or would not have caused any harmful symptoms in their lifetime. The unnecessary follow-up includes biopsies and possibly further surgery and chemotherapy.

Out-of-pocket cost

This describes the cost to you of participating in a screening programme. There are no plans to introduce charges for screening with the NHS. This cost reflects your out-of-pocket expenses incurred when participating in the screening programme. For example, transport costs, car parking, child care, time off work. In the UK, women will be screened 5 times over a 20 year period (a lifetime programme). We'll show the cost of the programmes over your life time and what that works amounts to per screen.

Each screening programme we show you will be described in terms of these three characteristics.

Each characteristic, however, has four different levels that will change.



National Insti Health R



detectedbyscreen

Women who will have cancers detected by screening



	$\langle $	
0%		100%

NHS National Institute for Health Research School for Primary Care Research

OR



detectedbyscreenNORISK

Women who will have cancers detected by screening



means that if 100 women are screened for 20 years, then 3 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 7 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 10 women will be identified as having cancer.

means that if 100 women are screened for 20 years, then 14 women will be identified as having cancer.






unnecessary

Women who will have an unnecessary follow-up



unnecessaryNORISK

Women who will have an unnecessary follow-up

0% 1% 5% 10% 20%

means that if 100 women are screened for 20 years, then 0 women will be followed-up unnecessarily.

means that if 100 women are screened for 20 years, then 1 woman will be followed-up unnecessarily.

means that if 100 women are screened for 20 years, then 5 women will be followed-up unnecessarily.

means that if 100 women are screened for 20 years, then 10 women will be followed-up unnecessarily.

means that if 100 women are screened for 20 years, then 20 women will be followed-up unnecessarily.



National Institute for Health Research



costs

Out-of-pocket cost to you of screening programme per year



£20 (or £100 over your lifetime)



£50 (or £250 over your lifetime)



£150 (or £750 over your lifetime)



£200 (or £1,000 over your lifetime)



NALIONAL INSTITUTE FOR Health Research



The University of Manchester

Exampleprog

Below is an example of one possible breast screening programme.





NHS National Institute for Health Research

OR

MANCHESTER 1824
The University of Manchester

ExampleprogNORISK

Below is an example of one possible breast screening programme.



National Institute for Health Research ichool for Primary Care Research



Examplechoice

Below is an example of the type of question we will ask you.

If these were your only options, which would you choose?

	Programme A	Programme B	No Screening
Women who will have cancers detected by screening	3% ***	14%	None: no cancers detected
Women who will have an unnecessary follow-up	0%	0%	None: no unnecessary follow-up
Out-of-pocket cost to you of screening programme per screen	£20 per screen (£100 over your lifetime)	E50 per screen (£250 over your lifetime)	None: no cost to you

You simply select the programme which you prefer.

		$\langle \neg \neg \rangle$	
	100%		0%
National Institut Medila Rese			

OR



ExamplechoiceNORISK

Below is an example of the type of question we will ask you.

Programme A	Programme B	No Screening
3%	14%	None: no cancers detected
		None:
0%	0%	no unnecessary follow-ups
£20 per screen	£50 per screen	None:
(£100 over your lifetime)	(£250 over your lifetime)	no cost to you
	Programme A 3% 0% £20 per screen (£100 over your lifetime)	Programme A Programme B 3% 14% 0% 0% E20 per screen (£100 over your lifetime) E50 per screen (£250 over your lifetime)

You simply select the programme which you prefer.





STARTDCE

Please press the "next" arrow below to begin the survey.



National Institute for Health Research School for Primary Care Research



DCE_Random1

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

(1 of 11 choice questions)

	Programme A	Programme B	No Screening
Women who will have cancers detected by screening	3%	14%	None: no cancers detected
Women who will have an unnecessary follow-up	1%	0%	None: no unnecessary follow-up
Out-of-pocket cost to you of screening programme per screen	£20 per screen (£100 over your lifetime)	£50 per screen (£250 over your lifetime)	None: no cost to you
	DCE_Random1=1	DCE_Random1=2	DCE_Random1=3

100%

National Institute for Health Research



0%



The University of Manchester

DCEnorisk_Random1

If these were your only options, which would you choose? Choose by clicking one of the buttons below:

(1 of 11 choice questions)

Women who will have cancers	20/	7 404	None:
screening	3%	14%	no cancers detected
Women who will have an unecessary follow-up	1%	0%	None: no unnecessary follow- ups
Out-of-pocket cost to you of screening programme per screen	E20 per screen D over your lifetime)	E50 per screen (£250 over your lifetime)	None: no cost to you
DC	Enorisk_Random1=1	DCEnorisk_Random1=2	DCEnorisk_Random1=3





Endofchoicetask

Thank you for providing your choices. You will now be asked some questions about the survey and a few questions about yourself.



National Institute for Health Research



FeedbackA

How confident are you that you would make the same choices if faced with the situations in real-life? FeedbackA=1 Very confident I would make the same choices FeedbackA=2 Quite confident I would make the same choices FeedbackA=3 Not confident I would make the same choices FeedbackA=4 Confident I would make different choices

lease explain your	answer		
		$\langle \neg \neg \rangle$	
	0%		100%

NHS National Institute for Health Research heal for Primary Care Research



The University of Manchester

FeedbackB

On a scale of 1-5, how easy or difficult did you find making choices between the alternatives?

1 (Very easy)	2	3 (Neither difficult nor easy)	4	5 (Very difficult)
FeedbackB_r1=1	FeedbackB_r1=2	FeedbackB_r1=3	FeedbackB_r1=4	FeedbackB_r1=5

FeedbackBexpl

Please explain your answer

$\langle \rightarrow \rangle$	

0% 100%

National Institute for Health Research chool for Primary Care Research



FeedbackB1

On a scale of 1-5, how easy or difficult did you find the survey to understand?

1		3 (Neither difficult		5
(Very easy)	2	nor easy)	4	(Very difficult)
FeedbackB1_r1=1	FeedbackB1_r1=2	FeedbackB1_r1=3	FeedbackB1_r1=4	FeedbackB1_r1=5

FeedbackB2

Was there anything in the survey that was particularly unclear?

FeedbackB2=1	No, I understood everything	
FeedbackB2=2	FeedbackB2_2_other There were some (please comment in the box below)	things that weren't clear or well explained

FeedbackBexpl1

Is there anything else about the survey we could change to make it easier to understand (e.g phrasing, explanations)?



National Institute for Health Research theol for Primary Care Research



FeedbackANA

Did you find yourself making choices based on one or two characteristics of the screening programme?

FeedbackANA=1 Yes, I based choices on one or two characteristics

FeedbackANA=2 No, I considered ALL characteristics

FeedbackANAdrop

FeedbackANAdrop1

FeedbackANAdrop2

Would you say you looked at the characteristic <u>cost</u>?
Please select

FeedbackANAexpl

 \cup

If you based your choices on one or two characteristics, why was this? (*Please tick all that apply*):

FeedbackANAexpl_1 There was just too many characteristics on which to make a decision

FeedbackANAexpl_2 The other characteristic(s) were too complicated or it was unclear what it meant

FeedbackANAexpl_3 FeedbackANAexpl_3_other Other reason (please explain)

	$\langle \neg \neg \rangle$	
0%		100%

NHS National Institute for Health Research



BackgroundB

What is your occupational status?



BackgroundC

What is your occupation? (Or previous occupation if retired?)

	· · · · ·	*	
- [
1			
		~	
		5	

0% 100%

National Institute for Health Research bool for Primary Care Research



BackgroundD

What is the highest level of education you have obtained?



National Institute for Health Research

	MANCHESTER	
	The University of Manchester	
BackgroundE		
What is your et	thnic group?	
BackgroundE=1	White British/Irish	
BackgroundE=2	White other	
BackgroundE=3	Mixed	
BackgroundE=4	Black/Black British	
BackgroundE=5	Asian/Asian British	
BackgroundE=6	BackgroundE_6_other Other (please specify)	
	$\langle \rangle$	
	0% 100%	

National Institute for Health Research School for Primary Care Research



BackgroundF

What is your religious group? BackgroundF=1 No religion BackgroundF=2 Christian BackgroundF=3 Buddhist BackgroundF=5 Hindu BackgroundF=6 Muslim BackgroundF=7 Sikh BackgroundF=8 BackgroundF_8_other Other (please specify)

> NHS National Institute for Health Research



BackgroundG

BackgroundG=1 BackgroundG=2 BackgroundG=2 No

	$\langle -$	
0%		100%

National Institute for Health Research hool for Primary Care Research



BackgroundH	

Do you have any daughters?

ackgroundH=2	No	
\bigcirc		

0% 100%

NAIS National Institute for Health Research ichool for Primary Care Research



ExperienceA

Have you ever had breast screening (a mammogram)?



ExperienceB

Have you received follow-up (such as an ultrasound, a repeat mammogram or a biopsy) for breast cancer?

ExperienceB=1	Yes
ExperienceB=2	No
ExperienceB=3	Don't know

	$\langle \neg \neg \rangle$	
0%		100%

National Institute for Health Research chool for Primary Care Research



transporttoscreen

Thinking about the last time you went for breast screening, what was your <u>main</u> form of transport to the screening?



costofscreen

Still thinking about your last breast screen, how much did it cost you for your return journey (even if your start and end points were different)? (If you did not pay any transport fare then please enter 0)

timeoffwork

Did you incur any other "costs"?

Wotkował institute for Howith Amagani School for Primary Care Research



ExperienceC

Have you, or has anyone in your family (blood relative or otherwise) had breast cancer?



ExperienceC1

Do you know of anybody else (e.g friends or colleagues) who has had breast cancer?

ExperienceC1=1	Yes	
ExperienceC1=2	No	
ExperienceC1=3	Don't know	
	0%	100%

National Institute fo Health Research Chool for Primary Care Research



The University of Manchester



RiskD1

How would you describe your future risk of breast cancer?



RiskE

To what extent are you concerned about your own risk of breast cancer?



RiskF

Is there anything else you would like to tell us about your risk of breast cancer?



Wational Institute for Hould's Beanard hould for Primary Care Research



The University of Manchester

FeelRiskA

On a scale of 1-5, how much do you agree with the following statements?

1 (Strongly disagree)	2 (Disagree)	3 (Neither agree nor disagree)	4 (Agree)	5 (Strongly agree)
FeelRiskA_r1=1	FeelRiskA_r1=2	FeelRiskA_r1=3	FeelRiskA_r1=4	FeelRiskA_r1=5

FeelRiskB

"I get a thrill when I	namble, even if it's	just the lottery"
i yeca unin when i	gamble, even in it s	just the lottery

	3				
1 2 (Neither agree nor				4	5
	(Strongly disagree)	(Disagree)	disagree)	(Agree)	(Strongly agree)
	FeelRisk8_r1=1	FeelRiskB_r1=2	FeelRiskB_r1=3	FeelRiskB_r1=4	FeelRiskB_r1=5
	\sim	0	\sim		

FeelRiskC

"Even if something is unlikely, I feel it will probably happen to me"

1 (Strongly disagree)	2 (Disagree)	3 (Neither agree nor disagree)	4 (Agree)	5 (Strongly agree)
FeelRiskC_r1=1	FeelRiskC_r1=2	FeelRiskC_r1=3	FeelRiskC_r1=4	FeelRiskC_r1=5
		$\langle \rangle$		

Wet/ovial Institute for Health Beanaith Ichool for Primary Care Research



NumeracyA

This next section will ask you some questions to understand how familiar you are with probabilities.

If you are unsure of the answer, you can skip the question.

Imagine I flip a coin 1,000 times. What is your best guess about how many times the coin would fall heads up in 1,000 flips?

NumeracyB

If there was a lottery where the chance of winning £10 is 1%, if 1,000 people bought a ticket, how many people would you expect to win?

NumeracyC

In another lottery, the chance of winning a car is 1 in 1,000. What percentage of tickets in the lottery will win a car?

0%		100%

Wotional Institute for Houlth Research choil for Primary Care Research



The University of Manchester

num2

You missed some of the previous questions about statistics and probability. Why was this?

0	Just too hard
num2=2	Were going to take me too long to answer
num2=3	Can't see the relevance to the study
num2=4	num2_4_other Other
	0%

Wet/oval Institute for Health Besnard



Comments2

Can you think of any thing else which might help us understand how you made your choices in the choice task?

	$\langle \rangle$	
0%		100%

Wational Institute for Health Bearaith



help

Thank you for completing this survey.

If you have been affected by any of the issues raised in this survey, the following organisations may be of assistance:

- National Hereditary Breast Cancer Helpline on 01629 813000 - Breast Cancer Care on 0808 800 6000

If you have any questions or concerns please contact Caroline Vass on 077 1675 1262 or caroline.vass@postgrad.manchester.ac.uk.

Manchester Centre for Health Economics, Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL

0%		100%

National Instit Health &

Appendix 3.17: Quantitative results of the DCE from the internet panel in the second pilot study

 Table A1: Response statistics

	Overall	Icon arrays	Percentages only
Completed to termination	58	30	29
Completed choice questions	59	28	31

Table A2: Effect of attributes on choice probabilities and testing for significance of the scale parameter (heteroskedastic conditional logit)

Effect of attributes on choice probabilities: pooled model		
Detect	0.107***	(0.13)
Risk	-0.013***	(0.01)
Cost	-0.070	(0.00)
IAP*detect	0.016***	(0.01)
IAP*risk	-0.028***	(0.01)
Observations	1914	
Scale term: IAP	7.574**	(2.43)

*p<0.05 ** p<0.01 *** p<0.001; Standard errors in parentheses

Appendix 4.1: ResearchNow® response rate in the large internet panel DCE

	n*	Unique panellists*
Emails Selected:	38,251	33,065
Emails Delivered:	23,082	23,082
Survey Loads:	2,205	2,205
Completes:	1,257	1,257
Screen Outs:	78	78
Quota Full:	511	511
Closed Survey:	60	60
Drop Outs:	336	336
Response Rate:	9.55%	9.55%
Qualifying completion rate:	57.01%	57.01%
Hit Rate:	5.45%	5.45%
Screenout rate	3.54%	3.54%
QuotaFull rate	23.17%	23.17%
Drop out Rate:	15.24%	15.24%
Incidence:	94.16%	94.16%

*Cannot be separated from pilot studies described in Chapter Five Response rate = (survey loads / emails delivered) Qualified Compl. rate = (completes / survey loads)

Hit rate = (completes / emails delivered) Drop out rate = (drop out / survey loads) Screenout rate = (Screenouts / survey loads) QuotaFull rate = (Quota Full / survey loads) Incidence = (completes / (completes + screenouts))

	Overall (both risk communication formats)				РО			IAP		
	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative	
Age group										
18-24	104	10.22%	10.22%	44	8.68%	8.68%	60	11.74%	11.74%	
25-34	103	10.12%	20.33%	51	10.06%	18.74%	52	10.18%	21.92%	
35-44	253	24.85%	45.19%	127	25.05%	43.79%	126	24.66%	46.58%	
45-49	253	24.85%	70.04%	131	25.84%	69.63%	122	23.87%	70.45%	
50+	305	29.96%	100.00%	154	30.37%	100.00%	151	29.55%	100.00%	
Ethnicity										
White British/Irish	872	86.42%	86.42%	435	86.65%	86.65%	437	86.19%	86.19%	
White other	53	5.25%	91.67%	27	5.38%	92.03%	26	5.13%	91.32%	
Mixed	14	1.39%	93.06%	7	1.39%	93.43%	7	1.38%	92.70%	
Black/Black British	16	1.59%	94.65%	7	1.39%	94.82%	9	1.78%	94.48%	
Asian/Asian British	45	4.46%	99.11%	22	4.38%	99.20%	23	4.54%	99.01%	
Other	9	0.89%	100.00%	4	0.80%	100.00%	5	0.99%	100.00%	
Religious Group										
No religion	382	37.90%	37.90%	188	37.45%	37.45%	194	38.34%	38.34%	
Christian	556	55.16%	93.06%	272	54.18%	91.63%	284	56.13%	94.47%	
Buddhist	7	0.69%	93.75%	5	1.00%	92.63%	2	0.40%	94.86%	
Jewish	10	0.99%	94.74%	6	1.20%	93.82%	4	0.79%	95.65%	
Hindu	10	0.99%	95.73%	5	1.00%	94.82%	5	0.99%	96.64%	
Muslim	19	1.88%	97.62%	10	1.99%	96.81%	9	1.78%	98.42%	
Sikh	4	0.40%	98.02%	3	0.60%	97.41%	1	0.20%	98.62%	
Other	20	1.98%	100.00%	13	2.59%	100.00%	7	1.38%	100.00%	
Occupational Status										
Employed full-time	343	33.99%	33.99%	167	33.27%	33.27%	176	34.71%	34.71%	
Employed part-time	211	20.91%	54.91%	114	22.71%	55.98%	97	19.13%	53.85%	
Self-employed	58	5.75%	60.65%	31	6.18%	62.15%	27	5.33%	59.17%	
Unemployed	30	2.97%	63.63%	17	3.39%	65.54%	13	2.56%	61.74%	
Retired	133	13.18%	76.81%	67	13.35%	78.88%	66	13.02%	74.75%	
Looking after home/ family	137	13.58%	90.39%	67	13.35%	92.23%	70	13.81%	88.56%	
Student	58	5.75%	96.13%	25	4.98%	97.21%	33	6.51%	95.07%	
Freelance/ temping	1	0.10%	96.23%	1	0.20%	97.41%	0	0.00%	95.07%	
Long-term sickness	37	3.67%	99.90%	12	2.39%	99.80%	25	4.93%	100.00%	
Temporarily laid off	1	0.10%	100.00%	1	0.20%	100.00%	0	0.00%	100.00%	

Appendix 4.2: Sample characteristics of respondents who completed the DCE in the final study

	Overall (both risk communication formats)				РО			IAP		
	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative	
Have children		0		A V	0		A V			
No	380	37.33%	37.33%	194	38.26%	38.26%	186	36.40%	36.40%	
Yes	638	62.67%	100.00%	313	61.74%	100.00%	325	63.60%	100.00%	
Education										
No formal qualifications	40	3.96%	3.96%	21	4.18%	4.18%	19	3.75%	3.75%	
1-4 O-levels	110	10.90%	14.87%	53	10.56%	14.74%	57	11.24%	14.99%	
5+ O-levels	112	11.10%	25.97%	54	10.76%	25.50%	58	11.44%	26.43%	
NVQs	105	10.41%	36.37%	57	11.35%	36.85%	48	9.47%	35.90%	
A-levels/AS	203	20.12%	56.49%	108	21.51%	58.37%	95	18.74%	54.64%	
Undergraduate degree	270	26.76%	83.25%	129	25.70%	84.06%	141	27.81%	82.45%	
Master's	88	8.72%	91.97%	48	9.56%	93.63%	40	7.89%	90.34%	
PhD	11	1.09%	93.06%	4	0.80%	94.42%	7	1.38%	91.72%	
Other qualification	70	6.94%	100.00%	28	5.58%	100.00%	42	8.28%	100.00%	
Experience mammogram										
Yes	436	43.25%	43.25%	217	43.23%	43.23%	219	43.28%	43.28%	
No	569	56.45%	99.70%	284	56.57%	99.80%	285	56.32%	99.60%	
Don't know	3	0.30%	100.00%	1	0.20%	100.00%	2	0.40%	100.00%	
Experienced follow-up										
Yes	148	14.68%	14.68%	75	14.94%	14.94%	73	14.43%	14.43%	
No	853	84.62%	99.31%	423	84.26%	99.20%	430	84.98%	99.41%	
Don't know	7	0.69%	100.00%	4	0.80%	100.00%	3	0.59%	100.00%	
Experience of breast cancer in f	family									
Yes	358	35.52%	35.52%	172	34.26%	34.26%	186	36.76%	36.76%	
No	618	61.31%	96.83%	318	63.35%	97.61%	300	59.29%	96.05%	
Don't know	32	3.17%	100.00%	12	2.39%	100.00%	20	3.95%	100.00%	
Experience of breast cancer wit	th friends									
Yes	701	69.54%	69.54%	356	70.92%	70.92%	345	68.18%	68.18%	
No	291	28.87%	98.41%	141	28.09%	99.00%	150	29.64%	97.83%	
Don't know	16	1.59%	100.00%	5	1.00%	100.00%	11	2.17%	100.00%	
Perceived current risk										
High or very high	17	1.69%	1.69%	6	1.20%	1.20%	11	2.17%	2.17%	
Quite high/ above average	86	8.54%	10.23%	41	8.18%	9.38%	45	8.89%	11.07%	
Average	474	47.07%	57.30%	246	49.10%	58.48%	228	45.06%	56.13%	
Slight or low	217	21.55%	78.85%	102	20.36%	78.84%	115	22.73%	78.85%	
No idea	213	21.15%	100.00%	106	21.16%	100.00%	107	21.15%	100.00%	

	Overall (both risk communication formats)				РО			IAP	
	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative	Frequency	Percentage	Cumulative
Perceived future risk									
High or very high	23	2.28%	2.28%	11	2.20%	2.20%	12	2.37%	2.37%
Quite high/ above average	98	9.73%	12.02%	48	9.58%	11.78%	50	9.88%	12.25%
Average	499	49.55%	61.57%	252	50.30%	62.08%	247	48.81%	61.07%
Slight or low	174	17.28%	78.85%	84	16.77%	78.84%	90	17.79%	78.85%
No idea	213	21.15%	100.00%	106	21.16%	100.00%	107	21.15%	100.00%
Concern about risk of breast c	cancer								
Not at all	132	13.11%	13.11%	62	12.38%	12.38%	70	13.83%	13.83%
A little	556	55.21%	68.32%	283	56.49%	68.86%	273	53.95%	67.79%
Quite a lot	162	16.09%	84.41%	77	15.37%	84.23%	85	16.80%	84.58%
Very much	93	9.24%	93.64%	47	9.38%	93.61%	46	9.09%	93.68%
No idea	64	6.36%	100.00%	32	6.39%	100.00%	32	6.32%	100.00%
Task difficulty									
Very easy	346	34.19%	34.19%	172	34.19%	34.19%	174	34.18%	34.18%
Easy	299	29.55%	63.74%	155	30.82%	65.01%	144	28.29%	62.48%
Neither easy nor difficult	260	25.69%	89.43%	125	24.85%	89.86%	135	26.52%	89.00%
Difficult	97	9.58%	99.01%	45	8.95%	98.81%	52	10.22%	99.21%
Very difficult	10	0.99%	100.00%	6	1.19%	100.00%	4	0.79%	100.00%

Survey risk version	Parameter estimate	Standard errors
Constant	0.291	(0.56)
18-24 (base category)	-	-
25-34	-0.494	(0.34)
35-44	-0.624*	(0.32)
45-49	-0.760*	(0.33)
50+	-0.755*	(0.37)
No formal qualifications (base category)	-	-
1-4 O-levels	0.272	(0.38)
5+ O-levels	0.178	(0.38)
NVQs	-0.070	(0.39)
A-levels/AS	-0.075	(0.37)
Undergraduate degree	0.204	(0.36)
Master's	-0.042	(0.41)
PhD	0.758	(0.73)
Other qualification	0.520	(0.42)
Employed full-time (base category)	-	-
Employed part-time	-0.281	(0.19)
Self-employed	-0.105	(0.30)
Unemployed	-0.382	(0.41)
Retired	-0.115	(0.26)
Looking after home/ family	-0.113	(0.22)
Student	-0.045	(0.37)
Freelance/ temping	0.000	(.)
Long-term sickness	0.705	(0.38)
Temporarily laid off	0.000	(.)
White British/Irish (base category)	-	-
White other	-0.093	(0.30)
Mixed	-0.005	(0.57)
Black/Black British	0.256	(0.52)
Asian/Asian British	0.343	(0.50)
Other	0.368	(0.69)
No religion (base category)	-	-
Christian	0.042	(0.14)
Buddhist	-1.179	(0.87)
Jewish	-0.377	(0.67)
Hindu	-0.495	(0.83)
Muslim	-0.569	(0.62)
Sikh	-1.366	(1.27)
Other	-0.592	(0.50)
Have children (dummy)	0.301	(0.16)
Experienced mammogram (dummy)	0.139	(0.20)
Experienced follow-up (dummy)	-0.084	(0.21)
Experienced breast cancer (dummy)	-0.065	(0.15)
Perceived current risk high (dummy)	-0.059	(0.15)
Perceived future risk high (dummy)	0.071	(0.14)
Concerned about risk (dummy)	-0.005	(0.07)

Appendix 4.3: Results of a simple logit regression to determine if any covariates were
a significant predictor of risk communication format

Survey risk version	Parameter estimate	Standard errors	
Found task difficult (dummy)	0.035	(0.06)	
Reported ANA to detect attribute (dummy)	-0.191	(0.41)	
Reported ANA to risk attribute (dummy)	0.306	(0.33)	
Reported ANA to cost attribute (dummy)	0.029	(0.19)	
Observations	1005		

* p<0.005; ** p<0.01; ***p<0.001

Appendix 4.4: Latent class analysis results from 6-preference-class (no scale-classes) model for the large internet panel DCE

Attributes	Pc-1	Pc-2	Pc-3	Pc-4	Pc-5	Pc-6
ASC (none)	-2.252***	- 2.083***	- 3.801***	- 0.874***	0.352**	- 1.612***
	(0.580)	(0.107)	(0.387)	(0.137)	(0.139)	(0.134)
Detect	0.481***	0.122** *	0.033** *	-0.003	-0.030*	0.022*
	(0.027)	(0.015)	(0.011)	(0.013)	(0.027)	(0.015)
Risk	-0.048***	- 0.201***	- 0.037***	0.011*	- 0.104***	- 0.076***
	(0.008)	(0.014)	(0.008)	(0.010)	(0.019)	(0.014)
Cost	-0.088***	- 0.062***	- 0.373***	0.039	- 0.185***	- 0.647***
	(0.014)	(0.018)	(0.026)	(0.022)	(0.033)	(0.051)
Preference class proportions	31.0%	24.1%	18.8%	9.7%	8.3%	8.2%

Appendix 5.1: Systematic review of qualitative research alongside healthcare DCEs

Background

Recent systematic reviews of published DCEs have found the designs are becoming increasingly complex with an increase in the number of choice sets presented and an exponential rise in the number containing complex attributes such as time or risk (de Bekker-Grob et al., 2012; Harrison et al., 2014; Ryan & Gerard, 2003b). The increased complexity of DCEs increases the potential for anomalous or inexplicable choices. As a result, a number of studies have explored the implications for the quantitative analyses; such as the inclusion or exclusion of responses which have failed tests for monotonicity or transitivity and the consequences of attribute non-attendance (Lancsar & Louviere, 2006; Lagarde, 2012). The use of tests for consistency in responses has also increased. A review by de Bekker-Grob et al. (2012) found most studies used some sort of internal validity tests to identify transitivity or continuity in choices. In health DCEs, in particular, the choice scenario presented may be about a good or service very unfamiliar to the respondent and the attributes or levels difficult to understand. Any increases in the cognitive burden of the task could potentially result in poor quality data and should be considered carefully (Hall et al., 2004).

