Does changing the way a discrete choice experiment (DCE) is presented to respondents affect results? An investigation in the context of health using between-subject designs

Wen Shi Alice Yu

Centre for Health Economics Research and Evaluation (CHERE)

University of Technology Sydney

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Submitted December 2021

Certification of Original Authorship

I, ***Wen Shi Alice Yu*** declare that this thesis, is submitted in fulfilment of the requirements for the award of ***Doctor of Philosophy*** in the ***Business School*** at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: **10/12/2021**

Contents

Certification of Original Authorship ii
Contents iii
List of Tables x
List of Figures xiv
Acknowledgements xvi
Thesis Abstract xviii

Chapter 1. Introduction 1
1.1 The case for using stated preference techniques in health services research 1
1.2 Type of stated preference techniques used in health services research 2
1.2.1 Contingent valuation (CV) studies 3
1.2.2 Standard gamble (SG) 3
1.2.3 Time trade-off (TTO) 4
1.2.4 Discrete choice experiments (DCEs) 5
1.2.5 Related techniques to DCEs: conjoint analysis (CA) 7
1.2.6 Related techniques to DCEs: best–worst scaling (BWS) DCEs 7
1.3 DCEs in health services research 8
1.4 Aims of thesis 9
1.5 Overview of chapters 10

Chapter 2. Methods 11
2.1 Definition of a DCE 11
2.2 How to create a choice set 12
2.2.1 Developing attributes and levels 12
2.2.2 Construction method of choice sets 13
2.3 DCEs and utility theory 14
2.4 Analysis of DCE data 15
2.4.1 Multinomial logit (MNL) model 16
Predicted choice probabilities 16
2.4.2 Latent class analysis (LCA) 17
2.4.3 Mixed logit (MXL) model 17
Interpretation of MXL model results: kernel density plots 18
2.4.5 Scaled MNL (S-MNL) model 18
2.5 Summary 19

Chapter 3. Different ways to present a DCE to respondents: A scoping review 20
3.1 Overview 20
3.2 Summary of previous reviews 20
3.2.1 Aims of scoping review 22
3.3 Framework used to investigate presentation differences in a DCE 23

3.3.1 What is considered part of the presentation of the DCE? 23
3.3.2 Between-subjects design 23
3.3.3 Presentation difference in a DCE 24
3.3.4 Main types of presentation differences used in DCEs 24
1. How information is presented to respondents in a DCE 25
1a. How information is presented to respondents in a DCE: differences in text 25
1b. How information is presented to respondents in a DCE: use of alternative presentation formats to text 26
2. Amount of information provided 27
3. Structure of choice sets 27
3.4 Methods 27
3.4.1 Literature search strategy 27
Inclusion criteria 27
Exclusion criteria 28
Search process 28
3.4.2 Data extraction details 31
3.5 Results: characteristics of studies 31
3.5.1 General study details 31
Country of study 31
Dissemination method of DCE 32
Population of interest 32
Study sample size 33
3.5.2 Study development details 33
Use of qualitative methods 33
Other methods used to develop study 34
3.5.3 Choice set presentation details 37
Choice sets 37
Number of options and using opt out 38
Number of attributes 39
Number of attributes vs number of choice sets per respondent 40
3.5.4 Choice set construction details 41
3.6 Results: Presentation differences and DCE outcomes 44
3.6.1 Arms in study 44
3.6.2 Which aspect/s of the DCE were different between arms? 46
3.6.3 How have presentation difference been used? 47
Summary of studies: differences in text 47
Summary of studies: amount of information provided 49
Summary of studies: differences in presentation formats 49
Summary of studies: structure of choice sets 49
Summary of studies: multiple presentation difference types 49
3.6.4 Analysis of the impact of presentation differences 50
Models used to analysis DCE data 50
Outcomes 51

3.7 Discussion 52
3.7.1 Summary of findings 52
3.7.2 Limitations of findings and directions for future research 53
3.7.3 Implications of findings 55
3.7.4 Next steps 55

Chapter 4. Preferences for the assessment of chemotherapy induced peripheral neuropathy (CIPN): The patient perspective 57
4.1 Introduction 57
4.2 Background 59
4.2.1 IN FOCUS Study 61
4.2.2 DCE theoretical framework 61
4.3 Methods 62
4.3.1 Initial attributes and levels 63
4.3.2 Refining attributes and levels: qualitative Interviews 64
Sample for cognitive interviews 65
Creation of the DCE for cognitive interviews 65
Information provided to participants in cognitive interviews 66
Creation of the survey for cognitive interviews 66
Cognitive interview process 66
Summary of results from cognitive interviews 67
4.3.3 Refining attributes and levels: consultation with health economists and DCE experts 68
Introduction 68
Key takeaways from consultation session 69
4.3.4 Finalised attributes and levels 69
4.3.5 Construction of choice sets 70
4.3.6 Using simulations to assess the performance of choice sets constructed 72
Purpose of simulations 72
Determining sets of priors 72
Determining number of responses for simulations 72
Assessment of simulations 73
4.3.7 Study 1: patient Sample 76
Ethics approval 76
Piloting 77
Survey flow 77
Sample 82
Survey launch and issues 82
Data analysis 83
4.4 Study 1: Patient sample results 83
4.4.1 General descriptive statistics of respondents 83
Cancer-related descriptive statistics 85
Summary of attributes and levels with abbreviations 88