While DCEs involve statistical analysis of quantitative choice data, other research areas, particularly in the social sciences, have used qualitative research methods to investigate people's beliefs and decision-making processes (Berg, 2007). Common data collection methods used include focus-groups, interviews or field-notes, which are usually transcribed and coded (Silverman, 1993). The process of coding transcripts and analysing the qualitative data depends on the theoretical approach chosen by the researchers (Green & Thorogood, 2004).

Although not frequently employed by economists, there is a growing body of research supporting the use of qualitative methods in health economics (Coast, 1999; Coast et al., 2004). In the context of DCEs, qualitative research methods have a number of potential roles. This could include identifying attributes and/or their associated levels, refining the terminology, cognitive piloting of the survey, exploring the use of aids and vignettes, gaining an understanding of respondents' decision-making processes, or estimating preference heterogeneity before modelling the data (Coast & Horrocks, 2007; de Bekker-Grob et al., 2012; Kløjgaard et al., 2012).

There is evidence that other stated-preference methods have also benefited from the use of

qualitative research in order to gain a deeper-understanding of their results (Baker et al., 2008). For example, a contingent valuation study used qualitative research methods to explain respondents' WTP valuations and the reasons motivating their responses (Chilton & Hutchinson, 2003). Likewise, a standard gamble study gained additional insight into people's understanding of the choice set through qualitative research (Baker & Robinson, 2004). In addition, guidelines for conducting DCEs advocate qualitative methods with specific recommendations for the development of attributes and levels (Bridges et al., 2011b; Coast & Horrocks, 2007; Lancsar & Louviere, 2008). However, there has been no explicit investigation of how well these recommendations regarding the use of qualitative methods in the preparation and testing of DCEs have been translated into practice.

Aims and objectives

The aim of this component of the thesis was to systematically identify all published studies in healthcare that reported the use of qualitative methods to inform the design and/or the interpretation of DCEs. The objectives were to: summarise the proportion of DCEs using qualitative methods; assess the context in which the research was applied; identify the methods and techniques used; and, where possible, appraise the quality of the research conducted.

Methods

This study used systematic review methods as advised by the Centre for Reviews and Dissemination (CRD) to identify all healthcare DCEs published in the last decade (since, and including, a previous systematic review) (de Bekker-Grob et al., 2012). Systematic reviews involve the identification of relevant studies, an appraisal of their quality, and a summary of evidence for a specific research questions in an explicit and methodical manner (Khan et al., 2003). These steps make systematic reviews distinctly different to literature reviews or commentaries which may not be standardised or rigorous in their approach.

The systematic review focussed on identifying DCEs rather than other stated preference methods such as CJA, ACA or contingent valuation because, as discussed in Chapter Two, these methods are grounded in different economic theories and are therefore not directly relevant to this review or the overall thesis (Louviere et al., 2010).

In line with previous systematic reviews, this review defined qualitative research methods as any exploration of peoples' thoughts or feelings through the collection of verbal or textual data (Baker et al., 2008). This definition did not include restrictions on free-text comments in DCE studies (although there is debate about whether these constitute qualitative research methods (O'Cathain & Thomas, 2004)).
An electronic search of Medline (Ovid, 1966 to date) was conducted in June 2012. Although other databases could have been searched, the strategy exactly replicated that of published reviews of DCEs (Ryan & Gerard, 2003b; de Bekker-Grob et al., 2012). The search terms used were: 'discrete choice experiment(s)', 'discrete choice model(l)ing', 'stated preference', 'part-worth utilities', 'functional measurement', 'paired comparisons', 'pairwise choices', 'conjoint analysis', 'conjoint measurement', 'conjoint studies', and 'conjoint choice experiment(s)'. The term 'conjoint analysis' was included to identify studies which had used discrete choices rather than those which required respondents to rate or rank alternatives. No search terms were used to directly identify qualitative studies as this was deemed to be too restrictive.

Table A1 shows a summary of the inclusion and exclusion criteria used in this review. The primary inclusion criteria were healthcare related and discrete choices (where respondents do not rank or rate, and there are no adaptive elements to the experimental design). Other literatures, such as environment, transport or food, were also excluded. Non-English articles and reviews, guidelines or protocols were not included. Following the initial screening, if an article could not be rejected with certainty on the basis of its abstract, the full text of the article was obtained for further evaluation. Retrieved papers were excluded if, despite the abstract, they still did not meet the inclusion criteria. Abstract screening was conducted by an initial reviewer (CV) and duplicated by a second reviewer (KP). The reviewers met to discuss results and papers which were disputed for inclusion were retrieved for further assessment. Papers were reviewed a second time to identify any articles relating to the same piece of research, thus limiting the problem of double counting a single study.

	Inclusion Criteria	Exclusion Criteria
Study focus	Healthcare interventions; healthcare services; healthcare jobs	Environmental; food; transport
Experiment type	Discrete choices; empirical study	Ranking or rating of choices; BWS; clinical trials; guidelines or reviews; discussion papers
Publication type	International and British studies; English language; published between 2002 and 2012	Foreign languages

Table A1:	Inclusion	and	exclusion	criteria	for the	review.
-----------	-----------	-----	-----------	----------	---------	---------

The main foci of the review were to: 1) identify and quantify the proportion of DCEs using qualitative methods; 2) investigate the stages in the DCE at which qualitative research is employed; 3) understand the methods and techniques currently used; and 4) where possible, evaluate the quality of the reporting of research. The studies were initially

categorised into three categories: 1) those which reported **no** qualitative research; 2) those which contained **basic** reporting which indicated some qualitative research may have been used; and 3) those which indicated an **extensive** qualitative component was conducted in direct relation to the DCE. This categorisation identified studies, in category three, which contained sufficient detail for critical appraisal. The categories are defined in Table A2.

	Aims	Methods	Analysis	Results
Definition	Clear statement	Indication of the	A description of	An explanation of
	of the purpose of	technique used	how the qualitative	the outcomes of
	the qualitative	to collect	data was	the qualitative
	research	qualitative data	examined/ software used	research
None	×	×	×	×
Basic	√ *	√ *	√ *	√ *
Extensive	\checkmark	\checkmark	\checkmark	\checkmark

Table A2: Categorisation of studies by level of reporting.

*Basic reporting to include at least one from: aims, methods, analysis, results.

Appraisal of studies

If a paper reported information on the aims, methods, analysis **and** results, they were deemed to contain 'extensive qualitative' research, suitable for a formal critical appraisal. The categorisation of studies was initially conducted by CV and repeated by two other researchers (Martin Eden and Eleanor Heather). Using qualitative research methods alongside DCEs is unusual and would not necessarily lead to any 'question formulation' but instead could reasonably be conducted with a specific aim (for example, informing attributes or checking the design). Similarly, data collected for informing attributes or levels would need to be reduced to a few key themes in contrast to the traditional 'expansionary' nature of qualitative data analysis.

Using pre-existing guides for appraisal of qualitative research methods in a traditional sense may have meant the extensive studies identified by this review would have been judged incorrectly or unfairly. It is, however, crucial that the qualitative research contained in the studies was formally assessed in a standardised and systematic way. Therefore a bespoke appraisal tool was developed to include broader issues which were not included in the traditional tools as advised by the CRD (CRD, 2008).

The tool was developed from a comparison of different appraisal forms (See: Critical Appraisal Skills Programme (CASP), 2006; Dixon-Woods, 2004; Long & Godfrey, 2004; NCDDR, 2004; Popay, Rogers, & Williams, 1998; Walsh & Downe, 2006) as recommended in the CRD guidelines for systematic reviews of qualitative research (CRD, 2008). In addition to the tools suggested by the CRD, the Joanna Briggs Institute

Qualitative Assessment and Review Instrument (JBI QARI see Pearson & Field (2007)) was also included in the comparison on the recommendation of experienced qualitative researchers in the department (personal communication with Dr Gavin Daker-White in January 2012).

As a starting point, the identified critical appraisal tools were summarised in a table to enable direct comparison of the suggested quality assessment criteria and coverage of the existing generic appraisal tools (the table can be found in Appendix 5.4). The iterative process involved two researchers (CV and Martin Eden) independently reviewing the tabulated list of quality assessment criteria and selecting criteria that matched the requirements for appraising the use of qualitative research methods in stated preference studies. The aim of this process was to identify commonalities and differences in the assessment criteria in a structured format. From this table, the relevance of each tool and quality appraisal criteria in relation to stated preference studies was established through an independent initial assessment and subsequent discussion by the two researchers to achieve consensus. Problems and factors discovered when testing the existing tools were used to focus these discussions and identify the relevance of the appraisal criteria for use in stated preference studies.

Results

Search results

One hundred and twenty four studies were already identified by a previous systematic review (de Bekker-Grob et al., 2012). The search resulted in 501 titles and abstracts since the previous review (2008 onwards) and 208 full papers were retrieved for further assessment and 148 papers met the inclusion criteria. Figure A1 shows the stages involved in screening and the reasons for rejection of the excluded papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) review guide (Moher et al., 2009). Therefore, in total, there were 277 papers included in the final review, which related to 254 empirical studies (because some studies were reported in more than one paper).

Each study was assigned a specific identity (ID). Previous reviews by de Bekker-Grob et al. (2012) and Ryan & Gerard (2003) also labelled studies with an ID which referred to either a single paper or group of papers of the same DCE. The ID also describes the amount of qualitative research reported in the paper(s) with the prefix A, B and C referring to no qualitative research reported, basic qualitative research reported or extensive qualitative research reported, respectively. The identified studies and associated ID are presented in Appendix 5.6.

Sixty two of the screened papers were rejected by CV and KP at the first stage because they were not related to healthcare. A further 232 papers were rejected because they did not contain an example of an empirical DCE (36 involved rating or ranking the alternatives; 11 were ACA studies; 22 were guidelines or overviews to conducting DCEs; and three related to prospective studies which had not begun). Ten of the papers were not in English, 27 were already covered by the previous systematic review, and 22 were duplicated in the search results. A detailed description of the reasons for rejection can be found in Appendix 5.3. Figure A1: Flow of studies through the systematic review



Critical appraisal tool

The existing appraisal tools covered eight broad themes: 1) background and overview; 2) description of the context; 3) sampling methods; 4) data collection; 5) reflexivity (or a consideration of how the investigators themselves can influence qualitative research); 6) ethical aspects; 7) data analysis; and 8) presentation of results. Within these eight themes, a set of 32 specific criteria were identified. Each tool differed in terms of the number of criteria included, ranging from a relatively small number of seven or eight criteria up to a more extensive list containing 44 items.

These existing tools were deemed inappropriate for a number of reasons. They failed to distinguish between the different study designs and the specific theoretical approaches for mixed methods studies; the appraisal tools were not validated for all qualitative approaches and the specific application for this thesis was not considered by any existent tool. Discrepancies in criteria were evident when the available appraisal tools were considered as a whole. There was also no 'best' all-rounder, resulting in a trade-off in some criteria (for example, emphasising the importance of congruity in the JBI QARI tool at the expense of detailed information about the analysis undertook as covered in CASP).

These existing tools failed to take into account the style of published papers in health economics journals. Stated preference studies tend to be published in health economics journals and are therefore not subject to the same rigorous peer-review for the qualitative component. Consequently, the details required about the philosophical foundations as found in the Long tool (Long & Godfrey, 2004), would penalise some good quality qualitative research in a stated preference study. Similarly, some tools (such as Long & Godfrey (2004) and Popay et al. (1998)) placed too much emphasis on technicalities such as the theoretical framework, the epistemological and ontological foundations, and evidence of conceptual adequacy. Details of these criteria may not be sufficiently explained in a stated preference paper where readers are unlikely to be familiar with qualitative philosophies or terms. To avoid penalising for a 'tick-box' exercise, these criteria were deemed unimportant.

In some respects the identified tools were deemed too specific; however, there were also some key criteria missing. When testing the tools on example papers it became apparent that additional criteria should be added with two key concepts in mind: 1) qualitative research methods are often employed in the design phase of stated preference studies and 2) qualitative work is conducted in this area in order to understand and provide evidence for validity of findings from stated preference studies. These two concepts both have implications for the quantitative findings: 1) reducing the data so that a few key attributes and levels are identified and that these are framed and presented in a way that respondents can easily understand; and 2) determining the extent to which results are in accordance with underlying economic theory and rationale for choice-based studies, and how this impacts on model selection, econometric tests or assumptions.

The bespoke tool was developed using an iterative process. As part of this trialling process additional criteria deemed important but absent from existing tools were added. These criteria were based on emphasising the importance of determining whether or not the qualitative component had explicitly been added to the study with the objective of informing study design and/or understanding of the stated preference study was considered paramount. Consequently, criteria were included in the tool which seek to highlight whether a clear intention and/or application of qualitative findings is discernible. Preference elicitation methods require multi-level synthesis of both qualitative and quantitative data which are often collected as separate streams of textual and choice data, respectively. How these inform each other in the development of the survey or the interpretation of results is unique and was not adequately covered in the existing tools. Although CASP (2006) asks: "How valuable is the research?" and the Long & Godfrey (2004) checklist covered "What are the implications for policy?" it was not believed these questions sufficiently covered studies who used the qualitative component to progress the DCE design. For example, how were many themes reduced to just a few attributes or levels?

Therefore the new appraisal tool allowed for parallel, multi-level synthesis of the qualitative data into their study and a clear description of how this was achieved was also considered to be a key criterion. How the results helped the design or interpretation of the stated preference study specifically was identified as an important consideration. A clear explanation of the consequential interpretation and influence was also included as a new criterion in the bespoke tool.

A further key recommendation to emerge from the development discussions was that the appraisal tool should be designed so as to encourage users to fully consider each criterion within a 'checklist' format. Therefore, the existing and additional criteria were formed into questions under a set of key headings. A researcher using the new critical appraisal tool could then elicit a yes/no answer for each criterion, which was then supported by an

additional 'comments' section.

A preliminary version of the tool was tested in a pre-selected sample of published DCEs studies which were in known existence before the review commenced (Cheraghi-Sohi et al., 2007; Ryan et al., 2009). Data were extracted from the published articles by two researchers (CV and Martin Eden) in accordance with the proposed quality assessment criteria of the appraisal tool. Face-to-face meetings between researchers were used to discuss the feasibility and acceptability of using the appraisal tool and further refinements were applied as a result of discussions. Appendix 5.5 shows the final version of the bespoke appraisal tool, and Appendix 5.9 presents an article describing the tool currently submitted to the journal Health Economics.

Identified DCEs

Details about the included studies and tabulation of the extracted data can be found in Appendix 5.7.

There was an exponential increase in the number of DCEs published over time, with over half of the studies (n=154, 56%) published since 2009. Half of the DCEs identified by this review were published in health services research journals such as Health Policy and Planning; Health Expectations; and Social Science and Medicine (n=139, 50%). A third were published in specialised medical journals (n=88) with the most popular areas being obstetrics and gynaecology (n=13, 15%), respiratory diseases (n=13, 15%) and oncology (n=9, 10%). The remaining articles were covered by general medical journals (n=31, 11%) or other areas (such as risk analysis). A full breakdown of journals publishing the DCEs can be found in Appendix 5.7.

Over half the DCEs published were conducted in Europe (n=186, 56%) and a quarter of the DCEs identified were carried out in the UK (n=84, 25%). Other popular countries included the United States of America (USA) (n=49, 15%), the Netherlands (n=38, 11%), Australia (n=26, 8%) and Canada (n=19, 6%). Nineteen studies (8%) asked residents of multiple countries to complete their DCE and two studies gathered data from delegates at an international conference thus incorporating a wide range of nationalities (A7, A73). Of the studies conducted within the UK (both solely and those looking at multiple countries) most (n=43, 51%) had a focus on primary care services or conditions usually treated by GPs. The settings of other countries were not included in this review due to inconsistencies in the definitions of 'primary care' across international healthcare systems.

For example, in some European countries, gynaecology is classified as a primary care service as women can attend a community clinic without referral and as a first point of contact.

Reporting of qualitative research methods

Overall, 111 (44%) studies did not report the use of any qualitative research methods; 114 (45%) reported minimal data on the use of qualitative methods; and 29 (11%) reported, or explicitly cited, the extensive use of qualitative methods. The trends in publishing qualitative research generally reflected the increases in the overall number of DCEs as demonstrated by Figure A2.





⁺²⁰¹² incomplete due to the year of search. *Overall includes papers rather than studies. None refers to studies which indicated no qualitative research. Basic refers to studies which indicated some qualitative research methods were employed. Extensive refers to studies which reported the qualitative component of their research in detail.

As shown in Figure A3, a variety of applications of qualitative research were identified. Applications included: the selection of attributes and/or levels (n=95, 66%); piloting the DCE (n=26, 18%); and understanding respondents' decision-making processes (n=4, 3%). Popular qualitative research methods included focus groups (n=64, 45%) and interviews (n=108, 76%). Semi-structured interviews (n=25, 17%), structured interviews (n=5, 3%) and cognitive debriefing techniques (n=12, 8%) were the most common approaches.



Figure A3: Potential use of qualitative research methods with DCEs

The next section will discuss in detail how these results were reported with either a basic or extensive description of the work undertaken.

Basic qualitative research

The following section describes the qualitative component of studies which indicated that they had used qualitative research methods alongside their DCE study but provided insufficient details to appraise the study.

Level of reporting

Almost all authors who reported using some qualitative research did so by stating in the methods section of the paper the nature of the qualitative component of their research (n=113, 99%). One study only mentioned that qualitative research was conducted but gave no indication of the collection methods used (B78), and two studies indicated that qualitative research was conducted in the design of the DCE but gave no further details about the context (B24, B34).

Sampling

The 16 studies (14%) which reported using qualitative research methods prior to implementing the DCE survey chose to sample from a different population to that finally used in their DCE study. Most often the qualitative research used a sample of researchers or healthcare professionals before distributing the final DCE to patients or the public.

Context

Almost all (n=113, 99%) of the studies which reported basic qualitative research reported using it before the DCE was implemented, either in the design or piloting phase. None of the studies reported using qualitative research methods alongside the final survey, although three studies reported using qualitative research at the end of the DCE to attain additional information on preferences (B83, B32, B25).

A variety of applications of qualitative research methods were identified. In the design of the DCE, researchers were most commonly seeking to identify attributes and/or assign levels (n=70, 61%), or validate attributes and/or levels identified through other methods (n=31, 27%). Researchers also used qualitative research methods more specifically to check terminology, vignettes and descriptions (n=9, 8%) and to confirm translations (n=2, 2%).

After the design phase, some studies also reported using qualitative research methods in the piloting of the DCE (n=24, 21%). In the pilot stage, the methods were specifically used to check for decision strategies (n=1, 1%) and also to determine an appropriate sample for

the final DCE. For example, study B5 used interviews to determine an appropriate age range for the final DCE. One study (B49) used the qualitative data acquired in the piloting stage to estimate preference heterogeneity and thus predict an appropriate model for the choice data, and another study (B11) used the qualitative research to predict the signs of the coefficients.

Data collection methods

Within the studies that had used basic reporting of qualitative methods, the most popular approach to qualitative data collection was interviews (n=89, 78%), comprising a range of techniques including structured and semi-structured interviews and focus groups. Ten studies also employed cognitive interviews which included debriefing questions at the end of the task as well as a verbal protocol analytical technique called 'think-aloud'. Focus groups were also another popular approach of data collection (n=50, 44%).

Analysis of qualitative data

Although a crucial step in drawing reliable and valid results from the qualitative data, only a minority of studies mentioned anything about the approach to the analysis of the qualitative data (n=15, 7%). Of these 15 studies, five reported using content analysis (B22, B25, B35, B51, B106) and two studies using framework analysis (B29, B31). Other analytical approaches included the use of grounded theory methods such as the constant comparative method (B33) and open-ended coding (B113). Three studies detailed the use of specialist qualitative software: two used NVivo® (B25, B65); and one used Atlas.ti® (B112). Table A3 shows the methods and context of the qualitative research reported in the DCEs. Table A3: Summary of methods and context of the qualitative research reported in DCEs

Methods	Studies (n)	%
Cognitive interviews	10	9%
Focus groups	50	44%
Free-text comments	2	2%
General interviews	59	52%
Semi-structured interviews	12	11%
Structured interviews	5	4%
Telephone interviews	3	3%
Total	141*	
Context		
To add additional information about preferences to support results	4	4%
To check for decision strategies	1	1%
To check terminology and descriptions	9	8%
To check translations	2	2%
To identify attributes and/or levels	70	61%
To inform general DCE design	3	3%
To inform the quantitative analysis	2	2%
To inform the sampling strategy	1	1%
To pilot/pre-test the DCE	24	21%
To validate attributes and/or levels identified through other methods	31	27%
Total	147*	

*Some studies used multiple methods or employed qualitative research methods in multiple contexts.

Extensive qualitative research

The following sections describes the qualitative component of studies which provided extensive details about the qualitative component of their study stating the aims, methods, analysis and results either in the text or with a citation to other work.

Level of reporting

Seven studies extensively described the use of qualitative research within the main text of the paper. Twenty-two further studies were identified as having conducted extensive qualitative research by checking the references to the qualitative component of the work. The cited qualitative research was published in a variety of sources including peer-reviewed journals (n=17) and National Institute for Health Research (NIHR) commissioned reports (n=4).

Context

Most studies (n=25, 86%) reported the use of qualitative research methods to identify attributes and/or levels for use in the DCE. Three studies (C19, C9 and C4) used qualitative research methods to understand more about how respondents completed the choice task presented. Two studies (C19 and C12) also used the qualitative methods as a

complement to the quantitative analysis. Other studies used qualitative research methods to pilot and refine the survey (C10, C15).

Data collection methods

The most common data collection approach was interviews (n=18, 62%). These interviews were mostly semi-structured (n=12, 41%) and face-to-face, although two studies used telephone interviews (C7 and C18). Of the three studies using qualitative research to understand more about how people completed the DCE task, two of these (C4 and C9) used an approach called think-aloud.

A number of studies also used focus groups or group interviews (n=12, 41%), and four studies used a combination of focus groups and interviews in their qualitative study (C22, C24, C14 and C27). One study (C10) used the results of an ethnographic direct observational study to identify attributes and levels for the DCE, and used semi-structured interviews to refine the training materials and descriptions.

Analysis of qualitative data

Most studies simply stated in the paper that they used thematic analysis (n=10, 34%) or content analysis (n=5, 17%) to categorise the qualitative data collected. One study also reported the use of a 'latent' content approach to discover underlying themes (C25). A type of thematic analysis, thematic synthesis was reported by C29 which involves a more explicit refinement of themes (possibly from multiple studies) and is an approach in line with reducing the qualitative data to develop a few attributes and levels. Other analytical approaches included framework analysis (n=3, 10%) and a related approach to qualitative analysis called charting (C4). Seven studies used some constant comparative analysis (n=4, 14%) or open-coding (n=3, 10%) at least in the initial stages. Two studies (C18 and C17) used interpretative phenomenological analysis (IPA) which often takes an open-coding approach rather than relying on pre-existing themes or frameworks.

The type of software used was not always reported but the most commonly reported packages were NVivo® (n=4, 14%) and Atlas.ti® (n=2, 7%). One study (C8) also used NUDist®, a software related to NVivo®. The citation of the qualitative research (either the main-text of the DCE or a previous publication); the application; the methods employed; and the analysis conducted are described in Table A4. The results of the qualitative analysis and how these influenced (or did not influence) the DCE study was unique to each paper and cannot readily be quantified or summarise.

ID, country, cited research	Methods		Context		Analysis
C16 (UK)	Focus groups		To identify attributes and level	ls	Thematic analysis
C18 (Multi-country)	Open-ended interviews	Open-ended telephone interviews	To identify attributes and levels	To compare to quantitative results	Frequency analysis and IPA
C19 (Germany)	Unstructured interv	views	To understand how respondent behaviour)	ts complete the choice task (trading	Thematic analysis
C9 (UK)	Think-aloud interv	iews	To understand how respondent behaviour)	ts complete the choice task (trading	Coded using a literature derived framework
C5 (USA) See also: Fitzpatrick et al. (2007); Fitzpatrick et al. (2008)	Semi-structured int	terviews	To identify attributes and level	ls	Constant comparative method in Atlas.ti® and 'open-coding'
C10 (UK) See also: Salisbury et al. (2007)	Ethnographical observation study	Semi-structured interviews	To identify attributes and levels	To refine descriptions of attributes	Constant comparative method in Atlas.ti®
C17 (Canada) See also: Grindrod et al. (2009)	Focus groups		To identify attributes and level	ls	IPA content analysis
C3 (Australia) See also: Haas et al. (2001)	Semi-structured int	terviews	To identify attributes and level	ls	Thematic analysis
C6 (Multi-country) See also: Haughney et al. (2004)	Semi-structured int	terviews	To identify attributes and level	ls	Thematic analysis
C13 (USA) See also: Herbild (2007)	Focus groups		To identify attributes and level	ls	Thematic analysis
C20 (USA) See also: Hsieh et al. (2010)	Focus groups		To identify attributes and level	ls	Content analysis using Atlas.ti®
C21 (Ghana) See also: Smith Paintain et al. (2011)	Semi-structured int	terviews	To identify attributes and level	ls	Content analysis using NVivo®
C11 (UK) See also: Nafees et al. (2006)	Focus groups		To validate attributes and leve	ls	Content analysis
C1 (USA) See also: Mark et al. (2003)	Focus groups		To identify attributes and level	ls	Thematic analysis
C28 (Australia) See also: Morton et al. (2010)	Semi-structured int	terviews	To identify attributes and level	ls	Thematic synthesis

Table A4: Description of the qualitative research contained in extensively detailed studies

ID, country, cited research	Methods		Context	Analysis
C29 (Australia) See also: Naik Panvelkar et al. (2010)	Semi-structured in	terviews	To identify attributes and levels	Thematic analysis
C22 (UK) See also: Fargher et al. (2007)	Semi-structured interviews	Focus groups	To identify attributes and levels	Constant comparative analysis
C12 (UK)	Focus groups		To identify attributes and levels Inform quantitative (subgroup) analysis	Constant comparative analysis
C23 (UK) See also: Netten et al. (2009)	Interviews		To identify attributes and levels	Thematic analysis
C24 (Vietnam) See also: Nghi et al. (2010)	Interviews	Focus groups	To identify attributes and levels	Theme-based code-book in NVivo®
C14 (UK) See also: Kennedy et al. (2005)	Focus groups	Semi-structured interviews	To identify attributes and levels	Framework analysis
C2 (UK)	Focus groups		To identify attributes and levels	Thematic analysis
C15 (Australia) See also: Netten et al. (2002)	Interviews		Piloting and pre-testing	Content analysis
C4 (UK)	Think-aloud interv	iews	To understand how respondents complete the choice task (trading behaviour)	Charting approach
C25 (UK) See also: Cheyne et al. (2007)	Group interviews		To identify attributes and levels	Latent content analysis
C7 (Australia) See also: Bryan et al. (2007)	Semi-structured interviews	Semi-structured interviews (telephone)	To identify attributes and levels	Open coding
C8 (UK) See also: Baker et al. (2006)	Semi-structured in	terviews	To identify attributes and levels	Open coding (moving to structured) using NUDist software and NVivo®
C26 (Multi-Country) See also: Dancet et al. (2011)	Focus groups		To identify attributes and levels	Constant comparative content analysis with a coding tree
C27 (Vietnam) See also: Witter et al. (2011)	Semi-structured interviews	Focus groups	To identify attributes and levels	Coding in NVivo®

'See also' directs to the cited qualitative research related to the DCE study.

Discussion

This systematic review identified all published health DCEs since 2003 and assessed the qualitative research reported in the paper. A total of 254 relevant studies were identified. A key finding of the systematic review was that a large proportion of DCE studies either do not conduct, or fail to explicitly report, using qualitative research methods. Some studies did acknowledge the lack of qualitative research was a study limitation.

Although the systematic review found that the most common application of qualitative research was to select attributes and levels for use in the DCE, other applications were also identified. Qualitative research was also frequently used in the pre-testing or piloting the DCE survey, and for refining or checking terminology. The review also found some studies using qualitative research methods in other applications. For example, to predict preference heterogeneity, select and specify a regression model, to identify the motives behind 'irrational responses', or to specifically test for breaks in the key axioms which support DCEs as a method (Kjaer & Gyrd-Hansen, 2008; Kjaer et al., 2006; Ryan et al., 2009). In light of these broad ranging applications, it is apparent that qualitative research methods are being used in ways beyond the advice of general conduct instructions or even specific guidelines.

The systematic review was the first systematic review to conduct an in-depth evaluation of the different applications, methods and quality of qualitative research reported in DCEs. Previous reviews of health DCEs have provided overviews of the design and analyses or assessed their content (Ryan & Gerard, 2003b; de Bekker-Grob et al., 2012; Harrison et al., 2014). Although de Bekker-Grob et al. (2012) gave a brief summary of qualitative research conducted alongside DCEs, but no details on the aims, methods or analysis were extracted from their papers.

Since the systematic review was conducted, another systematic review of DCEs in health has been published. Clark et al. (2014) also updated the review de Bekker-Grob et al. (2012) and running an identical search strategy, but again there was a more general focus and qualitative research methods were only briefly summarised.

Furthermore, the review by Clark et al. (2014) suffered from a number of serious limitations. The review focussed on the literature published in 2009-2012 but missed several DCE studies published in 2008 after the review by de Bekker-Grob et al. (2012). For example, studies A40, B43, B46 and B47 identified in the systematic review reported above were not included. Whilst missing some studies, the review by Clark et al. (2014) also suffered from double-counting as it included papers (such as B58 and A64) covered by the earlier review that they sought to update (de Bekker-Grob et al., 2012). Additionally, Clark et al. (2014) included some conjoint rating studies which did not meet their own inclusion criteria (such as Waltzman et al. 2011; and Bederman et al. 2010). A BWS type-2 study was also included in an apparent error (see Gunther et al. 2010).

Most importantly, the results of the systematic review identified almost sixty DCE studies not identified by Clark et al. (2014). A list of the studies included by this systematic review but missed by Clark et al. (2014) can be found in Appendix 5.8. It is unknown why these papers were not included by Clark et al. (2014).

Limitations

A limitation of the systematic review was the focus on papers recorded in one database, Medline. This search strategy was chosen because it updated a previously published review by de Bekker-Grob et al. (2012) and replicated their study. The authors of the review chose Medline as other databases such as Pubmed or Embase identified duplicate papers rather than missing studies. The authors state in their review: *"It was expected to identify the large majority of the health-related DCE studies published during the period"* (p. 146). A second limitation of the review is the lack of secondary data extraction. Abstract screening was repeated by a second researcher and the categorisation of papers into the extent of use of qualitative research methods was conducted by two other researchers. Final data extraction was conducted by CV only.

One of the most important limitations of the systematic review in this thesis was the reliance on what was reported in the published paper. It could be that rigorous qualitative research was being extensively conducted but the details were never reported in the final journal article, perhaps because of word restrictions. As a rebuttal to this limitation, a follow-up survey to authors of DCEs included in the review was conducted and is presented in Appendix 5.2.

Conclusion

The systematic review identified two papers (Cheraghi-Sohi et al., 2007; Ryan et al., 2009) which used concurrent verbal protocol analysis (called think-aloud) to understand more about people's choices. The think-aloud method was considered appropriate to answer the key research questions about understanding how respondents' complete a DCE.

There was a paucity of detail about the analysis of the qualitative data collected, however, when described, most studies used some sort of thematic analysis, which was most often a pre-defined framework to code the collected data from which themes were developed using NVivo® software. Therefore the analysis of the think-aloud data could follow a similar approach.

Bibliography

Baker, R., Freeman, G., Boulton, M., Brookes, O., Windridge, K., Tarrant, C., Low, J., Turner, D., Hutton, E. and Bryan, S. (2006). Continuity of Care: patients' and carers' views and choices in their use of primary care services. National Co-ordinating Centre for NHS Service Delivery and Organisation Research and Development. SDO/13b/20.

Baker, R. and Robinson, A. (2004). Responses to standard gambles: are preferences 'well constructed'? Health Economics. 13 (1). 37–48.

Baker, R., Robinson, A. and Smith, R. (2008). How do respondents explain WTP responses? A review of the qualitative evidence. Journal of Socio-Economics. 37 (4). 1427–1442.

Bederman, S.S., Mahomed, N.N., Kreder, H.J.H., McIsaac, W.J., Coyte, P.C. and Wright, J.G. (2010). In the eye of the beholder: preferences of patients, family physicians, and surgeons for lumbar spinal surgery. Spine. 35 (1). 108–115.

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

Berg, B. (2007). Qualitative Research Methods for the Social Sciences. 7th Ed. Pearson Education.

Bridges, J.F., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L., Regier, D. a, Johnson, F. and Mauskopf, J. (2011). Conjoint analysis applications in health-a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in Health. 14 (4). 403–413.

Bryan, S., Williams, I. and Mciver, S. (2007). Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. Health Economics. 16 (2). 179–193.

CASP (2006). 10 questions to help you make sense of qualitative research (Critical Appraisal Skills Programme). Public Health Resource Unit.

Cheraghi-Sohi, S., Bower, P., Mead, N., McDonald, R., Whalley, D. and Roland, M. (2007). Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique. Family Practice. 24 (3). 276–282.

Cheyne, B.H., Terry, R., Niven, C., Dowding, D., Hundley, V. and Mcnamee, P. (2007). 'Should I come in now ?': a study of women's early labour experiences. British Journal of Midwidery. 15 (10). 604–609.

Chilton, S.M. and Hutchinson, W.G. (2003). A qualitative examination of how respondents in a contingent valuation study rationalise their WTP responses to an increase in the quantity of the environmental good. Journal of Economic Psychology. 24 (1). 65–75.

Clark, M., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E.W. (2014). Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics. 32 (9). 883–902.

Coast, J. (1999). The appropriate uses of qualitative methods in health economics. Health economics. 8 (4). 345–353.

Coast, J. and Horrocks, S.A. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of health services research & policy. 12 (1). 25–30.

Coast, J., McDonald, R. and Baker, R. (2004). Issues arising from the use of qualitative methods in health economics. Journal of health services research & policy. 9 (3). 171–6.

CRD (2008). Systematic Reviews. CRD's (Centre for Reviews and Dissemination) guidance for undertaking reviews in health care. The University of York.

Dancet, E., Van Empel, I., Rober, P., Nelen, W., Kremer, J. and D'Hooghe, T. (2011). Patientcentred infertility care: a qualitative study to listen to the patient's voice. Human Reproduction. 26 (4). 827–33.

Dixon-Woods, M. (2004). The problem of appraising qualitative research. Quality and Safety in Health Care. 13 (3). 223–225.

Fargher, E.A., Eddy, C., Newman, W.G., Qasim, F., Tricker, K., Elliott, R.A. and Payne, K. (2007). Patients' and healthcare professionals' views on pharmacogenetic testing and its future delivery in the NHS. Pharmacogenomics. 8 (11). 1511–9.

Fitzpatrick, E., Angus, D., Durieux-Smith, A., Graham, I.D. and Coyle, D. (2008). Parents' needs following identification of childhood hearing loss. American journal of audiology. 17 (1). 38–49.

Fitzpatrick, E., Graham, I.D., Durieux-Smith, A., Angus, D. and Coyle, D. (2007). Parents' perspectives on the impact of the early diagnosis of childhood hearing loss. International Journal of Audiology. 46 (2). 97–106.

Green, J. and Thorogood, N. (2004). Qualitative Methods for Health Research. London: Sage Publications.

Grindrod, K.A., Rosenthal, M., Lynd, L., Marra, C.A., Bougher, D., Wilgosh, C. and Tsuyuki, R.T. (2009). Pharmacists' perspectives on providing chronic disease management services in the community — Part I: Current practice environment. Canadian Pharmacists Journal. 142 (5). 234–239.

Gunther, O.H., Kurstein, B., Riedel-Heller, S.G. and Konig, H.-H. (2010). The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. Health services research. 45 (1). 212–229.

Hall, J., Viney, R., Haas, M. and Louviere, J. (2004). Using stated preference discrete choice modeling to evaluate health care programs. Journal of Business Research. 57 (9). 1026–1032.

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J. and Payne, K. (2014). Risk as an attribute in discrete choice experiments: a critical review. The patient. 7 (2). 151–70.

Haughney, J., Barnes, G., Partridge, M. and Cleland, J. (2004). The Living & Breathing Study: a study of patients' views of asthma and its treatment. Primary Care Respiratory Journal. 13 (1). 28–35.

Herbild, L. (2007). Designing a DCE to outlay patients ' and the publics ' preferences for genetic screening in the treatment of depression. Health Economics Working Papers, University of Southern Denmark. 1. 1–31.

Hsieh, Y.-H., Hogan, M.T., Barnes, M., Jett-Goheen, M., Huppert, J., Rompalo, A.M. and Gaydos,C. a (2010). Perceptions of an ideal point-of-care test for sexually transmitted infections--aqualitative study of focus group discussions with medical providers. PloS one. 5 (11).

Kennedy, A., Gately, C. and Rogers, A. (2005). Examination of the implementation of the Expert Patients Programme within the structures and locality contexts of the NHS in England (PREPP study) - Report II. National Primary Care Research and Development Centre.

Khan, K., Kunz, R., Kleijnen, J. and Antes, G. (2003). Five steps to conducting a systematic review. Journal of the Royal Society of Medicine. 96. 118–121.

Kjaer, T. and Gyrd-Hansen, D. (2008). Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. Health policy. 85 (1). 124–132.

Kjaer, T., Gyrd-Hansen, D. and Willaing, I. (2006). Investigating patients' preferences for cardiac rehabilitation in Denmark. International Journal of Technology Assessment in Health Care. 22 (2). 211–218.

Kløjgaard, M., Bech, M. and Søgaard, R. (2012). Designing a stated choice experiment: the value of a qualitative process. Journal of Choice Modelling. 5 (2). 1–18.

Lagarde, M. (2012). Investigating attribute non-attendance and its consequences in choice experiments with latent class models. Health Economics. 22 (5). 554–567.

Lancsar, E. and Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. PharmacoEconomics. 26 (8). 661–677.

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

Long, A. and Godfrey, M. (2004). An evaluation tool to assess the quality of qualitative research studies. International Journal of Social Research Methodology. 7 (2). 181–196.

Louviere, J., Flynn, T. and Carson, R.T. (2010). Discrete choice experiments are not conjoint analysis. Journal of Choice Modelling. 3 (3). 57–72.

Mark, T.L., Kranzler, H.R., Poole, V.H., Hagen, C., McLeod, C. and Crosse, S. (2003). Barriers to the use of medications to treat alcoholism. American Journal on Addictions. 12 (4). 281–294.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 6 (7). e1000097.

Morton, R., Devitt, J., Howard, K., Anderson, K., Snelling, P. and Cass, A. (2010). Patient views about treatment of stage 5 CKD: a qualitative analysis of semistructured interviews. American Journal of Kidney Diseases. 55 (3). 431–440.

Nafees, B., Lloyd, A., Kennedy-Martin, T. and Hynd, S. (2006). How diabetes and insulin therapy affects the lives of people with type 1 diabetes. European Diabetes Nursing. 3 (2). 92–97.

Naik Panvelkar, P., Armour, C. and Saini, B. (2010). Community pharmacy-based asthma serviceswhat do patients prefer? The Journal of Asthma. 47 (10). 1085–93.

NCDDR (2004). What are the standards for quality research? National Center for the Dissemination of Disability Research. (9).

Netten, A., Burge, P., Malley, J., Brazier, J., Flynn, T. and Forder, J. (2009). Outcomes of Social Care for Adults (OSCA) Interim Findings. Personal Social Services Research Unit. Discussion.

Netten, A., Ryan, M., Smith, P., Skatun, D., Healey, A., Knapp, M. and Wykes, T. (2002). The development of a measure of social care outcome for older people. Personal Social Services Research Unit. Discussion.

Nghi, N.Q., LaMontagne, S., Bingham, A., Rafiq, M., Phoung Mai, L.T. and Lien, N.T.P. (2010). Human papillomavirus vaccine introduction in Vietnam: formative research findings. Sexual health. 7 (3). 262–270.

O'Cathain, A. and Thomas, K.J. (2004). 'Any other comments?' Open questions on questionnaires - a bane or a bonus to research? BMC medical research methodology. 4. 25.

Pearson, A. and Field, J. (2011). JBI critical appraisal checklist for qualitative research. Joanna Briggs Institute.

Popay, J., Rogers, A. and Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. Qualitative Health Research. 8 (3). 341–351.

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

Ryan, M., Watson, V. and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of a discrete choice experiment responses. Health Economics. 18. 321–336.

Salisbury, C., Banks, J., Goodall, S., Baxter, H., Montgomery, A., Pope, C., Gerard, K., Simons, L., Lattimer, V., Sampson, F., Pickin, M., Edwards, S., Smith, H. and Boudioni, M. (2007). An evaluation of Advanced Access in general practice. National Co-ordinating Centre for NHS Service Delivery and Organisation Research and Development.

Silverman, D. (1993). Interpreting Qualitative Data. Sage Publications.

Smith Paintain, L., Antwi, G., Jones, C., Amoako, E., Adjei, R.O., Afrah, N. a, Greenwood, B., Chandramohan, D., Tagbor, H. and Webster, J. (2011). Intermittent screening and treatment versus intermittent preventive treatment of malaria in pregnancy: provider knowledge and acceptability. PloS one. 6 (8).

Walsh, D. and Downe, S. (2006). Appraising the quality of qualitative research. Midwifery. 22 (2). 108–19.

Waltzman, J.T., Scholz, T. and Evans, G. (2011). What patients look for when choosing a plastic surgeon: an assessment of patient preference by conjoint analysis. Annals of plastic surgery. 66 (6). 643–647.

Witter, S., Thi Thu Ha, B., Shengalia, B. and Vujicic, M. (2011). Understanding the 'four directions of travel': qualitative research into the factors affecting recruitment and retention of doctors in rural Vietnam. Human Resources for Health. 9 (20).

Appendix 5.2: Survey to authors of published healthcare DCEs

Background

The systematic review presented Appendix 5.1 was constrained by its reliance on the details reported in published academic papers. It is well established that a bias exists in both medicine and social science, with studies exhibiting positive or interesting results being more likely to be published (Franco et al., 2014; Easterbrook et al., 1991) A survey was, therefore, designed to elicit information from authors of published DCEs in health about their views on the use of qualitative research methods alongside quantitative analytical methods.

Aims and objectives

The aim of this study was to try and reduce the effects of reporting bias through directly eliciting researchers' views and experience of conducting qualitative research alongside DCEs in a healthcare setting. The objectives of this study were to reveal more about researchers' experience of using qualitative research methods; their opinion of the usefulness of qualitative approaches in this context; and any explanation they had for the poor level of reporting found in the systematic review.

Methods

Study sample

Authors who indicated, in a published study, that they had used qualitative research methods but only reported basic details were identified as the most appropriate sample. In choosing authors who were known to have conducted qualitative research would generate interesting results and likely achieve a higher response rate than contacting authors who did not indicate they used qualitative research methods. There seemed little advantage to contacting authors of the studies which reported extensive details of the qualitative component to their research, as a comprehensive account of this aspect of their study had already been captured in the paper and appraised using the bespoke tool.

A total of 114 studies reported basic use of qualitative research methods and all authors were included in this review. As some corresponding authors had multiple studies included in the review, 91 individual authors were sampled.

Survey design

All of the journal articles provided an email address for the corresponding author. Therefore, the most feasible method of contacting authors and eliciting their views was to use an online survey. Whilst other methods such as telephone interviews or a postal survey could have been used, acquiring other contact details for all authors and given that a large proportion were abroad would have proved difficult. Therefore an online self-administered survey was developed.

The survey was designed using Sawtooth software (Sawtooth, 2012) and hosted on The University of Manchester's secure server. The final survey consisted of 12 questions and comprised a mix of closed-ended and free-text comment boxes. The questions asked authors about their **experience** and their **opinion** of: 1) using qualitative research methods alongside DCEs; and 2) communicating the qualitative work they conducted in a journal article. Additional questions included self-assessment of their and co-authors' expertise in qualitative research, the number of DCEs they had conducted, and whether they agreed with the key finding of the systematic review that qualitative research is not well reported in healthcare DCEs.

Pilot study

A preliminary study was devised by CV with input from the supervisory team and was piloted with researchers (n=3) experienced with DCEs but not included in this review (because their DCEs were unpublished or in non-health subjects). The pilot study identified that acquiring as much information about the authors' views was crucial. Therefore the preliminary questionnaire was amended so authors had the opportunity to add further comments to all of their answers, and at the end of the survey they were encouraged to highlight anything that was missing but important in a free-text comment box. The final survey can be found in Appendix 5.10.

Data collection

The email details and the corresponding authors' details were entered into a spreadsheet and double checked with the public online profiles of authors to ensure accuracy. When an email was returned as 'undeliverable', a link was sent through alternative means (such as Academia.edu, Researchgate and LinkedIn profiles). As a last resort when no online profile existed, an alternative author listed on the relevant paper was contacted.

The authors (n=91) were invited to participate via an email (or electronic message) which explained the systematic review and included a brief abstract covering the background, methods and results. The email contained a unique link to identify authors who had failed to either accept or decline the invitation. These authors were sent a reminder after 21 days. For the purpose of the reminder emails, the questionnaire respondent could be identified through the unique link; however, all responses were treated as anonymous. Templates of

the initial email and the reminder email can be found in Appendix 5.11 and 5.12, respectively.

Data analysis

The survey data were downloaded from the online server and analysed in Excel® (Microsoft, 2010). The analysis involved simple production of descriptive statistics for each of the questions and the generation of bar graphs for a visual summarisation. The authors' free-text comments were not thematically analysed because of the limited textual data available (some authors chose not to comment). Instead these free-text comments were collated into a word document and discussed with the supervisory team to generate core themes.

Results

After the first email sent on 1st May 2013, 38 authors completed the survey (an initial response rate of 42%). A total of 53 authors who had not responded to the initial e mail were then sent a reminder after three weeks. Four authors declined to take part (for reasons such as one author had not practiced in the field for a few years so could not sufficiently recall their experiences; another was a statistician who had only been involved with the DCE analysis). The questionnaire closed on the 30th June 2013, with a total of 53 completed or partially complete responses, resulting in an overall response rate of 58%. In addition, examples from the free-text comment box are used to provide additional contextual data provided by authors. These free-text comments are presented to illuminate the quantitative findings. The following sections provide a summary of the key results. Appendix 5.13 provides a tabulated breakdown of the authors' responses to each of the survey questions. As shown in Figure A1, most of the respondents to the survey were authors who had completed at least 10 DCEs.

Figure A1: Number of DCEs published as first author or co-author



Figure A2 shows that all the respondents perceived that using qualitative research methods had added value to the DCE they published. The authors also reported that the use of qualitative research methods added value to their experience of conducting DCEs in general, with 74% (n=31) stating it made a 'substantial improvement' to the study.





One respondent offered this comment, which suggests some antipathy towards the use of qualitative research methods in the DCE:

"Qualitative methods often require a subjective component that doesn't fit well with economics or quantitative methods. I am not convinced that qualitative work is always needed" (Author ID40; published 5-9 DCEs)

A key finding of the systematic review was poor reporting of qualitative research. The majority of survey respondents (n=42, 79%) agreed with this finding that the qualitative component was only briefly described in their DCE paper. Half of the respondents (n=26, 50%) believed the amount of qualitative work conducted was accurately reflected in the published paper as reported.

Figure A3 summarises the response of authors when explicitly asked about the reasons

behind the key finding of the systematic review (presented in Appendix 5.1) that qualitative research is not well reported in DCE papers. Almost half of authors (n=22, 44%) said they did not think the detail on qualitative research methods was of interest to journals:

"The subjects were asked WHY they made that choice. This yielded wonderful insight into the beliefs and misconceptions behind prevention choices. I haven't found a journal that is interested in publishing these more descriptive results" (Author ID22; published 3 DCEs)

"I have suggested to editors to incorporate a material section in the webpage including transcripts of the focus groups, and in general all the qualitative work, plus data and econometric model. This would allow replicability and learn more about real advantages and disadvantages of the methodology. Congrats for choosing such a relevant research topic and hopefully your efforts will improve the journal editorial process." (Author ID36; published 2 DCEs) One respondent suggested that reporting the qualitative research could jeopardise the

acceptance of their paper:

"Reporting these details could cause criticism by the reviewers, and

compromise acceptance of publication." (Author ID27; published 5-9 DCEs) Some respondents (n=11, 16%) stated that qualitative research would not be of interest to their peers. For example:

"Not all the work that was conducted is reported in the paper because of length limitations imposed by the Journal and because of the complexity of the topic to the main audience." (Author ID27; published 5-9 DCEs)

Some (n=9, 17%) of the survey respondents also used the free-text space to comment that the journal word limit was too low to accurately incorporate the qualitative work that was conducted as this comment illustrates:

"Journal space limitations made [it] impossible to report the extensive qualitative work we had to do in order to come up with a plausible valuation scenario, attributes definition and levels." (Author ID36; published 2 DCEs)

One author suggested that there has been too much focus on quantitative aspects of DCEs. For example:

"Perhaps there has been so much focus on the technical aspects of performing a DCE (design and analysis) that the most important step: identifying and reporting the methods for selecting attributes and their levels in a consistent manner has received less attention." (Author ID11; published 3 DCEs)

Figure A3 shows that over half of the respondents (n=26, 52%) considered qualitative

research was too complicated to report in detail and one fifth of the respondents (n=10, 20%) reported that they felt it was too time consuming to conduct properly.