4.4.2 MNL model results 89
4.4.3 MXL model results 91
4.4.4 Examining MXL model results: choice probabilities 94
4.4.5 Examining MXL model results: kernel density plots 97
4.4.6 Summary of responses to evaluation questions 100
4.5 Discussion of Study 1: patient sample results 101
4.5.1 Statement of principal findings 101
4.5.2 Strengths and weaknesses of the study 102
4.5.3 Study implications for clinicians or policy makers 102
4.6 Next steps 102

Chapter 5. Preferences for the assessment of chemotherapy induced peripheral neuropathy (CIPN): Contrasting the general population and patient perspectives 103
5.1 Background 103
5.2 Study 2: Methods 105
5.2.1 Overview of Study 2 105
5.2.2 Background information section of DCE 107
5.2.3 Experimental arms 107
5.2.4 Summary of attributes, levels and choice sets 107
5.2.5 Development of survey 108
5.2.6 Ethics approval 108
5.2.7 Recruitment 108
5.2.8 Planned sample size 109
5.2.9 Testing poolability of data from different sources 109
5.3 Study 2: Results 110
5.3.1 Sample characteristics 110
5.3.2 Analysis of general population sample 113
Overview of analyses 113
MNL model results by arm 114
Sensitivity analyses using the MNL model 115
S-MNL model results 115
MNL model: testing preference differences by arm 116
Final MNL model: pooled data 118
Latent class analysis (LCA): pooled data 119
MXL model results: pooled data 120
5.3.3 Responses to evaluation questions 126
5.3.4 Experience with illness and cancer 131
5.4 Comparison of Study 1 and Study 2: patient and general population preferences 134
5.4.1 Overview 134
5.4.2 Investigating scale differences between samples 134
5.4.3 Comparison using ratios 136
5.4.4 Comparison by utility ranking of possible assessment tools 138

5.5 Discussion 140
5.5.1 Study 2: Summary of findings 140
5.5.2 Comparison of Study 1 and 2: summary of main findings 141
5.5.3 Strengths and weaknesses of the study 142
5.5.4 Implications of findings and directions for future research 143

Chapter 6. Anchoring choice sets to immediate death or full health during the valuation of the EQ-5D-5L 146

6.1 Overview 146
6.2 Author involvement in the project 146
6.3 Background 147
6.3.1 Incorporating health effects into economic evaluation 147
6.3.2 Defining quality-adjusted life years (QALYs) 147
6.3.3 The use of multi-attribute utility instruments (MAUIs) 149
6.3.4 The EQ-5D-5L 149
6.3.5 Deriving QALY weights for the EQ-5D-5L 151
6.3.6 Using DCEs to derive QALY weights for the EQ-5D-5L 152
6.4 Aims 154
6.4.1 Details of the project 155
6.5 Methods 155
6.5.1 The data 155
6.5.2 Data collection: Peru 155
6.5.3 Data Collection: Denmark 156
6.5.4 Arms in study 156
6.5.5 List of attributes and levels 157
6.5.6 The task 158
6.5.7 Choice set construction method 160
6.6 Data Analysis 160
6.6.1 Modified utility specification 160
6.6.2 MNL model with duration 161
6.6.3 Reporting of MNL model parameter estimates: calculating QALY weights 161
6.6.4 MXL model with duration 162
6.7 Data cleaning and screening of respondents 163
6.7 Results 164
6.7.1 Poolability of arms across countries 164
6.7.2 MNL model results 164
Predicted health states utilities and position of 'dead' 165
6.7.3 MNL model results: Cross country comparison 168
6.7.4 MNL model results: significance of adjacent utility decrements 171
6.7.5 MXL model results 172
6.8 Discussion 176
6.8.1 Summary of findings 176
6.8.2 Limitations and direction for future research 177