Figure A3: Reasons for limited reporting of qualitative research methods

Two authors hypothesised that the quality of the qualitative research conducted alongside DCEs meant the work did not merit extensive reporting:

"Journals usually do not allow for much reporting of this, and the qualitative studies are not necessarily conducted in a sufficient stringent approach to be reported in separate papers." (Author ID3; published 5-9 DCEs) "What is presented in a manuscript often is limited because of word limits or because some of the qualitative research is somewhat informal." (Author ID19; published 10+ DCEs)

Some respondents (n=4, 8%) also reported that they did not believe qualitative research was of importance to funders. For example:

"It may not be appreciated by funders as a necessary component of doing high quality DCE studies." (Author ID1; published 5-9 DCEs)

Three quarters of the respondents (n=40, 75%) stated that they had no expertise in qualitative research methods. Some respondents (n=31, 58%) did have a qualitative researcher as part of their team, but others (n=8, 15%) did not.

In the free-text comments, one author stated there was a lack of guidance on reporting standards for the qualitative research conducted alongside a DCE:

"Helping to disseminate guidelines/suggestions useful to report accurately but concisely details on qualitative research, in a way that it is accepted by the reviewers/readers, could help improve the transparency of the studies, hence their quality and results reliability." (Author ID27; published 5-9 DCEs)

Discussion

A variety of reasons for the lack of reporting was identified. One respondent hypothesised that the lack of qualitative research was because it was deemed unimportant in terms of answering the research question posed. Another potential reason is lack of expertise because research teams conducting DCEs have quantitative economic backgrounds. This lack of expertise could be remedied with comprehensive guidelines on conducting qualitative research alongside a DCE study. Currently, detailed guidelines on the use of qualitative research methods only exist for the identification of attributes and levels..

The survey to authors was completed by researchers who had regular experience of designing and/or analysing DCEs, with most of the respondents publishing more than one DCE. This indicates the sample was knowledgeable and was likely to include experienced researchers whose views are likely to be representative of the wider research area. The survey found that **all** authors reported that qualitative research methods played a valuable role in the progression of their DCE study. This survey provided evidence that researchers designing and analysing DCEs regarded using qualitative research methods as beneficial in a health DCE study. The lack of reporting of a beneficial and informative component to a research study could be rectified with greater use of online appendices for reporting the qualitative research, particularly in word-restrictive journals.

Comparison with existing literature

None of the previous systematic reviews (Ryan & Gerard, 2003b; de Bekker-Grob et al., 2012; Clark et al., 2014) directly contacted authors to determine whether the results detailed in their papers were subject to reporting-bias. The results of the previous reviews were therefore unconditionally reliant on the textual accounts of the author as recorded in the article. Using a follow-up survey using the authors of published studies has allowed a thorough exploration of current and past DCE practice based on their own opinions.

Limitations

The survey used for this study (Appendix 5.10) had a number of limitations. Arguably, a more in-depth account of authors' views and experiences could have been collected (possibly through one-to-one interviews) and thorough thematic analysis of the free-text comments would have provided more robust results. However, the results of the questionnaire helped to explain the key findings of the systematic review, such as the drivers behind the absence of detail, and it is unlikely further analysis or review would have highlighted anything that would significantly alter the findings.

Conclusion

The results of the systematic review and survey to authors identified qualitative research methods were being used by DCE researchers to answer multiple research questions, and that these methods add value to a DCE study. This finding suggested that qualitative research would be an appropriate method to use alongside DCEs.

Bibilography

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

Clark, M., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E.W. (2014). Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics. 32 (9). 883–902.

Easterbrook, P.J., Berlin, J. a., Gopalan, R. and Matthews, D.R. (1991). Publication bias in clinical research. Lancet. 337 (8746). 867–876.

Franco, A., Malhotra, N. and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. Science. 345 (6203). 1502–1505.

Microsoft (2010). Excel.

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

Sawtooth (2012). Sawtooth Software SSI Web 8.3.8.

Appendix 5.3: Flow of papers through the systematic review

Figure A1: Search results from Medline database for systematic review of studies using qualitative research alongside DCEs, run on the 30th June 2012

253	•Discrete Choice Experiment(s)
18	Discrete Choice Model(1)ing
143	Stated Preference
4	Part-Worth Utilities
0	Functional Measurement
710	Paired Comparisons
8	Pairwise Choices
283	•Conjoint Analysis
43	•Conjoint Measurement
2	•Conjoint Studies
2	•Conjoint experiment(s)
1403	•TOTAL HITS
902	•Published pre-2008
501	•RETRIEVED ABSTRACTS



Appendix 5.4: Comparison of existing appraisal tools for assessing the quality of qualitative research

Table A1: Comparison of overview and context criteria covered by existing appraisal tools

	Clear aims or hypothesis	Importance and relevance of research	Appropriateness of qualitative methodology	Appropriate research design for aims?	Details of phenomena being studied	Thorough and systematic literature review to link research to current knowledge	Theoretical framework for study	Study perspectives (service, user carer etc)	Discussion of time horizon of for study
Critical Appraisal Skills Programme (CASP)	\checkmark	\checkmark	\checkmark	\checkmark			\checkmark	\checkmark	
Prompts for appraising qualitative research (Dixon-Woods)	1		\checkmark	\checkmark		\checkmark			
Evaluation Tool for Qualitative Studies (Long)	\checkmark				\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Quality Framework (Focus)	\checkmark		\checkmark	\checkmark	\checkmark		\checkmark		
Rationale and Standards (Popay)									
Appraising the Quality of Qualitative Research (Walsh)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		
JBI QARI			\checkmark	\checkmark			\checkmark		

	Discussion of recruitment strategy and technique (purposive, convenience etc)	Key character- istics of sample	Why these people were selected	Details of compare- ison groups	Appropriateness of setting	Clear description of data collection method (to replicate)	Justified collection methods	Explicit description of process (topic guides etc)	Reporting of modification of methods in study	Form of data (tape records)
Critical Appraisal Skills Programme (CASP)	\checkmark		\checkmark		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Prompts for appraising qualitative research (Dixon-Woods)	V		\checkmark			✓	V	~	✓	V
Evaluation Tool for Qualitative Studies (Long)	\checkmark	\checkmark	\checkmark	\checkmark	1	\checkmark		\checkmark	\checkmark	\checkmark
Quality Framework (Focus)	\checkmark	\checkmark	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Rationale and Standards (Popay)			\checkmark		\checkmark				\checkmark	
Appraising the Quality of Qualitative Research (Walsh)			\checkmark		\checkmark	\checkmark	\checkmark		\checkmark	
JBI QARI			\checkmark				\checkmark			

Table A2: Comparison of sampling and data collection criteria covered by existing appraisal tools

Table A3: Comparison of reflexivity and ethics criteria covered by existing appraisal tools

¥	Researcher(s) have discussed their own potential bias	Study demonstrates that ethical standards were maintained/ issues addressed	Handling of data and respondents post/during study	Approval from committee					
Critical Appraisal Skills Programme (CASP)	\checkmark	\checkmark	\checkmark	 ✓ 					
Prompts for appraising qualitative research (Dixon-Woods)									
Evaluation Tool for Qualitative Studies (Long)	\checkmark	\checkmark		\checkmark					
Quality Framework (Focus)	\checkmark	\checkmark	\checkmark						
Rationale and Standards (Popay)									
Appraising the Quality of Qualitative Research (Walsh)	\checkmark	\checkmark	\checkmark	\checkmark					
JBI QARI	\checkmark	\checkmark		\checkmark					
	Description of analysis process	Derivation of themes discussed	Evidence of iterative analysis	Steps to guard against selectivity	How presented data were selected	Adequacy of data to support findings	Implication of contradictory findings	Researchers' own influences or bias in analysis	Data and interpretations clearly integrated?
---	---------------------------------------	--------------------------------------	--------------------------------------	--	--	--	---	--	---
Critical Appraisal Skills Programme (CASP)	\checkmark	\checkmark	V	\checkmark	V	\checkmark	\checkmark	v v	
Prompts for appraising qualitative research (Dixon-Woods)	\checkmark	\checkmark	\checkmark	\checkmark	~	\checkmark			\checkmark
Evaluation Tool for Qualitative Studies (Long)	\checkmark		\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	
Quality Framework (Focus)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Rationale and Standards (Popay)						\checkmark		\checkmark	\checkmark
Appraising the Quality of Qualitative Research (Walsh)	√	\checkmark	√		\checkmark	\checkmark	\checkmark	1	\checkmark
JBI QARI						\checkmark		\checkmark	\checkmark

Table A4: Comparison of data analysis criteria covered by existing appraisal tools

	Explicit stating of the findings	Adequate discussion of evidence for/against	Credibility and validity demonstrated (triangulation etc)	Findings in relation to research question	Comparison to other studies	Contribution to existing literature	Recommendation for future research	Generalisability of results	Policy and practice implications
Critical Appraisal Skills Programme (CASP)	~	√	√	√	√	√	\checkmark		~
Prompts for appraising qualitative research (Dixon-Woods)						\checkmark			\checkmark
Evaluation Tool for Qualitative Studies (Long)		\checkmark	\checkmark		\checkmark	\checkmark		\checkmark	\checkmark
Quality Framework (Focus) Rationale and Standards (Popay)	\checkmark	\checkmark	\checkmark	✓	✓	\checkmark	\checkmark	\checkmark	~
Appraising the Quality of Qualitative Research (Walsh)		\checkmark	\checkmark	✓	✓	\checkmark	\checkmark	\checkmark	V
JBI QARI									

Table A5: Comparison of results and value of research criteria covered by existing appraisal tools

Appendix 5.5: Final critical appraisal tool for assessing the quality of qualitative research alongside DCEs

A CHECKLIST FOR QUALITATIVE RESEARCH ALONGSIDE STATED PREFERENCE	METH	IODS	
			UNCLEAR/ NA/
	YES	NO	COMMENTS
OVERVIEW			
Are there clear aims, objectives or hypotheses to the qualitative component?	u	u	
Is the use of qualitative methods appropriate and justified?			
Does the research fill a gap or link to current knowledge?			
CONTEXT			
Is there a theoretical framework/approach for the study? Grounded theory, phenomenology?			
Is there a clear application for the study? Does the setting match the aims?			
Is it clear and justified why the qualitative data was collected concurrently or retrospectively?			
SAMPLING			
Is there a clear description of the recruitment strategy?			
Are details of the sample's characteristics provided?			
Does the sample contribute to the generalisability of the study? Did the recruitment result in an			
unbiased sample?			
Does the study consider the appropriateness of the selected sample? Will this match the final			
or typical stated preference survey respondents?			
Did sampling continue until data saturation?			
Is the response rate detailed? Is the response rate poor? Why?			
COLLECTION			
Are there adequate descriptions of the methods and process for data collection? Could the			
study be replicated?	_		
Are the topic guides, schedules or prompts presented in, or with, the paper?	<u> </u>	<u> </u>	
Are examples of the survey (DCE or CV) questions presented? How were preferences elicited?		•	
Is there a clear description of the form of data? Tape recordings, videos, free-text comments?			
REFLEXIVITY			
Is there a description of the qualifications or experience of the researchers involved?			
ETHICS			
Is there indication of ethical approval if the qualitative study is dealing with a sensitive subject?			
ANALYSIS			
Is the analytical process adequately described? If themes were identified, how did this happen?			
Is there evidence of triangulation? Were others sources involved in the interpretation?			
Are raw data, such as verbatim quotes, presented or made available on request?			
Is there an explanation of why the presented data were selected?			
Is there congruity between the presented data and the interpretations drawn?			
RESULTS			
Are the findings of the qualitative research clearly stated?			
Has the namer considered the credibility and validity of results? Are there reports of reflexivity?			
Are conflicting qualitative coults evalue of 2			
Pre connecting quantative results explored ?			
Do the qualitative results relate to the original research question(s)?	-		
is there a comparison of the qualitative results to existing literature or similar studies?	-		
Are recommendations for future qualitative or quantitative research made?			
Is the generalisability of the qualitative results discussed?	<u> </u>		
Is there a clear application for the results of the qualitative research? Will it change or add to	-	-	
attributes and levels in a DCE? Does this inform the starting bid for a willingness-to-pay study?			

Appendix 5.6: Studies included in the systematic review of qualitative research conducted alongside DCEs

ID Full reference

- A1 McKenzie, L., Cairns, J., and Osman, L. M. (2001). Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. *Health Policy*, *57*(3), 193-204.
- A2 Osman, L. M., McKenzie, L., Cairns, J., Friend, J. A., Godden, D. J., Legge, J., and Douglas, J. (2001). Patient weighting of importance of asthma symptoms. *Thorax*, 56(2), 138–142.
- A3 Van der Pol, M., and Cairns, J. (2001). Estimating time preferences for health using discrete choice experiments. *Social Science and Medicine*, *52*(9), 1459-1470.
- A4 Moayyedi, P., Wardman, M., Toner, J., Ryan, M., and Duffett, S. (2002). Establishing patient preferences for gastroenterology clinic reorganisation using conjoint analysis. *European Journal of Gastroenterology and Hepatology*, *14*(4), 429-433.
- A5 Scott, A., Watson, M. S., and Ross, S. (2003). Eliciting preferences of the community for out of hours care provided by general practitioners: a stated preference discrete choice experiment. *Social Science and Medicine*, *56*(4), 803-814.
- A6 Bishop, A. J., Marteau, T. M., Armstrong, D., Chitty, L. S., Longworth, L., Buxton, M. J., and Berlin, C. (2004). Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. *BJOG*, *111*(8), 775-779.
- A7 Hundley, V., and Ryan, M. (2004). Are women's expectations and preferences for intrapartum care affected by the model of care on offer?. BJOG: An International Journal of Obstetrics and Gynaecology, 111(6), 550-560.
- A8 Watson, V., Ryan, M., Brown, C. T., Barnett, G., Ellis, B. W., and Emberton, M. (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. *Journal of Urology*, *172*(1), 2321-2325.
- A9 Weston, A., and Fitzgerald, P. (2004). Discrete choice experiment to derive willingness to pay for methylaminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. *PharmacoEconomics*, 22(18), 1195-1208.
- A10 Fraenkel, L., Constantinescu, F., Oberto-Medina, M., and Wittink, D. R. (2005). Women's preferences for prevention of bone loss. *Journal of Rheumatology*, *32*(6), 1086-1092.
- A11 Haughney, J., Partridge C., M. R. Vo., Larsson, T., Kessler, R., Stahl, E., Brice, R., and Lofdahl, C. G. (2005). Exacerbations of COPD: quantifying the patient's perspective using discrete choice modelling. *The European Respiratory Journal*, *26*(4), 623-629.
- A12 Huis in 't Veld, M. H., van Til, J. A., Ijzerman, M. J., and Vollenbroek-Hutten, M. M. (2005). Preferences of general practitioners regarding an application running on a personal digital assistant in acute stroke care. *Journal of Telemedicine and Telecare*, *11*(S1), 37-39.
- A13 Lee, A., Gin, T., Lau, A. S., and Ng, F. F. (2005). A comparison of patients' and health care professionals' preferences for symptoms during immediate post-operative recovery and the management of post-operative nausea and vomiting. *Anesthesie, Analgesie, Reanimation, 100*(1), 87-93.
- A14 Lloyd, A., McIntosh, E., and Price, M. (2005). The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. *PharmacoEconomics*, 23(11), 1167-1181.

- A15 Mantovani, L. G., Monzini, M. S., Mannucci, P. M., Scalone, L., Villa, M., and Gringeri,
 A. (2005). Differences between patients', physicians' and pharmacists' preferences for
 treatment products in haemophilia: a discrete choice experiment. *Haemophilia*, 11(6), 589–597.
- A16 Ryan, M., Diack, J., Watson, V., and Smith, N. (2005). Rapid prenatal diagnostic testing for Down syndrome only or longer wait for full karyotype: The views of pregnant women. *Prenatal Diagnosis*, 25(13), 1206-1211.
- A17 Ryan, M., Major, K., and Ska[°]tun, D. (2005). Using discrete choice experiments to go beyond clinical outcomes when evaluating clinical practice. *Journal of Evaluation in Clinical Practice*, *11*(4), 328-338.
- A18 Salkeld, G, Solomon, M., Butow, P., and Short, L. (2005). Discrete-choice experiment to measure patient preferences for the surgical management of colorectal cancer. *The British Journal of Surgery*, 92(6), 742-747.
- A19 Thompson, C. A., Foster, A., Cole, I., and Dowding, D. W. (2005). Using social judgement theory to model nurses' use of clinical information in critical care education. *Nurse Education Today*, *25*, 68-77.
- A20 Baltussen, R., Stolk, E., Chisholm, D., and Aikins, M. (2006). Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Economics*, *15*(7), 689-696.
- A21 Byrne, M. M., Souchek, J., Richardson, M., and Suarez-Almazor, M. (2006).
 Racial/ethnic differences in preferences for total knee replacement surgery. *Journal of Clinical Epidemiology*, 59(10), 1078-1086.
- A22 Gerard, K., Lattimer, V., Surridge, H., George, S., Turnbull, J., Burgess, A., Lathlean, J., et al. (2006). The introduction of integrated out-of-hours arrangements in England: a discrete choice experiment of public preferences for alternative models of care. *Health Expectations*, *9*(1), 60-69.
- A23 Johnson, F. R., and Backhouse, M. (2006). Eliciting stated preferences for healthtechnology adoption criteria using paired comparisons and recommendation judgments. *Value in Health*, 9(5), 303-311.
- A24 Lewis, S. M., Cullinane, F. M., Carlin, J. B., and Halliday, J. L. (2006). Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. *Australian and New Zealand journal of obstetrics and gynaecology*, 46(3), 205-211.
- A25 Longo, M. F., Cohen, D. R., Hood, K., Edwards, A., Robling, M., Elwyn, G., and Russell, I. T. (2006). Involving patients in primary care consultations: assessing preferences using discrete choice experiments. *The British Journal of General Practice*, *56*(522), 35-42.
- A26 Osoba, D., Hsu, M. A., Copley-Merriman, C., Coombs, J., Johnson, F. R., Hauber, B., Manjunath, R., et al. (2006). Stated preferences of patients with cancer for health-related quality-of-life (HRQOL domains during treatment. *Quality of Life Research*, *15*(2), 273-283.
- A27 Rubin, G., Bate, A., George, A., Shackley, P., and Hall, N. (2006). Preferences for access to the GP: a discrete choice experiment. *The British Journal of General Practice*, *56*(531), 743-748.
- A28 Wordsworth, S., Ryan, M., Skatun, D., and Waugh, N. (2006). Women's preferences for cervical cancer screening: a study using a discrete choice experiment. *International Journal of Technology Assessment in Health Care*, 22(3), 344-350.
- A29 Bishai, D., Brice, R., Girod, I., Saleh, A., and Ehreth, J. (2007). Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine. *PharmacoEconomics*, 25(2), 143-154.
- A30 Caldon, L. J., Walters, S. J., Ratcliffe, J., and Reed, M. W. (2007). What influences clinicians' operative preferences for women with breast cancer? An application of the discrete choice experiment. *European Journal of Cancer*, *43*(11), 1662-1669.

- A31 Goto, R., Nishimura, S., and Ida, T. (2007). Discrete choice experiment of smoking cessation behaviour in Japan. *Tobacco control*, *16*(5), 336-343.
- A32 Hjelmgren, J., and Anell, A. (2007). Population preferences and choice of primary care models: a discrete choice experiment in Sweden. *Health Policy*, *83*(2-3), 314-322.
- King, M. T., Hall, J., Lancsar, E., Fiebig, D. G., Hossain, I., Reddel, H. K., and Jenkins, C. R. (2007). Patient preferences for managing asthma: results from a discrete choice experiment. *Health Economics*, 16(7), 703–717.
- A34 Langenhoff, B. S., Krabbe, P. F., and Ruers, T. J. (2007). Computer-based decision making in medicine: a model for surgery of colorectal liver metastases. *European Journal of Surgical Oncology*, *33*(S2), S111-S117.
- A35 McGregor, J. C., Harris, A. D., Furuno, J. P., Bradham, D. D., and Perencevich, E. N. (2007). Relative influence of antibiotic therapy attributes on physician choice in treating acute uncomplicated pyelonephritis. *Medical Decision Making*, 27(4), 387-394.
- A36 Schwappach, D. L., and Strasmann, T. J. (2007). Does location matter? A study of the public's preferences for surgical care provision. *Journal of Evaluation in Clinical Practice*, *13*(2), 259-264.
- A37 Seston, E. M., Ashcroft, D. M., and Griffiths, C. E. (2007). Balancing the benefits and risks of drug treatment: a stated-preference, discrete choice experiment with patients with psoriasis. *Archives of Dermatology*, *143*(9), 1175-1179.
- A38 Van der Pol, M., Shiell, A., Au, F., Johnston, D., and Tough, S. (2008). Convergent validity between a discrete choice experiment and a direct, open-ended method: Comparison of preferred attribute levels and willingness to pay estimates. *Social science and medicine*, 67(12), 2043-2050.
- A39 Aspinall, P. A., Johnson, Z. K., Azuara-Blanco, A., Montarzino, A., Brice, R., and Vickers, A. (2008). Evaluation of quality of life and priorities of patients with glaucoma. *Investigative Ophthalmology and Visual Science*, *49*(5), 1907-1915.
- A40 Brau, R., and Lippi Bruni, M. (2008). Eliciting the demand for long-term care coverage: a discrete choice modelling analysis. *Health economics*, *17*(3), 411-433.
- A41 Lee, W. C., Joshi, A. V., Woolford, S., Sumner, M., Brown, M., Hadker, N., Pashos, C. L., et al. (2008). Physicians' preferences towards coagulation factor concentrates in the treatment of Haemophilia with inhibitors: a discrete choice experiment. *Haemophilia*, *14*(3), 454–465.
- A42 McTaggart-Cowan, H. M., Shi, P., Fitzgerald, J. M., Anis, A. H., Kopec, J. A., Bai, T. R., Soon, J. A., et al. (2008). An evaluation of patients' willingness to trade symptom-free days for asthma-related treatment risks: a discrete choice experiment. *Journal of Asthma*, 45(8), 630-638.
- A43 Mortimer, D., and Segal, L. (2008). Is the value of a life or life-year saved context specific? Further evidence from a discrete choice experiment. *Cost Effectiveness and Resource Allocation*, *6*, 8.
- A44 Snoek, G. J., van Til, J. A., Krabbe, P. F., and Ijzerman, M. J. (2008). Decision for reconstructive interventions of the upper limb in individuals with tetraplegia: the effect of treatment characteristics. *Spinal Cord*, *46*(3), 228-233.
- A45 Szeinbach, S. L., Harpe, S. E., Williams, P. B., and Elhefni, H. (2008). Testing for allergic disease: parameters considered and test value. *BMC Family Practice*, *9*, 47.
- A46 Wellman, G. S., and Vidican, C. (2008). Pilot study of a hierarchical Bayes method for utility estimation in a choice-based conjoint analysis of prescription benefit plans including medication therapy management services. *Research in Social and Administrative Pharmacy*, 4(3), 218-230.

- A47 Albada, A., and Triemstra, M. (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 Expectations*, *12*(1), 92-105.
- A48 Chan, Y. M., Sahota, D. S., Leung, T. Y., Choy, K. W., Chan, O. K., and Lau, T. K. (2009). Chinese women's preferences for prenatal diagnostic procedure and their willingness to trade between procedures. *Prenatal diagnosis*, *29*(13), 1270-1276.
- A49 Chuck, A., Adamowicz, W., Jacobs, P., Ohinmaa, A., Dick, B., and Rashiq, S. (2009). The willingness to pay for reducing pain and pain-related disability. *Value in Health*, *12*(4), 498-506.
- A50 Eberth, B., Watson, V., Ryan, M., Hughes, J., and Barnett, G. (2009). Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis. *Medical Decision Making*,29(6), 707-715.
- A51 Fiebig, D. G., Haas, M., Hossain, I., Street, D. J., and Viney, R. (2009). Decisions about Pap tests: What influences women and providers?. *Social science and medicine*, 68(10), 1766-1774.
- A52 Goto, R., Takahashi, Y., Nishimura, S., and Ida, T. (2009). A cohort study to examine whether time and risk preference is related to smoking cessation success. *Addiction*, *104*(6), 1018-1024.
- A53 Howard, K., and Salkeld, G. (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.
- A54 Petrou, S., and McIntosh, E. (2009). Women's preferences for attributes of firsttrimester miscarriage management: a stated preference discrete-choice experiment. *Value in Health*, *12*(4), 551-559.
- A55 Ratcliffe, J., Brazier, J., Tsuchiya, A., Symonds, T., and Brown, M. (2009). Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire.*Health economics*, *18*(11), 1261-1276.
- A56 Ringburg, A. N., Buljac, M., Stolk, E. A., van Lieshout, E. M., van Beeck, E. F., Patka, P., and Schipper, I. B. (2009). Willingness to pay for lives saved by Helicopter Emergency Medical Services. *Prehospital Emergency Care*, *13*(1), 37-43.
- A57 Robson, S. C., Kelly, T., Howel, D., Deverill, M., Hewison, J., Lie, M. L. S., ... and May, C. R. (2009). Randomised preference trial of medical versus surgical termination of pregnancy less than 14 weeks' gestation. *Health Technology Assessment*, 13(53), i-148.
- A58 Ryan, M., and Watson, V. (2009). Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health Economics*, *18*(4), 389-401.
- A59 Skjoldborg, U. S., Lauridsen, J., and Junker, P. (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.
- A60 Tinelli, M., Ryan, M., and Bond, C. (2009). Patients' preferences for an increased pharmacist role in the management of drug therapy. *International journal of pharmacy practice*, *17*(5), 275-282.
- A61 Van Helvoort-Postulart, D., Dellaert, B. G., Van Der Weijden, T., Von Meyenfeldt, M. F., and Dirksen, C. D. (2009). Discrete choice experiments for complex health-care decisions: does hierarchical information integration offer a solution?. *Health economics*, *18*(8), 903-920.

- A62 Van Til, J. A., Stiggelbout, A. M., and Ijzerman, M. J. (2009). The effect of information on preferences stated in a choice-based conjoint analysis. *Patient education and counseling*, *74*(2), 264-271.
- A63 Watson, V., Ryan, M., and Watson, E. (2009). Valuing experience factors in the provision of chlamydia screening: an application to women attending the family planning clinic. *Value in Health*, *12*(4), 621-623.
- A64 Witt, J., Scott, A., and Osborne, R. H. (2009). Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. *Health Economics*, *18*(6), 681–696.
- A65 Ahmed, A., and Fincham, J. E. (2010). Physician office vs retail clinic: patient preferences in care seeking for minor illnesses. *The Annals of Family Medicine*,8(2), 117-123.
- A66 Bhatt, M., Currie, G. R., Christopher Auld, M., and Johnson, D. W. (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. *Academic Emergency Medicine*,17(11), 1207-1215.
- A67 Bosworth, R., Cameron, T. A., and DeShazo, J. R. (2010). Is an ounce of prevention worth a pound of cure? Comparing demand for public prevention and treatment policies. *Medical Decision Making*, *30*(4), E40-E56.
- A68 Brown, D. S., Johnson, F. R., Poulos, C., and Messonnier, M. L. (2010). Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus. *Vaccine*, 28(7), 1702-1708.
- A69 Chen, T. T., Chung, K. P., Huang, H. C., Man, L. N., and Lai, M. S. (2010). Using discrete choice experiment to elicit doctors' preferences for the report card design of diabetes care in Taiwan–a pilot study. *Journal of evaluation in clinical practice*, *16*(1), 14-20.
- A70 Davison, S. N., Kromm, S. K., and Currie, G. R. (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. *Nephrology Dialysis Transplantation*, *25*(7), 2334-41.
- A71 Deverill, M., Lancsar, E., Snaith, V. B. A., and Robson, S. C. (2010). Antenatal care for first time mothers: a discrete choice experiment of women's views on alternative packages of care. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *151*(1), 33-37.
- Hendrix, M., Pavlova, M., Nieuwenhuijze, M. J., Severens, J. L., and Nijhuis, J. G.
 (2010). Differences in preferences for obstetric care between nulliparae and their partners in the Netherlands: a discrete-choice experiment. *Journal of Psychosomatic Obstetrics and Gynecology*, 31(4), 243-251.
- A73 Kievit, W., Van Hulst, L., Van Riel, P., and Fraenkel, L. (2010). Factors that influence rheumatologists' decisions to escalate care in rheumatoid arthritis: Results from a choice-based conjoint analysis. *Arthritis care and research*, *62*(6), 842-847.
- A74 Kiiskinen, U., Suominen-Taipale, A. L., and Cairns, J. (2010). Think twice before you book? Modelling the choice of public vs private dentist in a choice experiment. *Health economics*, *19*(6), 670-682.
- A75 Kimman, M. L., Dellaert, B. G., Boersma, L. J., Lambin, P., and Dirksen, C. D. (2010). Follow-up after treatment for breast cancer: one strategy fits all? An investigation of patient preferences using a discrete choice experiment. *Acta Oncologica*, *49*(3), 328-337.
- A76 Nieboer, A. P., Koolman, X., and Stolk, E. A. (2010). Preferences for long-term care services: Willingness to pay estimates derived from a discrete choice experiment. *Social science and medicine*, 70(9), 1317-1325.

- A77 Owen, K., Pettman, T., Haas, M., Viney, R., and Misan, G. (2010). Individual preferences for diet and exercise programmes: changes over a lifestyle intervention and their link with outcomes. *Public health nutrition*, *13*(02), 245-252.
- A78 Reese, P. P., Tehrani, T., Lim, M. A., Asch, D. A., Blumberg, E. A., Simon, M. K., ... and Halpern, S. D. (2010). Determinants of the decision to accept a kidney from a donor at increased risk for blood-borne viral infection. *Clinical Journal of the American Society of Nephrology*, 5(5), 917-923.
- A79 Scuffham, P. A., Whitty, J. A., Taylor, M., and Saxby, R. C. (2010). Health system choice. *Applied health economics and health policy*, 8(2), 89-97.
- A80 Stolk, E. A., Oppe, M., Scalone, L., and Krabbe, P. F. (2010). Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. *Value in Health*, *13*(8), 1005-1013.
- A81 Van der Pol, M., and McKenzie, L. (2010). Costs and benefits of tele-endoscopy clinics in a remote location. *Journal of telemedicine and telecare*, *16*(2), 89-94.
- A82 Van Der Pol, M., Shiell, A., Au, F., Jonhston, D., and Tough, S. (2010). Eliciting individual preferences for health care: a case study of perinatal care. *Health Expectations*, *13*(1), 4-12.
- A83 Van Mello, N. M., Mol, F., Opmeer, B. C., de Bekker-Grob, E. W., Essink-Bot, M. L., Ankum, W. M., ... and Hajenius, P. J. (2010). Salpingotomy or salpingectomy in tubal ectopic pregnancy: what do women prefer?. *Reproductive biomedicine online*, 21(5), 687-693.
- A84 Wyatt, J. C., Batley, R. P., and Keen, J. (2010). GP preferences for information systems: conjoint analysis of speed, reliability, access and users. *Journal of evaluation in clinical practice*, *16*(5), 911-915.
- A85 Ahmed, A., and Fincham, J. E. (2011). Patients' view of retail clinics as a source of primary care: Boon for nurse practitioners?. *Journal of the American Academy of Nurse Practitioners*, 23(4), 193-199.
- A86 Bech, M., Kjaer, T., and Lauridsen, J. (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment.*Health economics*, 20(3), 273-286.
- A87 Bøgelund, M., Vilsbøll, T., Faber, J., Henriksen, J. E., Gjesing, R. P., and Lammert, M. (2011). Patient preferences for diabetes management among people with type 2 diabetes in Denmark-a discrete choice experiment. *Current Medical Research and Opinion*, 27(11), 2175-2183.
- A88 Boonen, L. H., Donkers, B., and Schut, F. T. (2011). Channeling consumers to preferred providers and the impact of status quo bias: does type of provider matter?. *Health services research*, *46*(2), 510-530.
- A89 Brown, T. M., Pashos, C. L., Joshi, A. V, and Lee, W. C. (2011). The perspective of patients with haemophilia with inhibitors and their care givers: Preferences for treatment characteristics. *Haemophilia*, 17(3), 476–482.
- A90 Casciano, R., Malangone, E., Ramachandran, A., and Gagliardino, J. J. (2011). A quantitative assessment of patient barriers to insulin. *International journal of clinical practice*, 65(4), 408-414.
- A91 Darba, J., Restovic, G., Kaskens, L., Balbona, M. A., Carbonell, A., Cavero, P., ... and Padró, I. (2011). Patient preferences for osteoporosis in Spain: a discrete choice experiment. *Osteoporosis International*, 22(6), 1947-1954.

- A92 Faggioli, G., Scalone, L., Mantovani, L. G., Borghetti, F., Stella, A., and PREFER study group. (2011). Preferences of patients, their family caregivers and vascular surgeons in the choice of abdominal aortic aneurysms treatment options: the PREFER study. *European Journal of Vascular and Endovascular Surgery*, *42*(1), 26-34.
- A93 Golan, O., Hansen, P., Kaplan, G., and Tal, O. (2011). Health technology prioritization: which criteria for prioritizing new technologies and what are their relative weights?. *Health Policy*, *102*(2), 126-135.
- A94 Hong, S. H., Liu, J., Wang, J., Brown, L., and White-Means, S. (2010). Conjoint analysis of patient preferences on Medicare medication therapy management. *Journal of the American Pharmacists Association: JAPhA*, *51*(3), 378-387.
- A95 Johnson, F. R., Mohamed, A. F., Özdemir, S., Marshall, D. A., and Phillips, K. A. (2011). How does cost matter in health-care discrete-choice experiments?.*Health economics*, *20*(3), 323-330.
- A96 Marang-Van De Mheen, P. J., Dijs-Elsinga, J., Otten, W., Versluijs, M., Smeets, H. J., Vree, R., ... and Kievit, J. (2011). The relative importance of quality of care information when choosing a hospital for surgical treatment a hospital choice experiment. *Medical Decision Making*, *31*(6), 816-827.
- A97 Musters, A. M., de Bekker-Grob, E. W., Mochtar, M. H., van der Veen, F., and van Mello, N. M. (2011). Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF. *Human reproduction*, 26(9), 2425-2431.
- A98 Oteng, B., Marra, F., Lynd, L. D., Ogilvie, G., Patrick, D., and Marra, C. A. (2011). Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme. *Sexually transmitted infections*, 87(1), 52-57.
- A99 Robertson, R., and Burge, P. (2011). The impact of patient choice of provider on equity: analysis of a patient survey. *Journal of health services research and policy*, *16*(suppl 1), 22-28.
- A100 Sanders, P. M. H., Ijzerman, M. J., Roach, M. J., and Gustafson, K. J. (2011). Patient preferences for next generation neural prostheses to restore bladder function. *Spinal Cord*, *49*(1), 113-119.
- A101 Schaarschmidt, M. L., Schmieder, A., Umar, N., Terris, D., Goebeler, M., Goerdt, S., and Peitsch, W. K. (2011). Patient preferences for psoriasis treatments: process characteristics can outweigh outcome attributes. *Archives of dermatology*, *147*(11), 1285-1294.
- A102 Shafey, M., Lupichuk, S. M., Do, T., Owen, C., and Stewart, D. A. (2011). Preferences of patients and physicians concerning treatment options for relapsed follicular lymphoma: a discrete choice experiment. *Bone marrow transplantation*, *46*(7), 962-969.
- A103 Sweeting, M. K. R., Whitty, J. A., Scuffham, P. A., and Yelland, M. J. (2011). Patient Preferences for Treatment of Achilles Tendon Pain. *The Patient: Patient-Centered Outcomes Research*, 4(1), 45-54.
- A104 Twijnstra, A. R., Stiggelbout, A. M., de Kroon, C. D., and Jansen, F. W. (2011). Laparoscopic hysterectomy: eliciting preference of performers and colleagues via conjoint analysis. *Journal of minimally invasive gynecology*, *18*(5), 582-588.
- A105 Van Houtven, G., Johnson, F. R., Kilambi, V., and Hauber, A. B. (2011). Eliciting benefit–risk preferences and probability-weighted utility using choice-format conjoint analysis. *Medical Decision Making*, *31*(3), 469-480.
- A106 Vroomen, J. M., and Zweifel, P. (2011). Preferences for health insurance and health status: does it matter whether you are Dutch or German?. *The European Journal of Health Economics*, *12*(1), 87-95.

- A107 Wang, J., Hong, S. H., Meng, S., and Brown, L. M. (2011). Pharmacists' acceptable levels of compensation for MTM services: A conjoint analysis.*Research in Social and Administrative Pharmacy*, 7(4), 383-395.
- A108 Waschbusch, D. A., Cunningham, C. E., Pelham Jr, W. E., Rimas, H. L., Greiner, A. R., Gnagy, E. M., ... and Hoffman, M. T. (2011). A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naive children with ADHD. *Journal of Clinical Child and Adolescent Psychology*, *40*(4), 546-561.
- A109 Whitty, J. A., Scuffham, P. A., and Rundle-Thielee, S. R. (2011). Public and decision maker stated preferences for pharmaceutical subsidy decisions. *Applied health economics and health policy*, 9(2), 73-79.
- A110 King, M. T., Viney, R., Smith, D. P., Hossain, I., Street, D., Savage, E., ... and Armstrong, B. K. (2012). Survival gains needed to offset persistent adverse treatment effects in localised prostate cancer. *British journal of cancer*, 106(4), 638-645.
- Umar, N., Litaker, D., Schaarschmidt, M. L., Peitsch, W. K., Schmieder, A., and Terris, D. D. (2012). Outcomes associated with matching patients' treatment preferences to physicians' recommendations: study methodology. *BMC health services research*, 12(1), 1.
- B1 Longworth, L., Ratcliffe, J., and Boulton, M. (2001). Investigating women's preferences for intrapartum care: home versus hospital births. *Health and social care in the community*, *9*(6), 404-413.
- B2 Scott, A. (2001). Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics*, 20(3), 329–347.
- B3 Shackley, P., Slack, R., and Michaels, J. (2001). Vascular patients' preferences for local treatment: an application of conjoint analysis. *of Health Services Research and Policy*, 6(3), 151–157.
- B4 Aristides, M. (2002). WTP high for raltitrexed vs standard therapy in colorectal cancer. *PharmacoEconomics and Outcomes News*, 380, 14.
- B5 Gyrd-Hansen, D., and Slothuus, U. (2002). The citizen's preferences for financing public health care: a Danish survey. *International Journal of Health Care Finance and Economics*, 2(1), 25-36.
- B6 Hall, J., Kenny, P., King, M., Louviere, J., Viney, R., and Yeoh, A. (2002). Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. *Health Economics*, *11*(5), 457-465.
- B7 Phillips, K. A., Maddala, T., and Johnson, F. R. (2002). Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. *Health Services Research*, *37*(6), 1681-1705.
- B8 Ratcliffe, J., Van R., H., Buxton, M., Hardy, K., Colehan, J., and Partridge, M. (2002). Assessing patients' preferences for characteristics associated with homeopathic and conventional treatment of asthma: a conjoint analysis study. *Thorax*, *57*(6), 503-508.
- B9 Telser, H., and Zweifel, P. (2002). Measuring willingness-to-pay for risk reduction: an application of conjoint analysis. *Health Economics*, *11*(2), 129-139.
- B10 Ubach, C., Bate, A., Ryan, M., Porteous, T., Bond, C., and Robertson, R. (2002). Using discrete choice experiments to evaluate alternative electronic prescribing systems. *International Journal of Pharmacy Practice*, *10*(3), 191-200.
- B11 Bech, M. (2003). Politicians' and hospital managers' trade-offs in the choice of reimbursement scheme: a discrete choice experiment. *Health policy*, *66*(3), 261-275.
- B12 Salkeld, G., Solomon, M., Short, L., Ryan, M., and Ward, J. E. (2003). Evidence-based consumer choice: a case study in colorectal cancer screening. *Australian and New Zealand journal of public health*, 27(4), 449-455.

- B13 Taylor, S., and Armour, C. (2003). Consumer preference for dinoprostone vaginal gel using stated preference discrete choice modelling. *PharmacoEconomics*, *21*(10), 721-735.
- B14 Ubach, C., Scott, A., French, F., Awramenko, M., and Needham, G. (2003). What do hospital consultants value about their jobs? A discrete choice experiment. *British Medical Journal*, *326*(7404), 1432.
- B15 Aristides, M., Weston, A. R., FitzGerald, P., Reun, C. L., and Maniadakis, N. (2004).
 Patient Preference and Willingness-to-Pay for Humalog Mix25 Relative to Humulin 30/70: A Multicountry Application of a Discrete Choice Experiment. *Value in Health*, 7(4), 442-454.
- B16 Burge, P., Devlin, N., Appleby, J., Rohr, C., and Grant, J. (2004). Do patients always prefer quicker treatment?. *Applied Health Economics and Health Policy*, *3*(4), 183-194.
- B17 Dwight-Johnson, M., Lagomasino, I. T., Aisenberg, E., and Hay, J. (2004). Using conjoint analysis to assess depression treatment preferences among low-income Latinos. *Psychiatric Services*, *55*(8), 934-936.
- B18 Ratcliffe, J., Buxton, M., McGarry, T., Sheldon, R., and Chancellor, J. (2004). Patients' preferences for characteristics associated with treatments for osteoarthritis. *Rheumatology*, *43*(3), 337-345.
- B19 Sculpher, M., Bryan, S., Fry, P., de Winter, P., Payne, H., and Emberton, M. (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. *Bmj*, 328(7436), 382.
- B20 Wordsworth, S., Skåtun, D., Scott, A., and French, F. (2004). Preferences for general practice jobs: a survey of principals and sessional GPs. *British Journal of General Practice*, *54*(507), 740-746.
- B21 Gerard, K., and Lattimer, V. (2005). Preferences of patients for emergency services available during usual GP surgery hours: a discrete choice experiment. *Family Practice*, 22(1), 28-36.
- B22 Hanson, K., McPake, B., Nakamba, P., and Archard, L. (2005). Preferences for hospital quality in Zambia: results from a discrete choice experiment. *Health Economics*, *14*(7), 687-701.
- B23 Akkazieva, B., Gulacsi, L., Brandtmuller, A., Péntek, M., and Bridges, J. F. (2006). Patients' preferences for healthcare system reforms in Hungary: a conjoint analysis. *Applied Health Economics and Health Policy*, 5(3), 189-198.
- B24 Arana, J. E., Leon, C. J., and Quevedo, J. L. (2006). The effect of medical experience on the economic evaluation of health policies: A discrete choice experiment. *Social Science and Medicine*, *63*(2), 512-524.
- B25 Ashcroft, D. M., Seston, E., and Griffiths, C. E. (2006). Trade-offs between the benefits and risks of drug treatment for psoriasis : a discrete choice experiment with U.K dermatologists. *British Journal of Dermatology*, 155(6), 1236-1241.
- B26 Berchi, C., Dupuis, J. M., and Launoy, G. (2006). The reasons of general practitioners for promoting colorectal cancer mass screening in France. *European Journal of Health Economics*, 7(2), 91-98.
- B27 Mahadevia, P., Shah, S., Mannix, S., Brewster-Jordan, J., Kleinman, L., Liebman, C., and O'Downd, L. (2006). Willingness to pay for sensory attributes of intranasal corticosteroids among patients with allergic rhinitis. *Journal of Managed Care Pharmacy*, *12*(2), 143–151.
- B28 Peacock, S., Apicella, C., Andrews, L., Tucker, K., Bankier, A., Daly, M. B., and Hopper, J. L. (2006). A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. *British Journal of Cancer*, *95*(10), 1448-1453.

- B29 Porteous, T., Ryan, M., Bond, C. M., and Hannaford, P. (2006). Preferences for selfcare or professional advice for minor illness: a discrete choice experiment. *The British Journal of General Practice*, *56*(533), 911-917.
- B30 Baltussen, R., ten Asbroek, A. H., Koolman, X., Shrestha, N., and Bhattarai, P. (2007).
 Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health policy and planning*, 22(3), 178-185.
- B31 Burr, J. M., Kilonzo, M., Vale, L., and Ryan, M. (2007). Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. *Optometry and Vision Science*, 84(8), 797-808.
- B32 Caldow, J., Bond, C., Ryan, M., Campbell, N. C., Miguel, F. S., Kiger, A., and Lee, A. (2007). Treatment of minor illness in primary care: a national survey of patient satisfaction, attitudes and preferences regarding a wider nursing role. *Health Expectations*, 10(1), 30-45.
- B33 Gidman, W., Elliott, R. A., Payne, K., Meakin, G. H., and Moore, J. (2007). A comparison of parents and pediatric anesthesiologists' preferences for attributes of child daycase surgery: A discrete choice experiment. *Paediatric Anaesthesia*, 17(11), 1043-1052.
- B34 Hitchcock, W., Mellon, M., Memran, M., Parasuraman, B., Ramachandran, S., and Walzer, S. (2007). Caregiver preferences for pediatric asthma treatment delivery systems. *Advances in Therapy*, 24(6), 1240-1253.
- B35 Johnson, F. R., Ozdemir, S., Manjunath, R., Hauber, A. B., Burch, S. P., and Thompson, T. R. (2007). Factors that affect adherence to bipolar disorder treatments: a stated-preference approach. *Medical Care*, 45(6), 545-552.
- B36 Lancsar, E. J., Hall, J. P., King, M., Kenny, P., Louviere, J. J., Fiebig, D. G., Hossain, I., et al. (2007). Using discrete choice experiments to investigate subject preferences for preventive asthma medication. *Respirology*, *12*(1), 127-136.
- B37 Lloyd, A., McIntosh, E., Rabe, K. F., and Williams, A. (2007). Patient preferences for asthma therapy: a discrete choice experiment. *Primary Care Respiratory Journal*, *16*(4), 241-248.
- B38 Ossa, D. F., Briggs, A., McIntosh, E., Cowell, W., Littlewood, T., and Sculpher, M. (2007). Recombinant erythropoietin for chemotherapy-related anaemia: Economic value and health-related quality-of-life assessment using direct utility elicitation and discrete choice methods. *PharmacoEconomics*, 25(3), 223-237.
- B39 Papanikolaou, P., Lyne, P., and Ratcliffe, J. (2007). Using the discrete choice experimental design to investigate decision-making about pressure ulcer prevention by community nurses. *Health and Social Care in the Community*, *15*(6), 588-598.
- B40 Scott, A., Bond, C., Inch, J., and Grant, A. (2007). Preferences of community pharmacists for extended roles in primary care: a survey and discrete choice experiment. *PharmacoEconomics*, *25*(9), 783-792.
- B41 Seston, E. M., Elliott, R. A., Noyce, P. R., and Payne, K. (2007). Women's preferences for the provision of emergency hormonal contraception services. *Pharmacy World and Science*, *29*(3), 183–189.
- B42 Walzer, S. (2007). What do parents want from their child's asthma treatment? *Therapeutics and Clinical Risk Management*, *3*(1), 167-175
- B43 Araña, J. E., León, C. J., and Hanemann, M. W. (2008). Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly. *Journal of health economics*, 27(3), 753-769.
- B44 Danishevski, K., McKee, M., Sassi, F., and Maltcev, V. (2008). The decision to perform Caesarean section in Russia. *International Journal for Quality in Health Care*, 20(2), 88-94.

- B45 de Bekker-Grob, E. W., Essink-Bot, M., Meerding, W. J., Pols, H. A., Koes, B. W., and Steyerberg, E. W. (2008). Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. *Osteoporosis International*, *19*(7), 1029-1037
- B46 Grant, A. M., Wileman, S. M., Ramsay, C. R., Bojke, L., Sculpher, M. J., Kilonzo, M. M., ... and REFLUX Trial Group. (2008). The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease–a UK collaborative study. *Health Technology Assessment*, 12(31).
- B47 Grutters, J. P., Joore, M. A., Kessels, A. G., Davis, A. C., and Anteunis, L. J. (2008).
 Patient preferences for direct hearing aid provision by a private dispenser a discrete choice experiment. *Ear and Hearing*, 29(4), 557-564
- B48 Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of health economics*,27(4), 1078-1094.
- B49 Kjaer, T., and Gyrd-Hansen, D. (2008). Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. *Health Policy*, 85(1), 124-132.
- B50 Lloyd, A., Doyle, S., Dewilde, S., and Turk, F. (2008). Preferences and utilities for the symptoms of moderate to severe allergic asthma. *European Journal of Health Economics*, 9(3), 275-284.
- B51 Mangham, L. J., and Hanson, K. (2008). Employment preferences of public sector nurses in Malawi: results from a discrete choice experiment. *Tropical Medicine and International Health*, *13*(12), 1433-1441.
- B52 Negri'n, M. A., Pinilla, J., and Leo'n, C. (2008). Willingness to pay for alternative policies for patients with Alzheimer's Disease. *Health Economics, Policy and Law, 3*(Pt 3), 257-275.
- B53 Tinetti, M. E., McAvay, G. J., Fried, T. R., Foody, J. M., Bianco, L., Ginter, S., and Fraenkel, L. (2008). Development of a Tool for Eliciting Patient Priority from Among Competing Cardiovascular Disease, Medication-Symptoms, and Fall Injury Outcomes. *Journal of the American Geriatrics Society*, *56*(4), 730-736.
- B54 Bijlenga, D., Birnie, E., and Bonsel, G. J. (2009). Feasibility, Reliability, and Validity of Three Health-State Valuation Methods Using Multiple-Outcome Vignettes on Moderate-Risk Pregnancy at Term. *Value in Health*, *12*(5), 821-827.
- B55 Brown, D. S., Finkelstein, E. A., Brown, D. R., Buchner, D. M., and Johnson, F. R. (2009). Estimating older adults' preferences for walking programs via conjoint analysis. *American journal of preventive medicine*, *36*(3), 201-207.
- B56 Clark, M., Moro, D., and Szczepura, A. (2009). Balancing patient preferences and clinical needs: Community versus hospital based care for patients with suspected DVT. *Health Policy*, *90*(2-3), 313-319.
- B57 Cunningham, C. E., Deal, K., Rimas, H., Chen, Y., Buchanan, D. H., and Sdao-Jarvie, K. (2009). Providing information to parents of children with mental health problems: a discrete choice conjoint analysis of professional preferences. *Journal of abnormal child psychology*, 37(8), 1089-1102.
- B58 Green, C., and Gerard, K. (2009). Exploring the social value of health-care interventions: a stated preference discrete choice experiment. *Health Economics*, *18*(8), 951-976.
- B59 Griffith, J. M., Lewis, C. L., Hawley, S., Sheridan, S. L., and Pignone, M. P. (2009).
 Randomized trial of presenting absolute v. relative risk reduction in the elicitation of patient values for heart disease prevention with conjoint analysis.*Medical Decision Making*.