6.8.3 Implications of findings 178
Chapter 7. Discussion 179
Introduction 179
Chapter 3: Scoping review 180
Summary of findings 180
Implications of findings 181
Chapters 4 and 5: Preferences for the assessment of CIPN 182
Summary of findings 182
Implications of findings 184
Chapter 6: Anchoring to immediate death versus full health 186
Summary of findings 186
Implications of findings 187
Limitations of findings and directions for future research 188
Conclusions 189
What are the main ways a DCE can be presented to respondents? 190
Does changing how the DCE is presented to respondents lead to differences in results? 190
Final remarks 191
References 192
Appendices 205
Appendix 3A: Studies included in Scoping Review 205
Appendix 3B: Definitions and scope of general study details extracted 212
Appendix 3C: Definitions and scope of study development details 213
Appendix 3D: Definitions and scope of validity check details 214
Appendix 3E: Definitions and scope of choice set presentation details 215
Appendix 3F: Definitions and scope of choice set construction details 216
Appendix 3G: Definitions and scope of arm details 217
Appendix 3H: Definitions and scope of data analysis details 218
Appendix 4A: Ethics approval for qualitative interviews 219
Appendix 4B: Cognitive interview protocol, Interviews 1-4 220
Appendix 4C: Cognitive interview protocol, Interviews 5-6 222
Appendix 4D: Prompting questions 225
Appendix 4E: Summary of cognitive interviews 226
Appendix 4F: Consultation session with health economists and DCE experts 229
Appendix 4G: Sets of priors used in simulations 232
Appendix 4H: Ethics approval No. ETH19-3464: copy of email 234
Appendix 4I: Copy of full survey 236
Appendix 4J: Copy of email newsletter by IN FOCUS 251
Appendix 4K: Copy of the wording of email to relaunch survey on 1st of November 2019 253
Appendix 4L: Respondent comments 254
Appendix 5A: CIPN DCE general population sample copy of survey 256

Appendix 5B: CIPN DCE general population sample ethics approval (copy of letter) ... 274
Appendix 5C: Excluded respondents comment box text ... 275
Appendix 5D: Sensitivity analysis: MNL model results with fastest 10% removed by arm ... 276
Appendix 5E: Sensitivity analysis: Arm 2 with 4 respondents that could not view the video removed ... 277
Appendix 5F: S-MNL results general population arms combined ... 278
Appendix 6A: Durations used in choice sets ... 279
Appendix 6B: Test for poolability of arms by country ... 280
Appendix 6C: MNL parameter estimates on latent scale ... 281

List of Tables

Table 3.1 Search terms....29
Table 3.2 Searches by research database....29
Table 3.3 Studies included by year....31
Table 3.4 Country of study....32
Table 3.5 Dissemination method of DCE....32
Table 3.6 Population of interest....33
Table 3.7 Study sample size summary....33
Table 3.8 No. of qualitative methods used....34
Table 3.9 Type of qualitative methods used....34
Table 3.10 Rationale for qualitative methods....34
Table 3.11 No. of methods used....35
Table 3.12 Type of method used....35
Table 3.13 Rationale for method use....35
Table 3.14 Types of validity checks....36
Table 3.15 Total number of choice sets....37
Table 3.16 No. of choice sets per respondent....38
Table 3.17 How were choice sets assigned?....38
Table 3.18 No. of options....39
Table 3.19 What is the construction method designed to model?....42
Table 3.20 What was the construction method optimised for?....42
Table 3.21 What software was used for construction?....43
Table 3.22 Method to construct choice sets....43
Table 3.23 Method of construction: other....43
Table 3.24 No. of arms in study....44
Table 3.25 How were arms assigned?....45
Table 3.26 By convenience: details....45
Table 3.27 By criteria: details....45
Table 3.28 No. of aspects of the DCE different between arms....46
Table 3.29 List of parts different between arms....46
Table 3.30 No. of presentation differences used per study....47
Table 3.31 Frequency of different presentation difference types....47

Table 3.32 No. of models used 50
Table 3.33 List of models used 51
Table 3.34 Other modelling procedures used 51
Table 3.35 Any differences in DCE results by arm? 52
Table 4.1 CIPN DCE: initial attributes and levels 63
Table 4.2 Project 1: vignette for choice sets 70
Table 4.3 Project 1: finalised attributes and levels 71
Table 4.4 Summary of respondent numbers used 72
Table 4.5 Standardised bias summary 76
Table 4.6 Education level of respondents 84
Table 4.7 Summary of assignment to block 84
Table 4.8 Number of females and males by assigned block 84
Table 4.9 Mean and median completion times by block to the nearest minute (mins) 85
Table 4.10 Years since first cancer diagnosis 86
Table 4.11 Cancer type (can select more than one) 86
Table 4.12 Type of chemotherapy drugs received (can select more than one) 87
Table 4.13 List of attributes with abbreviations 88
Table 4.14 Summary of MNL model results 89
Table 4.15 Predicted choice probabilities (MNL) 90
Table 4.16 MXL model 1 91
Table 4.17 MXL model variations: AIC and BIC summary 92
Table 4.18 Final MXL model 94
Table 4.19 Choice probabilities: attributes with no heterogeneity 95
Table 4.20 Choice probabilities: attributes with heterogeneity 96
Table 4.21 Responses to question 'How did you decide whether Assessment A or B was better in each choice question?' 100
Table 5.1 Survey status 111
Table 5.2 Frequency of gender and age of respondents 111
Table 5.3 Education level 112
Table 5.4 Block assignment 112
Table 5.5 Completion time 113
Table 5.6 MNL results by arm 114
Table 5.7 S-MNL model results 116