- B60 Guimaraes, C., Marra, C. A., Colley, L., Gill, S., Simpson, S. H., Meneilly, G. S., ... and Lynd, L. D. (2009). A valuation of patients' willingness-to-pay for insulin delivery in diabetes. *International journal of technology assessment in health care*, 25(03), 359-366.
- B61 Hauber, A. B., Mohamed, A. F., Johnson, F. R., and Falvey, H. (2009). Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents. *Diabetic Medicine*, 26(4), 416-424.
- B62 Hyde, S., Gansky, S. A., Gonzalez-Vargas, M. J., Husting, S. R., Cheng, N. F., Millstein, S. G., and Adams, S. H. (2009). Developing an acceptability assessment of preventive dental treatments. *Journal of public health dentistry*, 69(1), 18-23.
- B63 Kruijshaar, M. E., Essink-Bot, M. L., Donkers, B., Looman, C. W., Siersema, P. D., and Steyerberg, E. W. (2009). A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. *BMC medical research methodology*, 9(1), 31.
- B64 Kruk, M. E., Paczkowski, M., Mbaruku, G., de Pinho, H., and Galea, S. (2009). Women's preferences for place of delivery in rural Tanzania: a population-based discrete choice experiment. *American journal of public health*, *99*(9), 1666.
- B65 Lim, J. N., and Edlin, R. (2009). Preferences of older patients and choice of treatment location in the UK: a binary choice experiment. *Health policy*, *91*(3), 252-257.
- B66 Marshall, D. A., Johnson, F. R., Kulin, N. A., Özdemir, S., Walsh, J. M., Marshall, J. K.,
 ... and Phillips, K. A. (2009). How do physician assessments of patient preferences for
 colorectal cancer screening tests differ from actual preferences? A comparison in Canada and
 the United States using a stated-choice survey. *Health economics*, 18(12), 1420-1439.
- B67 Mühlbacher, A. C., Rudolph, I., Lincke, H. J., and Nübling, M. (2009). Preferences for treatment of attention-deficit/hyperactivity disorder (ADHD): a discrete choice experiment. *BMC health services research*, *9*(1), 149.
- B68 Pavlova, M., Hendrix, M., Nouwens, E., Nijhuis, J., and van Merode, G. (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.
- B69 Ratcliffe, J., Bekker, H. L., Dolan, P., and Edlin, R. (2009). Examining the attitudes and preferences of health care decision-makers in relation to access, equity and cost-effectiveness: a discrete choice experiment. *Health Policy*, *90*(1), 45-57.
- B70 Regier, D. A., Friedman, J. M., Makela, N., Ryan, M., and Marra, C. A. (2009). Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children. *Clinical genetics*, 75(6), 514-521.
- B71 Scalone, L., Mantovani, L. G., Borghetti, F., Von Mackensen, S., and Gringeri, a. (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
- B72 Blaauw, D., Erasmus, E., Pagaiya, N., Tangcharoensathein, V., Mullei, K., Mudhune, S.,
 ... and Lagarde, M. (2010). Policy interventions that attract nurses to rural areas: a
 multicountry discrete choice experiment. *Bulletin of the World Health Organization*, 88(5), 350-356.
- B73 Bunge, E. M., de Bekker-Grob, E. W., van Biezen, F. C., Essink-Bot, M. L., and de Koning, H. J. (2010). Patients' preferences for scoliosis brace treatment: a discrete choice experiment. *Spine*, *35*(1), 57-63.
- B74 de Bekker-Grob, E. W., Hofman, R., Donkers, B., van Ballegooijen, M., Helmerhorst, T. J., Raat, H., and Korfage, I. J. (2010). Girls' preferences for HPV vaccination: a discrete choice experiment. *Vaccine*, 28(41), 6692-6697.

- B75 Bekker-Grob, D., Esther, W., Hol, L., Donkers, B., Van Dam, L., Habbema, J. D. F., ... and Steyerberg, E. W. (2010). Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value in Health*, *13*(2), 315-323.
- B76 Essers, B. A., Dirksen, C. D., Prins, M. H., and Neumann, H. A. (2010). Assessing the Public's Preference for Surgical Treatment of Primary Basal Cell Carcinoma: A Discrete-Choice Experiment in the South of The Netherlands. *Dermatologic Surgery*, *36*(12), 1950-1955.
- B77 Hauber, A. B., Johnson, F. R., Grotzinger, K. M., and Özdemir, S. (2010). Patients' benefit-risk preferences for chronic idiopathic thrombocytopenic purpura therapies. *Annals of Pharmacotherapy*, *44*(3), 479-488.
- B78 Hauber, A. B., Mohamed, A. F., Johnson, F. R., Oyelowo, O., Curtis, B. H., and Coon, C. (2010). Estimating importance weights for the IWQOL-Lite using conjoint analysis. *Quality of Life Research*, 19(5), 701-709.
- B79 Hol, L., de Bekker-Grob, E. W., van Dam, L., Donkers, B., Kuipers, E. J., Habbema, J.
 D. F., ... and Essink-Bot, M. L. (2010). Preferences for colorectal cancer screening strategies: a discrete choice experiment. *British journal of cancer*, *102*(6), 972-980.
- B80 Johnson, F. R., Hauber, A. B., Özdemir, S., and Lynd, L. (2010). Quantifying Women's Stated Benefit–Risk Trade-Off Preferences for IBS Treatment Outcomes. *Value in Health*, 13(4), 418-423.
- B81 Johnson, F. R., Hauber, B., Ozdemir, S., Siegel, C. A., Hass, S., and Sands, B. E. (2010).
 Are gastroenterologists less tolerant of treatment risks than patients? Benefit-risk preferences in Crohn's disease management. *J Managed Care Pharm*, 16(8), 616-28.
- B82 Koopmanschap, M. A., Stolk, E. A., and Koolman, X. (2010). Dear policy maker: have you made up your mind? A discrete choice experiment among policy makers and other health professionals. *International journal of technology assessment in health care*, 26(02), 198-204.
- B83 Kruk, M. E., Johnson, J. C., Gyakobo, M., Agyei-Baffour, P., Asabir, K., Kotha, S. R., ... and Dzodzomenyo, M. (2010). Rural practice preferences among medical students in Ghana: a discrete choice experiment. *Bulletin of the World Health Organization*, 88(5), 333-341.
- B84 Nayaradou, M., Berchi, C., Dejardin, O., and Launoy, G. (2010). Eliciting population preferences for mass colorectal cancer screening organization.*Medical Decision Making*, *30*(2), 224-233.
- B85 Opuni, M., Bishai, D., Gray, G. E., McIntyre, J. A., and Martinson, N. A. (2010). Preferences for characteristics of antiretroviral therapy provision in Johannesburg, South Africa: results of a conjoint analysis. *AIDS and Behavior*, *14*(4), 807-815.
- B86 Torbica, A., and Fattore, G. (2010). Understanding the impact of economic evidence on clinical decision making: a discrete choice experiment in cardiology. *Social science and medicine*, 70(10), 1536-1543.
- B87 Youngkong, S., Baltussen, R., Tantivess, S., Koolman, X., and Teerawattananon, Y.
 (2010). Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC health services research*, 10(1), 197.
- B88 Bonnichsen, M. O. (2011). Elicitation of ostomy pouch preferences. *The Patient: Patient-Centered Outcomes Research*, 4(3), 163-175.
- B89 Bridges, J. F., Selck, F. W., Gray, G. E., McIntyre, J. A., and Martinson, N. A. (2011).
 Condom avoidance and determinants of demand for male circumcision in Johannesburg, South Africa. *Health policy and planning*, 26(4), 298-306.

- B90 Damen, T. H., de Bekker-Grob, E. W., Mureau, M. A., Menke-Pluijmers, M. B., Seynaeve, C., Hofer, S. O., and Essink-Bot, M. L. (2011). Patients' preferences for breast reconstruction: a discrete choice experiment. *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 64(1), 75-83.
- B91 Diaby, V., Kakou, H. D., and Lachaine, J. (2011). Eliciting preferences for reimbursed drugs selection criteria in Côte d'Ivoire. *The Patient: Patient-Centered Outcomes Research*, 4(2), 125-131.
- B92 Fegert, J. M., Slawik, L., Wermelskirchen, D., Nübling, M., and Mühlbacher, A. (2011). Assessment of parents' preferences for the treatment of school-age children with ADHD: a discrete choice experiment.
- B93 Guo, N., Marra, C. A., FitzGerald, J. M., Elwood, R. K., Anis, A. H., and Marra, F. (2011). Patient preference for latent tuberculosis infection preventive treatment: a discrete choice experiment. *Value in Health*, 14(6), 937-943.
- B94 Hauber, A. B., Gonzalez, J. M., Schenkel, B., Lofland, J. H., and Martin, S. (2011). The value to patients of reducing lesion severity in plaque psoriasis. *Journal of Dermatological Treatment*, 22(5), 266-275.
- B95 Howard, K., Salkeld, G., Pignone, M., Hewett, P., Cheung, P., Olsen, J., ... and Roberts-Thomson, I. C. (2011). Preferences for CT colonography and colonoscopy as diagnostic tests for colorectal cancer: a discrete choice experiment. *Value in Health*, *14*(8), 1146-1152.
- B96 Kolstad, J. R. (2011). How to make rural jobs more attractive to health workers. Findings from a discrete choice experiment in Tanzania. *Health economics*,20(2), 196-211.
- B97 Kruk, M. E., Rockers, P. C., Tornorlah Varpilah, S., and Macauley, R. (2011). Population preferences for health care in Liberia: insights for rebuilding a health system. *Health services research*, *46*(6pt2), 2057-2078.
- B98 Laver, K., Rehab, M., Ratcliffe, J., George, S., Lester, L., Walker, R., ... and Crotty, M. (2011). Early rehabilitation management after stroke: what do stroke patients prefer?. *Journal of Rehabilitation Medicine*, *43*(4), 354-358.
- B99 Lloyd, A., Hodgkins, P., Dewilde, S., Sasané, R., Falconer, S., and Sonuga Barke, E. (2011). Methylphenidate delivery mechanisms for the treatment of children with attention deficit hyperactivity disorder: Heterogeneity in parent preferences. *International journal of technology assessment in health care*, 27(03), 215-223.
- B100 Lloyd, A., Nafees, B., Barnett, A. H., Heller, S., Ploug, U. J., Lammert, M., and Bøgelund, M. (2011). Willingness to pay for improvements in chronic long-acting insulin therapy in individuals with type 1 or type 2 diabetes mellitus. *Clinical therapeutics*, *33*(9), 1258-1267.
- B101 Mentzakis, E., Ryan, M., and McNamee, P. (2011). Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. *Health economics*, 20(8), 930-944.
- B102 Mentzakis, E., Stefanowska, P., and Hurley, J. (2011). A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. *Health Economics, Policy and Law, 6*(03), 405-433.
- B103 Mühlbacher, A. C., and Nübling, M. (2011). Analysis of physicians' perspectives versus patients' preferences: direct assessment and discrete choice experiments in the therapy of multiple myeloma. *The European Journal of Health Economics*, *12*(3), 193-203.
- B104 Nathan, H., Bridges, J. F., Schulick, R. D., Cameron, A. M., Hirose, K., Edil, B. H., ... and Pawlik, T. M. (2011). Understanding surgical decision making in early hepatocellular carcinoma. *Journal of Clinical Oncology*, *29*(6), 619-625.

- B105 Pereira, C. C., Mulligan, M., Bridges, J. F., and Bishai, D. (2011). Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis. *Vaccine*,29(7), 1443-1447.
- B106 Schwappach, D. L., Mülders, V., Simic, D., Wilm, S., and Thürmann, P. A. (2011). Is less more? Patients' preferences for drug information leaflets.*Pharmacoepidemiology and drug safety*, *20*(9), 987-995.
- B107 Swinburn, P., Lloyd, A., Ali, S., Hashmi, N., Newal, D., and Najib, H. (2011).
 Preferences for antimuscarinic therapy for overactive bladder. *BJU international*, *108*(6), 868-873.
- B108 Yil, D., Ryanl, M., Campbelll, S., Elliottl, A., Torrancel, N., Chambersl, A., ... and Smithl, B. H. (2011). Using discrete choice experiments to inform randomised controlled trials: an application to chronic low back pain management in primary care. *European Journal of Pain*, 15(5), 531-e1.
- B109 Benjamin, L., Cotté, F. E., Philippe, C., Mercier, F., Bachelot, T., and Vidal-Trécan, G. (2012). Physicians' preferences for prescribing oral and intravenous anticancer drugs: a discrete choice experiment. *European Journal of Cancer*, 48(6), 912-920.
- B110 Chancellor, J., Martin, M., Liedgens, H., Baker, M. G., and Müller-Schwefe, G. H. (2012). Stated preferences of physicians and chronic pain sufferers in the use of classic strong opioids. *Value in Health*, 15(1), 106-117.
- B111 Eisingerich, A. B., Wheelock, A., Gomez, G. B., Garnett, G. P., Dybul, M. R., and Piot, P. K. (2012). Attitudes and acceptance of oral and parenteral HIV preexposure prophylaxis among potential user groups: a multinational study. *PLoS One*, 7(1), e28238.
- B112 Hodgkins, P., Swinburn, P., Solomon, D., Yen, L., Dewilde, S., and Lloyd, A. (2012).
 Patient Preferences for First-Line Oral Treatment for Mild-to-Moderate Ulcerative Colitis. *The Patient-Patient-Centered Outcomes Research*, 5(1), 33-44.
- B113 Lim, M. K., Bae, E. Y., Choi, S. E., Lee, E. K., and Lee, T. J. (2012). Eliciting public preference for health-care resource allocation in South Korea. *Value in Health*, 15(1), S91-S94.
- B114 Mohamed, A. F., Johnson, F. R., Hauber, A. B., Lescrauwaet, B., and Masterson, A. (2012). Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey, and Italy. *European journal of gastroenterology and hepatology*, 24(4), 419-426.
- C1 Mark, T. L., and Swait, J. (2003). Using stated preference modeling to forecast the effect of medication attributes on prescriptions of alcoholism medications. *Value in Health*, 6(4), 474-482.
- C2 Roux, L., Ubach, C., Donaldson, C., and Ryan, M. (2004). Valuing the benefits of weight loss programs: an application of the discrete choice experiment. *Obesity Research*, *12*(8), 1342-1351.
- C3 Hall, J., Fiebig, D. G., King, M. T., Hossain, I., and Louviere, J. J. (2006). What influences participation in genetic carrier testing? Results from a discrete choice experiment. *Journal of Health Economics*, 25(3), 520-37.
- C4 Ryan, M., Netten, A., Ska[°]tun, D., and Smith, P. (2006). Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people. *Journal of Health Economics*, 25(5), 927-944.
- C5 Fitzpatrick, E., Coyle, D. E., Durieux-Smith, A., Graham, I. D., Angus, D. E., and Gaboury, I. (2007). Parents' preferences for services for children with hearing loss: a conjoint analysis study. *Ear and hearing*, 28(6), 842-849.

- C6 Haughney, J., Fletcher, M., Wolfe, S., Ratcliffe, J., Brice, R., and Partridge, M. R. (2007). Features of asthma management: quantifying the patient perspective. *BMC pulmonary medicine*, *7*(1), 16.
- C7 Tappenden, P., Brazier, J., Ratcliffe, J., and Chilcott, J. (2007). A stated preference binary choice experiment to explore NICE decision making. *PharmacoEconomics*, 25(8), 685-693.
- C8 Turner, D., Tarrant, C., Windridge, K., Bryan, S., Boulton, M., Freeman, G., and Baker, R. (2007). Do patients value continuity of care in general practice? An investigation using stated preference discrete choice experiments. *Journal of Health Services Research and Policy*, 12(3), 132-137.
- C9 Cheraghi-Sohi, S., Hole, A. R., Mead, N., McDonald, R., Whalley, D., Bower, P., and Roland, M. (2008). What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. *The Annals of Family Medicine*, 6(2), 107-115.
- C10 Gerard, K., Salisbury, C., Street, D., Pope, C., and Baxter, H. (2008). Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. *Journal of Health Services Research and Policy*, *13*(S2), 3-10.
- C11 Lloyd, A., Penson, D., Dewilde, S., and Kleinman, L. (2008). Eliciting patient preferences for hormonal therapy options in the treatment of metastatic prostate cancer. *Prostate Cancer and Prostatic Diseases*, *11*(2), 153-159.
- C12 Pitchforth, E., Watson, V., Tucker, J., Ryan, M., van Teijlingen, E., Farmer, J., Ireland, J., et al. (2008). Models of intrapartum care and women's trade-offs in remote and rural Scotland: a mixed methods study. *BJOG*, *115*(5), 560-569.
- C13 Herbild, L., Bech, M., and Gyrd-Hansen, D. (2009). Estimating the Danish populations' preferences for pharmacogenetic testing using a discrete choice experiment. The case of treating depression. *Value in Health*, *12*(4), 560-567.
- C14 Richardson, G., Bojke, C., Kennedy, A., Reeves, D., Bower, P., Lee, V., Middleton, E., et al. (2009). What outcomes are important to patients with long term conditions? A discrete choice experiment. *Value in Health*, *12*(2), 331-339.
- C15 Ryan, M., Watson, V., and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health economics*, *18*(3), 321-336.
- C16 Baker, R., Bateman, I., Donaldson, C., Jones-Lee, M., Lancsar, E., Loomes, G., ... and Wildman, J. (2010). Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health technology assessment 14*(27), 1-178.
- C17 Grindrod, K. A., Marra, C. A., Colley, L., Tsuyuki, R. T., and Lynd, L. D. (2010). Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy. *Annals of Pharmacotherapy*, *44*(10), 1554-1564.
- C18 Bridges, J. F., Gallego, G., Kudo, M., Okita, K., Han, K. H., Ye, S. L., and Blauvelt, B. M. (2011). Identifying and prioritizing strategies for comprehensive liver cancer control in Asia. *BMC health services research*, *11*(1), 298.
- C19 Bridges, J., Kinter, E. T., Schmeding, A., Rudolph, I., and Muhlbacher, A. (2011). Can patients diagnosed with schizophrenia complete choice-based conjoint analysis tasks? *Patient*, 4(4), 267–275.
- C20 Hsieh, Y. H., Gaydos, C. A., Hogan, M. T., Uy, O. M., Jackman, J., Jett-Goheen, M., ... and Rompalo, A. M. (2011). What qualities are most important to making a point of care test desirable for clinicians and others offering sexually transmitted infection testing?. *PLoS One*, *6*(4), e19263.

- C21 Lagarde, M., Paintain, L. S., Antwi, G., Jones, C., Greenwood, B., Chandramohan, D., ... and Webster, J. (2011). Evaluating health workers' potential resistance to new interventions: a role for discrete choice experiments. *PloS one*, *6*(8), e23588.
- C22 Payne, K., Fargher, E. A., Roberts, S. A., Tricker, K., Elliott, R. A., Ratcliffe, J., and Newman, W. G. (2011). Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. *Value in Health*, *14*(1), 121-134.
- C23 Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., and Brazier, J. E. (2011). Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Social science and medicine*, 72(10), 1717-1727.
- C24 Poulos, C., Yang, J. C., Levin, C., Van Minh, H., Giang, K. B., and Nguyen, D. (2011). Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam. *Social Science and Medicine*, *73*(2), 226-234.
- C25 Scotland, G. S., McNamee, P., Cheyne, H., Hundley, V., and Barnett, C. (2011). Women's preferences for aspects of labor management: results from a discrete choice experiment. *Birth*, *38*(1), 36-46.
- Van Empel, I. W., Dancet, E. A., Koolman, X. H., Nelen, W. L., Stolk, E. A., Sermeus, W., ... and Kremer, J. A. (2011). Physicians underestimate the importance of patient-centredness to patients: a discrete choice experiment in fertility care. *Human reproduction*, 26(3), 584-593.
- C27 Vujicic, M., Shengelia, B., Alfano, M., and Thu, H. B. (2011). Physician shortages in rural Vietnam: Using a labor market approach to inform policy. *Social Science and Medicine*, *73*(7), 970-977.
- C28 Morton, R. L., Snelling, P., Webster, A. C., Rose, J., Masterson, R., Johnson, D. W., and Howard, K. (2012). Factors influencing patient choice of dialysis versus conservative care to treat end-stage kidney disease. *Canadian Medical Association Journal*, *184*(5), 277-283.
- C29 Naik-Panvelkar, P., Armour, C., Rose, J., and Saini, B. (2012). Patients' value of asthma services in Australian pharmacies: the way ahead for asthma care. *Journal of Asthma*, 49(3), 310-316.

Appendix 5.7: Tabulated results from the systematic review of DCEs

Year	n	(%)	
2001	7	(3%)	
2002	8	(3%)	
2003	7	(3%)	
2004	12	(4%)	
2005	12	(4%)	
2006	20	(7%)	
2007	27	(10%)	
2008	25	(9%)	
2009	41	(15%)	
2010	42	(16%)	
2011	58	(21%)	
2012*	11	(4%)	

 Table A1: Number and percentage of studies included in the systematic review by year of publication

*2012 up to search date

Table A2:Number and percentage of studies included in the systematic review by country

Country	n	%	Country	n	%
Australia	23	8.5%	Multi Country	20	7.4%
Canada	16	5.9%	Nepal	1	0.4%
Denmark	9	3.3%	Netherlands	33	12.2%
Finland	1	0.4%	Russia	1	0.4%
France	3	1.1%	South Africa	2	0.7%
Germany	7	2.6%	South Korea	1	0.4%
Ghana	3	1.1%	Spain	4	1.5%
Hong Kong	2	0.7%	Sweden	2	0.7%
Hungary	1	0.4%	Switzerland	1	0.4%
Ireland	1	0.4%	Taiwan	1	0.4%
Italy	5	1.8%	Tanzania	2	0.7%
Ivory Coast	1	0.4%	Thailand	1	0.4%
Japan	4	1.5%	UK	76	28.0%
Liberia	1	0.4%	USA	43	15.9%
Malawi	1	0.4%	Vietnam	2	0.7%
Missing	2	0.7%	Zambia	1	0.4%

Health Services Research1-51% 39 Health Economics7 57% 29% 9 99 43% 22% Health services research and health policy6 43% 22% journals00 38% 22% Specialist medicine1- 38% 02 2% 0.7% Cardiology1 1% 0.4% Dentistry1 1% 0.4% Dermatology5 5% 1.8% Diet2 2% 0.7% Endocrinology2 2% 0.7% Gastroenterology2 2% 0.7% Genetics1 1% 0.4% Hematology4 4% 1.4%		n	% of sub discipline	% of total studies
39 Health Economics 7 9 Health services research and health policy journals $57%$ 9 $43%$ $22%$ $22%$ Specialist medicine 1 02 $-$ $38%$ Mark O2 02 $02%$ $0.7%$ Anesthesiology 2 $2%$ $2%$ $0.7%$ Dentistry 1 $1%$ $1%$ $0.4%$ Dermatology 5 $5%$ $5%$ $1.8%$ $0.1%$ Diet 2 $2%$ $2%$ $0.7%$ Gastroenterology 2 $2%$ $2%$ $0.7%$ $0.7%$ Genetics 1 $1%$ $1%$ $0.4%$ Hematology 4 $4%$ $4%$ $1.4%$	Health Services Research	1	-	51%
Health Services research and health policy 7 37.6 27.96 Health services research and health policy 6 43% 22% Specialist medicine 1 $ 38\%$ O2 02 02 02% Anesthesiology 2 2% 0.7% Cardiology 1 1% 0.4% Dentistry 1 1% 0.4% Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Health Economics	39 7	57%	20%
Health services research and health policy journals6 43% 22% 0 Specialist medicine1- 38% 020202Anesthesiology22%Cardiology11%Dentistry11%Dermatology55%Diet22%Cardiology22%0.1%0.4%Dermatology55%0.18%0.16%0.16%2%0.17%0.17%0.18%0.16%0.18%0.16%0.18%0.4%0.11%0.1%0.11%0.1%0.11%	Teatur Economies	9	5770	2970
journals 0 Specialist medicine 1 - 38% 02 02 0.7% Anesthesiology 2 2% 0.7% Cardiology 1 1% 0.4% Dentistry 1 1% 0.4% Dentistry 1 1% 0.4% Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Health services research and health policy	6	43%	22%
Operation Operation <t< td=""><td>Journals</td><td><u> </u></td><td></td><td>38%</td></t<>	Journals	<u> </u>		38%
Anesthesiology 2 2% 0.7% Cardiology 1 1% 0.4% Dentistry 1 1% 0.4% Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Specialist medicine	02	-	5070
Cardiology 1 1% 0.4% Dentistry 1 1% 0.4% Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Anesthesiology	2	2%	0.7%
Dentistry 1 1% 0.4% Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Cardiology	1	1%	0.4%
Dermatology 5 5% 1.8% Diet 2 2% 0.7% Endocrinology 2 2% 0.7% Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Hematology 4 4% 1.4%	Dentistry	1	1%	0.4%
Diet22%0.7%Endocrinology22%0.7%Gastroenterology22%0.7%Genetics11%0.4%Geriatrics11%0.4%Hematology44%1.4%	Dermatology	5	5%	1.8%
Endocrinology22%0.7%Gastroenterology22%0.7%Genetics11%0.4%Geriatrics11%0.4%Hematology44%1.4%	Diet	2	2%	0.7%
Gastroenterology 2 2% 0.7% Genetics 1 1% 0.4% Geriatrics 1 1% 0.4% Hematology 4 4% 1.4%	Endocrinology	2	2%	0.7%
Genetics 1 1% 0.4% Geriatrics 1 1% 0.4% Hematology 4 4% 1.4%	Gastroenterology	2	2%	0.7%
Geriatrics 1 1% 0.4% Hematology 4 4% 1.4%	Genetics	1	1%	0.4%
Hematology 4 4% 1.4%	Geriatrics	1	1%	0.4%
	Hematology	4	4%	1.4%
Immunology 1 1% 0.4%	Immunology	1	1%	0.4%
Nephrology 2 2% 0.7%	Nephrology	2	2%	0.7%
Neurology 2 2% 0.7%	Neurology	2	2%	0.7%
Obstetrics and gynecology113%4.7%	Obstetrics and gynecology	1	13%	4.7%
3	Orresterry	3	00/	2.20/
Ontcology 9 9% 5.2%	Oncology	9	9%	3.2%
Optimizing 2 2% 0.7%	Opthalmology	2	2%	0.7%
Orthopedics 3 3% 1.1%	Orthopedics	3	3%	1.1%
Otolaryngology 2 2% 0.7%	Otolaryngology	2	2%	0.7%
Physical medicine and renabilitation 1 1% 0.4%	Physical medicine and renabilitation	1	1%	0.4%
Psychiatry 3 3% 1.1%	Psychiatry	3	3%	1.1%
Psychology 2 2% 0.7%	Psychology	2	2%	0.7%
Reproductive medicine 4 4% 1.4%	Reproductive medicine	4	4%	1.4%
Respiratory medicine 1 13% 4.7%	Respiratory medicine	1	13%	4.7%
Rheumatology 3 3% 1.1%	Rheumatology	3	3%	1.1%
Surgery 2 2% 0.7%	Surgery	2	2%	0.7%
Urology 1 1% 0.4%	Urology	1	1%	0.4%
Venerology 1 1% 0.4%	Venerology	1	1%	0.4%
Telemedicine 3 3% 1.1%	Telemedicine	3	3%	1.1%
Emergency Care33%1.1%	Emergency Care	3	3%	1.1%
Nursing 1 1% 0.4%	Nursing	1	1%	0.4%
Pharmacology 1 10% 3.6%	Pharmacology	1	10%	3.6%
0		0		
General Medicine 3 11%	General Medicine	3 0		11%

Table A3: Number and percentage of studies included in the systematic review by journal type

Appendix 5.8: List of studies not included in the Clark et al. (2014) systematic review

- BNF (2015). Adverse reactions to drugs. British National Formulary. 62.
- Han, P., Klein, W., Killam, B., Lehman, T., Massett, H. and Freedman, A.N. (2012). Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. *Patient education and counseling*. 86 (1). 106–113.
- Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J. and Payne, K. (2014). Risk as an attribute in discrete choice experiments: a critical review. *The patient*. 7 (2). 151–70.
- Hutton, J.L. (2009). Number needed to treat and number needed to harm are not the best way to report and assess the results of randomised clinical trials. *British journal of haematology*. 146 (1). 27–30.
- Keller, C., Siegrist, M. and Visschers, V. (2009). Effect of risk ladder format on risk perception in highand low-numerate individuals. *Risk Analysis*. 29 (9). 1255–64.
- Lipkus, I. and Hollands, J. (1999). The visual communication of risk. JNCI monographs. 25. 149–163.
- McCaffery, K., Dixon, A., Hayen, A., Jansen, J., Smith, S. and Simpson, J.M. (2012). The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study. *Medical Decision Making*. 32 (4). 532–44.
- Moore, R.A., Derry, S., McQuay, H.J. and Paling, J. (2008). What do we know about communicating risk? A brief review and suggestion for contextualising serious, but rare, risk, and the example of cox-2 selective and non-selective NSAIDs. *Arthritis research & therapy*. 10 (1). 1–16.
- Paling, J. (2003). Strategies to help patients understand risks. BMJ British Medical Journal. 745-748.
- Schapira, M.M., Nattinger, A.B. and McHorney, C.A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*. 21 (6). 459– 467.
- Schmidt, M. (2004). Investigating risk perception: a short introduction. In: *Loss of agro-biodiversity in Vavilov centers, with a special focus on the risks of genetically modified organisms (GMOs).* Vienna: Schmidit PhD Thesis, 1–16.
- Schwartz, L., Woloshin, S., Black, W. and Welch, H.G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*. 127 (11). 1023–1028.
- Sprague, D., Russo, J., Lavallie, D.L. and Buchwald, D.S. (2012). Influence of framing and graphic format on comprehension of risk information among American Indian tribal college students. *Journal of Cancer Education*. 27. 752–8.

Appendix 5.9: The submitted paper describing the development of the appraisal tool

Health Economics



CHEQUAL: a checklist for the use of qualitative methods in stated preference valuation studies

Journal:	Health Economics
Manuscript ID	HEC-15-0445
Wiley - Manuscript type:	Health Economics Letter
Keywords:	Qualitative research, stated preference, discrete choice experiment, contingent valuation method, checklist

SCHOLARONE^{**} Manuscripts

http://mc.manuscriptcentral.com/hec

Title: CHEQUAL: a checklist for the use of qualitative methods in stated preference valuation studies

Key words: Qualitative research, stated preference, discrete choice experiment, contingent valuation method, critical appraisal, checklist

http://mc.manuscriptcentral.com/hec

59 60

Page 2 of 22

Summary [200 words]
Qualitative research methods can inform the design and interpretation of stated preference
valuation studies. The potential strengths of qualitative research can only be realised if
studies are conducted appropriately and reported with clarity. This study had three aims: to
identify whether existing qualitative research critical appraisal tools are suitable for use
regarding qualitative research conducted as part of a stated preference valuation study; make
recommendations on the modification of existing tools to produce a new checklist; and test
the proposed checklist in an exemplar systematic review of stated preference studies. An
iterative process of identifying, testing and critiquing existing appraisal tools was used to
produce a checklist for use when reporting qualitative methods used alongside a stated
preference valuation study. Seven established qualitative appraisal tools were identified in the
health services research literature. However, these tools were judged to be inappropriate for
use in respect to stated preference valuation studies. A checklist was designed for health
economists using qualitative methods in stated preference valuation studies ('CHEQUAL').
The new checklist was tested in a systematic review of healthcare discrete choice
experiments. CHEQUAL offers a transparent mechanism for extracting, summarising and
critiquing the reported qualitative research used in stated preference valuation studies.

http://mc.manuscriptcentral.com/hec

458

2			
3			
4			
6			
7			
8			
9			
10			
11			
12			
14			
15			
16			
17			
18			
19			
20			
22			
23			
24			
25			
26			
27			
29			
30			
31			
32			
33			
34			
30			
37			
38			
39			
40			
41			
42			
44			
45			
46			
47			
48			
50			
51			
52			
53			
54			
55			
57			
58			
59			
60			

1. Introduction

The need to promote rigour in published research has stimulated the development of checklists for the appraisal of study designs used in health economics (for example, the Consolidated Health Economic Evaluation Reporting Standards (Husereau et al., 2013; Xie et al., 2015). These checklists form the basis for appraisal tools that can be used to inform the design of studies and preparation of manuscripts for submission to journals or the critical appraisal of published studies in the context of systematic reviews. Similarly, peer-reviewers, funders and decision-makers, interested in assessing or using the research, can use these checklists to provide a structured view on the quality of published studies as reported.

⁴Qualitative research' is increasingly advocated in the field of health economics (Coast, 1999; Coast et al., 2004). The term 'qualitative research' refers to a broad range of philosophies, approaches and methodologies used to acquire an in-depth understanding or explanation of people's perceptions (Braun & Clarke, 2006; Corbin & Strauss, 2008; Hsieh & Shannon, 2005; Pope & Mays, 2008). Some commentators have made specific recommendations for application of qualitative research in the design and interpretation of quantitative stated preference (SP) studies (see figure 1 for an overview of types of SP studies) (Coast et al., 2012; Kløjgaard et al., 2012; Baker & Robinson, 2004; Baker et al., 2008). A key strength of qualitative methods is being able to collate important contextual, alongside quantitative preference, data (Cheraghi-Sohi et al., 2007; Ryan et al., 2009). The potential strengths of qualitative research can only be realised if studies are conducted appropriately and reported with sufficient clarity to understand the methods and interpretation of the findings.

<Figure 1 here>

http://mc.manuscriptcentral.com/hec

Quantitative SP studies are survey-based valuation methods increasingly used in health services research to elicit individuals' preferences for goods or services (Clark et al., 2014; de Bekker-Grob et al., 2012; Smith & Sach, 2010). Guidelines for the design and conduct of SP studies recommend the use of qualitative methods (Bridges et al., 2011; Lancsar & Louviere, 2008). Existing guidelines have not provided a checklist on how to report the use of qualitative research alongside SP valuation studies. This study had three objectives to: identify whether existing qualitative critical appraisal tools are suitable for use with SP studies; if appropriate, make recommendations on how existing critical appraisal tools could be modified appropriately to produce a checklist; and, test the feasibility of using the suggested checklist in an exemplar systematic review of published healthcare SP valuation studies.

2. Methods

1

58 59 60 Existing qualitative methods critical appraisal tools (hereafter termed 'tools') were identified from guidelines published by the Centre for Reviews and Dissemination (CRD)(CRD, 2008). Manual searching of qualitative websites was combined with personal communication with two researchers with expertise in conducting qualitative healthcare studies. A tool was included if it was designed for appraising qualitative research generically, rather than for specific methods (e.g. ethnographic studies, interview studies).

The identified tools were summarised (independently by researchers XXX and XX) in a tabular format to facilitate direct comparison of the suggested quality assessment criteria and coverage of the existing tools. In addition, two published SP studies known to include a substantial qualitative aspect (Cheraghi-Sohi et al., 2007; Ryan et al., 2009) were used

http://mc.manuscriptcentral.com/hec

2 3 4 5 6 7 8 9 10 (independently by two researchers: XXX and XX) to test the identified tools that offered checklists to assess the quality of published qualitative studies. Discussions were then held between members of the research team (with primary expertise 11 12 13 in: health economics (n=3); economics (n=1); qualitative research (n=2)) and informed by the tabulated summary criteria with the exemplar use of data extracted using the identified tools. This iterative process aimed to reach consensus on the relevance of each tool to inform a checklist of the qualitative elements of SP studies. The process identified that important criteria were absent but some included criteria were afforded an unwarranted importance; consequently concluding that no existing tool adequately met the requirements for the purpose of assessing qualitative research used alongside SP studies. Discussion of existing tools produced a new set of criteria to be included in a preliminary version of a checklist. The feasibility and practical use of the proposed checklist was then tested in two published DCEs (Ryan et al., (2009) and Cheraghi-Sohi et al., (2007)) and two contingent valuation studies (Smith, (2007) and Shiell & Gold (2002)) all of which reported the use of qualitative research methods in either the survey design or interpretation of the results. Data were extracted from articles in accordance with the proposed checklist and further discussions were held between research team members to review how capable the checklist was of capturing the dimensions of the qualitative research conducted. This iterative process of testing, refining and discussing continued until an acceptable list of criteria was reached. The final version of the checklist was then agreed and further tested in a systematic review of qualitative research methods used alongside published healthcare DCEs.

3. Results

59 60

http://mc.manuscriptcentral.com/hec

Health Economics

 Seven published qualitative methods critical appraisal tools ('tools') were identified and reviewed (See Appendix A). The tools covered eight broad themes: 1) background and overview; 2) description of the context; 3) sampling methods; 4) data collection; 5) reflexivity (or a consideration of how the investigators themselves can influence qualitative research); 6) ethical aspects; 7) data analysis and 8) presentation of results. Across these eight themes a set of 32 specific criteria were identified. The number of criteria included in the tools ranged from seven to 42.

There was agreement (amongst all members of the research team) that none of the existing tools were appropriate to inform a checklist of qualitative methods in SP valuation studies for three key reasons that centred on the: rationale for using qualitative methods; reporting of philosophical foundations and theoretical underpinnings of the selected qualitative method; and practical use of the checklist. Certain aspects of some of tools were identified to be potentially relevant and served as the starting point for a new checklist.

The available tools did not allow an assessment of whether a clear rationale for using qualitative methods was reported in the SP study. This is an important omission regarding qualitative methods' use with SP valuation studies. It should be explicit whether qualitative research methods were employed in: 1) the design phase and/or 2) order to interrogate the validity of the quantitative data. These two rationale, respectively, have distinct implications for the study findings, as qualitative methods should have been used and reported with sufficient clarity to understand whether: 1) the data had been reduced to generate a final list of attributes and levels that were framed and presented in a way that respondents could understand; and 2) the results were in accordance with underlying economic theory and

http://mc.manuscriptcentral.com/hec

 $\begin{array}{c} 6\\ 7\\ 8\\ 9\\ 10\\ 11\\ 12\\ 13\\ 1\\ 15\\ 16\\ 17\\ 19\\ 20\\ 22\\ 22\\ 22\\ 22\\ 22\\ 22\\ 22\\ 22\\ 20\\ 31\\ 32\\ 33\\ 33\\ 35\\ 36\\ 37\\ 38\\ 39\\ 40\\ \end{array}$

Health Economics

rationale for the use of choice-based studies and the impact on model selection, econometric tests or assumptions.

One of the existing tools placed substantial emphasis on the need to report the philosophical foundations of using qualitative research methods (Long & Godfrey, 2004). Similarly, two of the tools (Long & Godfrey (2004) and Popay et al. (1998)) placed emphasis on reporting: the theoretical framework; the epistemological and ontological foundations; and evidence of conceptual adequacy. Details of these criteria may not be reported in a SP valuation study due to journal word restrictions and/or a perception that an economic audience is not sufficiently familiar or interested in these aspects. It was unanimously agreed that the omission of these aspects would not detract from an assessment of the quality of the qualitative research in the context of SP valuation studies.

In the 1990s, a consistent view emerged that mixed methods (qualitative combined with quantitative) studies in health services research would add value to knowledge and understanding (Barbour, 1999; Pope & Mays, 1993, 1995; O'Cathain et al., 2007). However, the use of mixed methods studies did not result in the desired effect, with criticism of each component, either in isolation or when combined, due to inadequate design and/or reporting (O'Cathain et al., 2008; Pope & Mays, 2009). The proposed new checklist, unlike existing tools, was designed to determine if, and how, qualitative and quantitative data were synthesised and included in the study design and/or interpretation.

http://mc.manuscriptcentral.com/hec

Health Economics

1

59 60 Page 8 of 22

Two further issues with using existing tools were identified. The tools failed to take into account the style of studies published in medical or health economics journals and the tension between reporting the mixed methods employed within tight word limits. The restriction of word count can, in part, be addressed by judicious use of supplementary appendices, but clarity is still required on methods and analysis. Another practical issue is that most health economists, who design SP valuation studies, are not trained in the use of qualitative methods. This is often addressed by working with qualitative researchers but then another tension emerges which is whether the study report should be focussed on the qualitative or quantitative aspects. The proposed checklist is designed to provide a means of rigorous peer-review of the qualitative component alongside the quantitative aspects of a SP valuation study by a non-expert user.

The iterative process produced a new checklist called 'CHEQUAL'; a mechanism to facilitate the systematic reporting and evaluation of the use of qualitative methods in SP valuation studies (see Appendix B). CHEQUAL could also guide researchers seeking to design a qualitative component of a SP study.

The final version of CHEQUAL was tested by using it to summarise if, and how, qualitative research methods were reported in published healthcare DCEs, generating a tabular summary using the 32 identified criteria. All healthcare DCEs published between 2001 and 2012 were identified (n=254) using an electronic search strategy based on de Bekker-Grob et al (2012). Of the 254 DCEs, 143 studies explicitly reported some use of qualitative methods and of these, 29 reported an extensive qualitative component either in the article or as a citation to previous work. A study was defined as having an extensive qualitative component (either in

http://mc.manuscriptcentral.com/hec

60

the article or in cited related work) when three conditions were met: a clear statement of the purpose of the qualitative research; an indication of the technique used to collect qualitative data; a description of how the qualitative data was examined/software used; and an explanation of the outcomes of the qualitative research. The feasibility of using CHEQUAL was demonstrated and the checklist was judged to identify key pieces of information which were, and were not, reported.

4. Discussion

The use of qualitative techniques to inform the design or understanding of SP studies is a relatively new but increasingly popular methodological direction within health economics. This study identified a need for a checklist designed specifically to check the extent to which, and how, qualitative research was conducted alongside SP studies. A practical and feasible checklist (CHEQUAL) was developed and tested. It is anticipated that this checklist will be of use to health economists, and their qualitative research co-investigators, to aid the clear and transparent reporting of the design, analysis and use of qualitative methods used in existing SP valuation studies. It could also prove a useful starting point for researchers embarking on a new SP study to inform the design of the qualitative component of the study. The proposed checklist also has a role in the conduct of systematic reviews of published SP valuation studies to critically appraise how qualitative methods were used alongside the quantitative aspects.

5. Conclusion

The development of CHEQUAL is timely given the increased use of qualitative research methods alongside SP valuation studies. The proposed new checklist was designed specifically to offer a transparent, fair and balanced mechanism for standardising the

http://mc.manuscriptcentral.com/hec

Health Economics

reporting of qualitative research alongside SP valuation studies. SP valuation studies have a vital role to play in health services research and as a source of information for resource allocation decision-making. The benefits of using qualitative methods alongside quantitative methods can only be realised if details are reported with clarity. CHEQUAL has been pretested in the context of healthcare DCEs but to meet multiple research and policy objectives the checklist should be further tested in wider application in the field of health economics.

59 60

http://mc.manuscriptcentral.com/hec

Acknowledgements

We would like to acknowledge the discussions with colleagues who attended a seminar that presented the potential use of CHEQUAL at the Health Economics Unit, University of Birmingham. In particular, we want to thank Dr Emma Frew for her considered suggestions at this seminar.

Declaration: The authors have no financial conflicts of interests. No ethical approval was required for the research conducted in this study.

http://mc.manuscriptcentral.com/hec

References

1

58 59 60

Baker, R. and Robinson, A. (2004). Responses to standard gambles: are preferences "w	vell
constructed"? Health Economics. 13 (1). 37-48.	

- Baker, R., Robinson, A. and Smith, R. (2008). How do respondents explain WTP responses? A review of the qualitative evidence. *Journal of Socio-Economics*, 37 (4), 1427–1442.
- Barbour, R. (1999). The case for combining qualitative and quantitative approaches in health services research. *Journal of Health Services Research & Policy*. 4 (1). 39–43.
- De Bekker-Grob, E.W., Ryan, M. and Gerard, K. (2012). Discrete Choice Experiments in Health Economics: A Review of the Literature. *Health Economics*. 21 (2). 145–172.
- Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3 (2), 77–101.
- Bridges, J.F., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L., Regier, D. a, Johnson, F. and Mauskopf, J. (2011). Conjoint analysis applications in health-a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health*. 14 (4), 403–413.
- Carson, R.T. and Louviere, J. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*. 49 (4), 539–559.
- CASP (2006). 10 questions to help you make sense of qualitative research. Public Health Resource Unit.
- Cheraghi-Sohi, S., Bower, P., Mead, N., McDonald, R., Whalley, D. and Roland, M. (2007). Making sense of patient priorities: applying discrete choice methods in primary care using "think aloud" technique. *Family Practice*. 24 (3). 276–282.
- Clark, M., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E.W. (2014). Discrete choice experiments in health economics: a review of the literature. *PharmacoEconomics*. 32 (9). 883–902.
- Coast, J. (1999). The appropriate uses of qualitative methods in health economics. *Health economics*, 8 (4), 345–353.
- Coast, J., Al-Janabi, H., Sutton, E., Horrocks, S.A., Vosper, J., Swancutt, D.R. and Flynn, T. (2012). Using qualitative methods for attribute development for discrete choice experiments : issues and recommendations. *Health Economics*. 21 (6), 730–741.
- Coast, J., McDonald, R. and Baker, R. (2004). Issues arising from the use of qualitative methods in health economics. *Journal of health services research & policy*. 9 (3). 171– 6.
- Corbin, J. and Strauss, A. (2008). Basics of qualitative research: Techniques and procedures for developing grounded theory. 3rd editio. Sage Publications.

http://mc.manuscriptcentral.com/hec
1

59 60

- CRD (2008). Systematic Reviews. CRD's (Centre for Reviews and Dissemination) guidance for undertaking reviews in health care. The Univer.
- Dixon-Woods, M. (2004). The problem of appraising qualitative research. Quality and Safety in Health Care. 13 (3), 223–225.
- Hsieh, H.-F. and Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*. 15 (9). 1277–88.
- Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A., Mauskopf, J. and Loder, E. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ British Medical Journal*. 346, f1049–f1049.
- Kløjgaard, M., Bech, M. and Søgaard, R. (2012). Designing a stated choice experiment: the value of a qualitative process. *Journal of Choice Modelling*. 5 (2). 1–18.
- Lancsar, E. and Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *PharmacoEconomics*. 26 (8). 661–677.
- Long, A. and Godfrey, M. (2004). An evaluation tool to assess the quality of qualitative research studies. *International Journal of Social Research Methodology*, 7 (2), 181–196.
- NCDDR (2004). What are the standards for quality research? National Center for the Dissemination of Disability Research. (9).
- O'Cathain, A., Murphy, E. and Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of health services research & policy*. 13 (2). 92–8.
- O'Cathain, A., Murphy, E. and Nicholl, J. (2007). Why, and how, mixed methods research is undertaken in health services research in England: a mixed methods study. *BMC health* services research, 7, 85.
- Pearson, A. and Field, J. (2011). JBI critical appraisal checklist for qualitative research. Joanna Briggs Institute.
- Popay, J., Rogers, A. and Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*. 8 (3), 341–351.
- Pope, C. and Mays, N. (2009). Critical reflections on the rise of qualitative research. BMJ British Medical Journal. 339.
- Pope, C. and Mays, N. (1993). Opening the black box : an encounter in the corridors ofhealth services research. BMJ British Medical Journal, 315–318.
- Pope, C. and Mays, N. (2008). Qualitative research in health care. 3rd editio. John Wiley & Sons.

http://mc.manuscriptcentral.com/hec

13

1 2 3 4 5 6 7 8 9	 Pope, C. and Mays, N. (1995). Qualitative research: reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. <i>BMJ British Medical Journal</i>. 311. Ryan, M., Watson, V. and Entwistle, V. (2009). Rationalising the "irrational": a think aloud study of a discrete choice experiment responses. <i>Health Economics</i>. 18, 321–336.
10 11 12 13 14	Shiell, A. and Gold, L. (2002). Contingent valuation in health care and the persistence of embedding effects without the warm glow. <i>Journal of Economic Psychology</i> . 23. 251– 262.
15 16 17	Smith, R.D. (2007). The relationship between reliability and size of willingness to pay values: a qualitative insight. <i>Health economics</i> . 216 (2). 211–216.
18 19 20 21	Smith, R.D. and Sach, T.H. (2010). Contingent valuation: what needs to be done? <i>Health economics, policy and law.</i> 5 (Pt 1). 91–111.
22 23 24	Walsh, D. and Downe, S. (2006). Appraising the quality of qualitative research. <i>Midwifery</i> . 22 (2). 108–19.
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 58 57 58 58 57 58 58 57 58 58 57 58 58 57 58 58 57 58 58 57 58 58 57 58 58 58 58 58 58 58 58 58 58	Xie, F., Pickard, a S., Krabbe, P.F.M., Revicki, D., Viney, R., Devlin, N. and Feeny, D. (2015). A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). <i>PharmacoEconomics</i> , 33 (8). 867–77.
60	14

http://mc.manuscriptcentral.com/hec

14

Appendix 5.10: The questionnaire sent to authors of DCE studies to assess their experience and views of using qualitative research methods



MANCHESTER 1824 The University of Manchester
Thank you for agreeing to complete this survey.
The purpose of this survey is to identify your views on the role and reporting of Qualitative Research Methods (QRMs) in
 Designing a discrete choice experiment (DCE) and/or Understanding the responses to a DCE.
NHS National Institute for Health Research National School for Primary Care Research
$\langle \neg \neg \rangle$
0% 100%

	MANCHESTER 1824	
	The University of Manchester	
How many DCEs in health	n care have you published, either as the first au	thor or as a co-author?
 1 2 3 4 5 - 9 10 or more 		
	National School for Primary Care Research	National Institute for Health Research
	$\langle \Rightarrow \Rightarrow \rangle$	
	0%	

MANCHESTER 1824 The University of Manchester	
For this question, please think about the study mentioned in the the invitation email.	
Do you think the paper accurately reflected the amount of qualitative research undertaken in the	
If no, could you please briefly explain your answer.	
Yes	
No No	
Opn't know	
NHS National Institute for	
Health Research National School for Primary Care Research	
$\langle \neg \downarrow \rangle$	
0%	

MANCHESTER 1824	
The University of Manchester	
For this question, please think about the study mentioned in the the invitation email	
When a member of the recentral team an event is suplicitive recentral?	
Regarding expertise in gualitative research	
If no, could you please briefly explain your answer.	
National Institute for	
Health Research	
National School for Primary Care Research	

	MANCHESTER 1824 The University of Manchester
A key finding of limited or no qua Does this findir O Yes No (please	my systematic review of DCEs in health care was that many studies report either litative research methods. g agree with your experience of reading or conducting health care DCEs? explain why)
	NHS National Institute for Health Research National School for Primary Care Research >



MANCHESTER 1824
The University of Manchester
Do you feel that the qualitative research completed as part of this health care DCE added value?
☐ It made a substantial improvement
 It added a little value
None, it did not add any value at all
\bigcirc No, it hindered the study and had a negative role
NHS National Institute for Health Research
$\langle \neg \neg \rangle$
02/

	MANCHESTER 1824 The University of Manchester	
Thinking about the study ment research completed as part of It made a substantial impro It added a little value None, it did not add any va No, it hindered the study ar	ioned in the the invitation email,do y your DCE added value? vement lue at all nd had a negative role	ou feel that the qualitative
Nat	ional School for Primary Care Resea	NHS National Institute for Health Research Irch
09/5	$\langle \neg \neg \rangle$	00%

MANCHESTER 1824 The University of Manchester
Thinking about all DCEs you have conducted, in general, how would you describe the impact of
It makes a substantial improvement
It adds a little value
O It has no impact at all
\bigcirc It has a negative impact and hinders the study progression
NHS National Institute for Health Research National School for Primary Care Research
0%

MANCHESTER 1824 The University of Manchester	
Is there anything you would like to add regarding the issues this surve questions we have asked you?	y has addressed or the
National School for Primary Care Research	NH5 National Institute for Health Research
0%	

ļ	MANCHESTER 1824 The University of Manchester	
	Would you like a copy of the results of this review? Yes No	
	National School for Primary Care Research	National Institute for Health Research
	0%	

MANCHESTER 1824 The University of Manchester
receive a copy of the results, please provide you email address in the space below.
NHS National Institute for Health Research National School for Primary Care Research
 0%



Appendix 5.11: The invitation to authors of DCE studies to complete the questionnaire to assess their experience and views of qualitative research methods alongside DCEs

Dear <TitleA> <Last Name>,

I am a PhD student sponsored by the National Institute for Health Research (NIHR) National School for Primary Care Research and based in the Manchester Centre for Health Economics at The University of Manchester, UK. I am supervised by Professor Katherine Payne looking at the framing of risk attributes in discrete choice experiments (DCEs).

I am writing to ask if you would complete an on-line survey that should take approximately **5 minutes** of your time. The purpose of this survey is identify your views on the role and reporting of qualitative research methods in DCEs

I have conducted a systematic review of the use of qualitative research methods in published DCEs. A summary of this systematic review is shown at the end of this email. A limitation of this systematic review is that it can only summarise the **reported** qualitative research methods. This is why I am contacting you, as an author of a published DCE, to complete this short survey. I want to understand more about researchers' views of using qualitative methods in DCEs with particular reference to your paper: <Title> in <Journal> (<Year>).

To complete the survey please click on this link <link>

Please note that your responses to this survey will be anonymously reported as a summary of the key findings.

Thank you very much for taking the time to read this email and I hope you are able to complete my survey for me.

Kind regards,

Caroline Vass

Manchester Centre for Health Economics NIHR School for Primary Care Research Doctoral Researcher Institute of Population Health, Jean McFarlane Building The University of Manchester Oxford Road, Manchester M13 9PL

Appendix 5.12: The reminder invitation to authors of DCE studies to complete the questionnaire to assess their experience and views of qualitative research methods alongside DCEs

Dear <TitleA> <Last Name>,

This is a friendly reminder that the invitation to complete this survey is still open. Your response would be greatly appreciated.

I am a PhD student sponsored by the National Institute for Health Research (NIHR) National School for Primary Care Research and based in the Manchester Centre for Health Economics at The University of Manchester, UK. I am supervised by Professor Katherine Payne looking at the framing of risk attributes in discrete choice experiments (DCEs).

I am writing to ask if you would complete an on-line survey that should take approximately **5 minutes** of your time. The purpose of this survey is identify your views on the role and reporting of qualitative research methods in DCEs

I have conducted a systematic review of the use of qualitative research methods in published DCEs. A summary of this systematic review is shown at the end of this email. A limitation of this systematic review is that it can only summarise the **reported** qualitative research methods. This is why I am contacting you, as an author of a published DCE, to complete this short survey. I want to understand more about researchers' views of using qualitative methods in DCEs with particular reference to your paper: <Title> in <Journal> (<Year>).

To complete the survey please click on this link <link>

Please note that your responses to this survey will be anonymously reported as a summary of the key findings.

Thank you very much for taking the time to read this email and I hope you are able to complete my survey for me.

Kind regards,

Caroline Vass

Manchester Centre for Health Economics NIHR School for Primary Care Research Doctoral Researcher Institute of Population Health, Jean McFarlane Building The University of Manchester Oxford Road, Manchester M13 9PL

Appendix 5.13: Author responses to the questionnaire of their experience and views of conducting qualitative research alongside DCEs

How many DCEs in health care have you published, either as the first author or as a co-author?				
	n	%		
1	9	17%		
2	10	19%		
3	9	17%		
4	0	0%		
5-9	16	30%		
10+	9	17%		

Why do you think there is this apparent absence / limited reporting of qualitative research methods? Some possible reasons are listed below, please tick all which apply. Qualitative research in DCEs...

Is not of interest to most of my peers	11	22%
Is not of interest to journals	22	44%
Is not of interest to funders	4	8%
Is not important in the design of health DCEs	3	6%
Does not affect the study outcomes	2	4%
Is too complicated to report in detail	26	52%
Is too time consuming to conduct properly	10	20%
Other reasons	28	56%

Do you feel that the qualitative research completed as part of this health care DCE added value?

	In this I	DCE	In DCEs generally		
	n	%	n	%	
It made a substantial improvement	29	58%	31	74%	
It added a little value	21	42%	11	26%	
None, it did not add any value at all	0	0%	0	0%	
No, it hindered the study and had a negative role	0	0%	0	0%	
For this question, please think about the study mentioned in the invitation email. Was a member of the					
research team an expert in qualitative research?					
Yes, I have expertise in qualitative research	13		25%		
Yes, a member of the research team had expertise in qualitative research	31		58%		
Do not know	1		2%		
No, there was no expert in qualitative research in the research team 8 15%					
Do you think the paper accurately reflected the amount of qualitative research undertaken in the study?					
Yes	26		50%		
No	24		46%		
Don't know	2		4%		
A key finding of my systematic review of DCEs in health care was that many studies report either limited					

A key finding of my systematic review of DCEs in health care was that many studies report either limited or no qualitative research methods. Does this finding agree with your experience of reading or conducting health care DCEs?

Yes	42	79%
No	11	21%

Appendix 5.14: Paper advertisement to the general public for participation in the think-aloud interviews

Understanding risk in healthcare: an interview study

Volunteers Required

How we travel, what we choose to eat, all and we would like to know your opinion of

We are looking for female volunteers aged 18-70 take place at a location convenient for you.

please contact Caroline Vass on the details

0771 675 1262 Tel

0 caroline.vass@postgrad.manchester.ac.uk

The University of Manchester

MANCHI

NHS

National Institute for Health Research

(2)	11 12	a a	1 1	1	1 S	1 1	
Caroline Vass 077 1675 126 Caroline Vass 077 1675 126 Caroline Vass	Caroline Vass 077 1675 126 Caroline Vass 077 1675 126	077 1675 126 Caroline Vass 077 1675 126 Caroline Vass	Caroline Vass 077 1675 126 Caroline Vass	077 1675 126 Caroline Vass	077 1675 126 Caroline Vass 077 1675 126 Caroline Vass	077 1675 126 Caroline Vass 077 1675 126 Caroline Vass	

Appendix 5.15: Email advertisements to The University of Manchester staff for participation in the think-aloud interviews

To: Distribution Lists

Subject: Call for Research Participants

Dear Staff,

I am a PhD student in the Manchester Centre for Health Economics. This is an invitation to take part in an interview that aims to understand your views about risk in healthcare and identify how balance the risks and benefits of breast cancer screening.

The interview will take approximately 45 minutes and can take place at a location convenient for you. The full information sheet about what to expect should you choose to participate is attached.

Please feel free to forward this email to anyone else, in or outside the university, who might be interested in participating in this research.

Many thanks,

Caroline Vass

PhD Student in Health Economics The Manchester Centre for Health Economics Institute of Population Health University of Manchester 4.306, Jean McFarlane Building Oxford Road Manchester M13 9PL Tel: +44 (0)161 306 7970 http://www.population-health.manchester.ac.uk/students/CarolineVass

Appendix 5.16: Think-aloud interview schedule

Introduction: My name is Caroline Vass and I'm a PhD student at the Manchester Centre for Health Economics. My PhD is about how people make a balance between risks and benefits when making choices in health care. I am particularly interested in how we can use surveys to understand their views on risks and benefits.

This interview is about how you balance risks and benefits in health care. I will ask you to complete a questionnaire and to say out-loud how you have come to your answers.

The interview should take about 45 minutes, and there will be imaginary examples looking at breast cancer screening programmes.

Have you read the information sheet? Please complete this consent form. So, are you ok to continue?

Topic 1: Warm-Up Exercise

I am going to ask you a question and I would like you to think aloud as you answer it. What I mean by think aloud is basically for you to say aloud everything that you would normally say to yourself or are thinking about silently.

I know that it is not something you would normally do but it will help me understand what you are thinking and how you came up with your answer. Do you understand what I would like you to do? Question 1: Would you mind telling me how many windows do you have in your house?

Topic 2: The DCE

Here is a survey made up of questions about imaginary breast cancer screening programmes. Take a few minutes to read through the information and definitions at the front of the survey. Do you have any questions at this point?

Question 1: As you complete the survey I want you to talk aloud by trying to say out-loud what you're thinking as you're completing this questionnaire.

Don't worry about making sense or talking to me, I am merely going to listen to what you have to say. If you go a bit quiet, I might ask some questions to get you talking again. Is that ok?

Probes:

- What are you thinking now?
- Why did you choose that one?
- Would you choose that if I wasn't here?

• Are you considering all of the information presented?

[If respondent is choosing A's or B's, probe why...]

[If respondent fails internal tests [QUESTION 6], probe why...]

Topic 3: Ease of Completion

Question 1: What did you think about the questions?

Question 2: Did you find yourself concentrating on a particular characteristic or do you think you

weighed them up evenly?

Question 3: Would you change any of your answers?

Question 4: Do you think you are satisfied with your choices?

Topic 4: Further questions

These questions are just so I can understand if different people see things differently.

Question 1: What is your current employment status? What kind of work do / did you do? Was/Is that full

or part-time? Do you have to use numbers in this job?

Question 2: Do you play games that involve a gamble? If not clear, suggest betting on horses or bingo.

Where do you play <insert game>? Is that online?

Topic 5: Final thoughts

Do you have any other feedback or thoughts?

Thank you for completing this interview.

Appendix 5.17: Flow of recruited study participants for the think-aloud interviews and the eye-tracking study



Appendix 6.1: Paper advertisement to the general public for participation in the eye-tracking study

Understanding risk in healthcare:

an eye-tracking study

Volunteers Required

We make choices that involve risk all the time – deciding how to travel and what to eat and drink can all involve choices about risk. This is true in healthcare – for example deciding whether to undergo a treatment that could improve our health, but might involve side effects. To help patients decide what to do, doctors sometimes provide information that helps explain different risks

We are looking for female volunteers aged 18 to 70 to take part in a study at the University in which volunteers will be asked to look at information about breast cancer screening programmes. We'll ask their opinions about the information provided and check what information people's eyes are drawn to. It should take no more than 25 minutes.

For more information please contact Caroline Vass



NHS National Institute for Health Research

Tel 0771 675 1262

caroline.vass@postgrad.manchester.ac.uk

	1824
The University	of Mancheste

										1			
077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1263	077 1675 1262	077 1675 1263	077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1262	077 1675 1262
Caroline Vass													

Appendix 6.2: Email advertisement to The University of Manchester staff for participation in the eye-tracking study

To: Distribution Lists

Subject: Call for Research Participants

Dear Staff,

I am a PhD student in the Manchester Centre for Health Economics. This is an invitation to take part in an interview that aims to understand your views about risk in healthcare and identify how balance the risks and benefits of breast cancer screening.

The interview will take approximately 45 minutes and can take place at a location convenient for you. The full information sheet about what to expect should you choose to participate is attached.

Please feel free to forward this email to anyone else, in or outside the university, who might be interested in participating in this research.

Many thanks,

Caroline Vass

PhD Student in Health Economics
The Manchester Centre for Health Economics
Institute of Population Health
University of Manchester
4.306, Jean McFarlane Building
Oxford Road
Manchester M13 9PL
Tel: +44 (0)161 306 7970
http://www.population-health.manchester.ac.uk/students/CarolineVass

Appendix 6.3: Information sheet for the eye-tracking study participants



NHS National Institute for Health Research

National School for Primary Care

Understanding risk in healthcare: an eye-tracking study

You are being invited to take part in a research study that aims to understand how people complete surveys. The example we will use is choosing a screening programme for breast cancer. Before you decide whether to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the aim of the research?

This study wants to understand how people make choices about healthcare. The aim is to explore people balance the risks and benefits when making a decision about breast cancer screening.

Why have I been chosen?

This study aims to explore the views of women. Any female member of the public who is able to complete a survey in English is able to take part in this study.

What would I be asked to do if I took part?

You will be invited to partake in an eye-tracking study at the University of Manchester. The study should take no more than 15 minutes. During the study I will ask you to complete a questionnaire with different imaginary breast cancer screening programmes and you will be asked which you prefer. Whilst you are completing the questionnaire, your eye-movements will be monitored, which is like having an eye-test at an opticians. At the end of the questionnaire I will ask you to complete a short survey containing questions about the choices you made.

What happens to the data collected?

The eye-tracking data will be collected to see if there are patterns in people's answers and where they were looking whilst completing the survey. The answers to the survey will be used to help us understand how breast cancer screening programmes can be better provided. The survey questions will help us understand why different people might make different choices. Your name will not be recorded and instead you will be given an anonymous identification number.

How is confidentiality maintained?

The law called the Data Protection Act (1998) tells us how to keep your information secure. If you provide personal information (name, address, contact details) it will remain confidential and we will not give your details to anyone else.

Will I be paid for participating in the research?

We are not able to pay you for taking part in this study. We can reimburse your travel expenses for your journey to the campus in line with university policy. To thank you for your time, you will receive an Amazon voucher for participating in this research.

Where will the research be conducted?

The research will take place on The University of Manchester's main campus, in the Zochonis Building on Brunswick Street. This building is number 60 on the campus maps which are located around the campus.

Will the outcomes of the research be published?

The main outcome of this research will be a PhD thesis for Caroline Vass, a student based in the Manchester Centre for Health Economics at The University of Manchester. In addition, we may want to report the findings at conferences or in a published journal article.

My question is not covered here, who do I contact for more information?

Please contact Caroline Vass (PhD Student in the Centre for Health Economics) on 077 1675 1262 or email caroline.vass@postgrad.manchester.ac.uk.

<u>Thank you for reading this information sheet</u> <u>Please keep this information sheet for your records</u>

Appendix 6.4: A description of the eye-tracking pilot study

A pilot eye-tracking DCE was designed to understand the preferences of female members of the public (recruited by posters in local cafes) for a breast screening programme described by three attributes (probability of detecting a cancer, risk of unnecessary treatment, and out-of-pocket cost) each with four levels. Two survey versions were used that varied how the risk attributes (probability of detecting a cancer and risk of unnecessary treatment) were presented as: (1) a percentage or (2) a percentage and icon array. Eye-movements were recorded as a series of co-ordinates 1,000 times a second. Eye-tracking data were analysed in terms of direction of motion and total visual attention (dwell time) to pre-defined areas of interest using descriptive statistics. Immediately after completing the last choice question, respondents were asked a series of debriefing questions. The effect of each attribute on the women's preferences were analysed using a conditional logit model.

Fifteen completed the DCE in the eye-tracking experiment. Results of the pilot study found respondents gave significantly more visual attention, indicating information processing, to both risk attributes when risk was communicated with an icon array rather than solely as a percentage with a mean dwell time of 6316 and 5043 milliseconds, respectively. Respondents to the icon array version also exhibited significantly more upwards and downwards eye-movements (43% v 38% of saccades) suggesting calculations were made in line with expected utility theory possibly reflecting a greater understanding of the risk information. The eye-tracking data confirmed the self-reported attribute non-attendance as stated by respondents when asked the de-briefing questions with significantly lower (by almost 70%) mean dwell times to these attributes. The results of the conditional logit revealed both probability of detecting a cancer and the risk of unnecessary treatment were significant in women's decision to partake in breast screening.

Appendix 6.5: Consent form for participants in the eye-tracking study



NHS National Institute for Health Research

School for Primary Care Research

Consent form Understanding risk in healthcare

		Please tick box
 I confirm that I have read the attached info project and have had the opportunity to cons questions and had these answered satisfact 	ormation sheet on the abov sider the information and as orily.	sk
2. I understand that my participation in the st free to withdraw at any time without giving a	tudy is voluntary and that I reason	am
3. I agree to the use of anonymous quotes ir	n academic articles.	
I agree to take part in the above project		
Name of participant	Date	Signature
Name of person taking consent	Date	Signature

Appendix 6.6: iPad survey for the eye-tracking study



NH5 National Institute for Health Research chool for Primary Care Researd



The University of Manchester

FeedbackB

On a scale of 1-5, how easy or difficult did you find making choices between the alternatives?

(very easy)	2 nor	easy) 4	(Very difficult)
FeedbackB_r1=1 Feedba	ckB_r1=2 Feedba	ckB_r1=3 FeedbackB	_r1=4 FeedbackB_r1=5

FeedbackBexpl

Please explain your answer

$\langle \rightarrow \rangle$	

0% 100%

National Institute for Health Research chool for Primary Care Research



FeedbackB1

On a scale of 1-5, how easy or difficult did you find the survey to understand?

1		3 (Neither difficult		5
(Very easy)	2	nor easy)	4	(Very difficult)
FeedbackB1_r1=1	FeedbackB1_r1=2	FeedbackB1_r1=3	FeedbackB1_r1=4	FeedbackB1_r1=5

FeedbackB2

Was there anything in the survey that was particularly unclear?

FeedbackB2=1	No, I understood everything	
FeedbackB2=2	FeedbackB2_2_other There were some (please comment in the box below)	things that weren't clear or well explained

FeedbackBexpl1

Is there anything else about the survey we could change to make it easier to understand (e.g phrasing, explanations)?



National Institute for Health Research hool for Primary Care Research



FeedbackANA

Did you find yourself making choices based on one or two characteristics of the screening programme?

FeedbackANA=1 Yes, I based choices on one or two characteristics

FeedbackANA=2 No, I considered ALL characteristics

FeedbackANAdrop

FeedbackANAdrop1

Would you say you looked at the characteristic <u>risk of unnecessary treatment</u>?
Please select

FeedbackANAdrop2

Would you say you looked at the characteristic <u>cost</u>?
Please select

FeedbackANAexpl

 \cup

If you based your choices on one or two characteristics, why was this? (*Please tick all that apply*):

FeedbackANAexpl_1 There was just too many characteristics on which to make a decision

FeedbackANAexpl_2 The other characteristic(s) were too complicated or it was unclear what it meant

FeedbackANAexpl_3 FeedbackANAexpl_3_other Other reason (please explain)

() (%)

> NHS National Institute for Health Research



BackgroundB

What is your occupational status?



BackgroundC

What is your occupation? (Or previous occupation if retired?)

$\langle \rangle$	

0%

National Institute for Health Research bool for Primary Care Research



BackgroundD

What is the highest level of education you have obtained?



NHS National Institute for Health Research col for Primary Care Research

		MANCHESTER	
		The University of Manchester	
BackgroundE			
What is your e	thnic group?		
BackgroundE=1	White British/Irish		
BackgroundE=2	White other		
BackgroundE=3	Mixed		
BackgroundE=4	Black/Black British		
BackgroundE=5	Asian/Asian British		
BackgroundE=6	BackgroundE_6_other Other (please specify)		
		2	
	0%		100%

National Institute for Healtin Research School for Primary Care Research



BackgroundF

What is your religious group? BackgroundF=1 No religion BackgroundF=2 Christian BackgroundF=3 Buddhist BackgroundF=5 Hindu BackgroundF=6 Muslim BackgroundF=7 Sikh BackgroundF=8 BackgroundF_8_other Other (please specify)

> NHS National Institute for Health Research



BackgroundG

BackgroundG=1 BackgroundG=2 BackgroundG=2 No

	$\langle \rangle$	
0%		100%

National Institute for Health Research theol for Primary Care Research



BackgroundH	

Do you have any daughters?

BackgroundH=2	No	
\odot	NO	

0% 100%

NHS National Institute for Health Research School for Primary Care Research



ExperienceA

Have you ever had breast screening (a mammogram)?



ExperienceB

Have you received follow-up (such as an ultrasound, a repeat mammogram or a biopsy) for breast cancer?

ExperienceB=1	Yes
ExperienceB=2	No
ExperienceB=3	Don't know

	$\langle \rangle$	
0%		100%

National Institute for Health Research theol for Primary Care Research



transporttoscreen

Thinking about the last time you went for breast screening, what was your <u>main</u> form of transport to the screening?



costofscreen

Still thinking about your last breast screen, how much did it cost you for your return journey (even if your start and end points were different)? (If you did not pay any transport fare then please enter 0)

timeoffwork

Did you incur any other "costs"?

Wet/onel institute for Health Research School for Primary Gre Research



ExperienceC

Have you, or has anyone in your family (blood relative or otherwise) had breast cancer?



ExperienceC1

Do you know of anybody else (e.g friends or colleagues) who has had breast cancer?

ExperienceC1=1	Yes	
ExperienceC1=2	No	
ExperienceC1=3	Don't know	
	0%	100%

National Institute fo Health Research Chool for Primary Care Research


The University of Manchester



RiskD1

How would you describe your future risk of breast cancer?



RiskE

To what extent are you concerned about your own risk of breast cancer?



RiskF

Is there anything else you would like to tell us about your risk of breast cancer?



Warloval Destinate for Hould's Beanard hould for Primary Care Research



The University of Manchester

FeelRiskA

On a scale of 1-5, how much do you agree with the following statements?

1 (Strongly disagree)	2 (Disagree)	3 (Neither agree nor disagree)	4 (Agree)	5 (Strongly agree)
FeelRiskA_r1=1	FeelRiskA_r1=2	FeelRiskA_r1=3	FeelRiskA_r1=4	FeelRiskA_r1=5

FeelRiskB

"I ae	et a thrill when	n I gamble, eve	en if it's just	the lottery"
1 90	ca unni whe	i i gambie, eve		are routery

		3		
1	2	(Neither agree nor	4	5
(Strongly disagree)	(Disagree)	disagree)	(Agree)	(Strongly agree)
FeelRisk8_r1=1	FeelRiskB_r1=2	FeelRiskB_r1=3	FeelRiskB_r1=4	FeelRiskB_r1=5
	0	\sim		

FeelRiskC

"Even if something is unlikely, I feel it will probably happen to me"

1 (Strongly disagree)	2 (Disagree)	3 (Neither agree nor disagree)	4 (Agree)	5 (Strongly agree)
FeelRiskC_r1=1	FeelRiskC_r1=2	FeelRiskC_r1=3	FeelRiskC_r1=4	FeelRiskC_r1=5
		$\langle \neg \neg \rangle$		

Wet/ovial institute for Health Beanaith School for Primary Care Research



NumeracyA

This next section will ask you some questions to understand how familiar you are with probabilities.

If you are unsure of the answer, you can skip the question.

Imagine I flip a coin 1,000 times. What is your best guess about how many times the coin would fall heads up in 1,000 flips?

NumeracyB

If there was a lottery where the chance of winning £10 is 1%, if 1,000 people bought a ticket, how many people would you expect to win?

NumeracyC

In another lottery, the chance of winning a car is 1 in 1,000. What percentage of tickets in the lottery will win a car?

	$\langle $	
0%		100%

Wotional Institute for Houlth Research choil for Primary Care Research



The University of Manchester

num2

You missed some of the previous questions about statistics and probability. Why was this?

0	Just too hard
num2=2	Were going to take me too long to answer
num2=3	Can't see the relevance to the study
num2=4	num2_4_other Other
	0%

Wet/oval Institute for Health Respond



Comments2

Can you think of any thing else which might help us understand how you made your choices in the choice task?

	$\langle \rangle$	
0%		100%

Wational Institute for Health Bearaith



help

Thank you for completing this survey.

If you have been affected by any of the issues raised in this survey, the following organisations may be of assistance:

- National Hereditary Breast Cancer Helpline on 01629 813000 - Breast Cancer Care on 0808 800 6000

If you have any questions or concerns please contact Caroline Vass on 077 1675 1262 or caroline.vass@postgrad.manchester.ac.uk.

Manchester Centre for Health Economics, Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL

0%		100%

National Instit Health &

Appendix 6.7: Heteroskedastic conditional logit results for the eye-tracking study

Heteroskedastic conditional logit model (linear risk, pooled sample)					
	ASC (none)	-1.565***	(0.41)		
	Detect	0.141***	(0.03)		
Utility	Risk	-0.076***	(0.02)		
	Cost	-0.031	(0.03)		
	IAP*detect	1.789	(4.29)		
	IAP*risk	-0.901	(2.07)		
Scale	Icon arrays and percentages(IAP)	-2.152	(2.12)		
	Log likelihood	n	Obs		
	-329.87321	42	1386		

* p<0.005; ** p<0.01; ***p<0.001; standard errors in parentheses