Table 5.8 MNL results with dummy indicator for arm 117
Table 5.9 MNL results 118
Table 5.10 LCA results 120
Table 5.11 MXL model with all parameters entered as random 121
Table 5.12 Summary of AIC and BIC 122
Table 5.13 Final MIXL model for pooled data 122
Table 5.14 Helpfulness of introductory information 126
Table 5.15 Arm 2 video quality rating 127
Table 5.16 Level of agreement with helpfulness of the physical test type explanations 127
Table 5.17 Testing for whether level of agreement with helpfulness of physical test explanations is dependent on arm 128
Table 5.18 Level of agreement with ability to imagine symptoms 128
Table 5.19 Level of agreement with ability to imagine impacts on daily activities 129
Table 5.20 Testing for whether ability to imagine is dependent on arm 129
Table 5.21 Strategy for choosing preferred assessment 130
Table 5.22 Level of agreement: ease of identifying differences between assessment tools 130
Table 5.23 Level of agreement: ease of choosing a preferred assessment tool 130
Table 5.24 Testing for whether ease of identifying/choosing is dependent on arm 131
Table 5.25 Frequency of conditions 132
Table 5.26 Frequency of cancer diagnosis type 132
Table 5.27 Someone you know with cancer? 132
Table 5.28 What was their cancer diagnosis? 133
Table 5.29 S-MNL: patient and general population samples 135
Table 5.30 Ratios of parameters 137
Table 5.31 Ratios of MRS 137
Table 6.1 EQ-5D-5L levels with health state 13245 underlined and bolded for illustration 151
Table 6.2 Arms in study 157
Table 6.3 List of attributes and levels 157
Table 6.4 Arm death_arm: mock-up of health state A vs B (different durations) 159
Table 6.5 Arm death_arm: mock-up of health state B vs C (anchored to immediate death) 159

Table 6.6 Total number of respondents by arm and country 164
Table 6.7 Total number of choice sets by arm 164
Table 6.8 MNL results 165
Table 6.9 Example of calculation of utility weight using the PIT state 166
Table 6.10 Example of calculation of utility weight with constrained values 166
Table 6.11 Summary of health state utilities for Peruvian respondents 167
Table 6.12 Dummy coding for five level attribute 171
Table 6.13 Incremental dummy coding for five level attribute 171
Table 6.14 MNL model results: significance of adjacent utility decrements 172
Table 6.15 Peru: AIC and BIC 173
Table 6.16 Denmark: AIC and BIC 173
Table 6.17 Peru: MXL model latent scale results 174
Table 6.18 Denmark: MXL model raw parameter estimates 175
Appendix Table 1 Arm death_arm duration combinations for health state A and B 279
Appendix Table 2 Arm fullhealth_arm duration combinations for health states A and B 279
Appendix Table 3 Arm fullhealth_arm frequency of durations for health states B vs C (full health) 279
Appendix Table 4 MNL parameter estimates on latent scale 281

List of Figures

Figure 3.1 Main types of presentation differences 25
Figure 3.2 Screening process flow diagram 30
Figure 3.3 No. of attributes presented to respondents 39
Figure 3.4 Scatterplot: No. of choice sets vs. attributes* 41
Figure 4.1 Project outline 59
Figure 4.2 Developing the DCE flow chart 63
Figure 4.3 Boxplots of betas and standard errors 74
Figure 4.4 Screenshot: explanation of peripheral neuropathy 78
Figure 4.5 Screenshot: information about attributes Part 1 78
Figure 4.6 Screenshot: information about attributes: Part 2 79
Figure 4.7 Screenshot: types of physical tests Part 1 79
Figure 4.8 Screenshot: types of physical tests Part 2 80
Figure 4.9 Screenshot: types of physical tests Part 3 80
Figure 4.10 Screenshot: types of physical tests Part 4 81
Figure 4.11 Example choice set 81
Figure 4.12 Age of respondents (years) 85
Figure 4.13 Kernel density plot of the conditional mean distribution for Det 2 97
Figure 4.14 Kernel density plot of the conditional mean distribution for CT 4 98
Figure 4.15 Overlaid kernel density plot of distribution of conditional means for Res 2 and Res 3 98
Figure 5.1 Outline of CIPN DCE Project: Study 1 and 2 104
Figure 5.2 Survey flow for arms 1 and 2 in Study 2 106
Figure 5.3 Boxplots of completion time by arm (in minutes) 113
Figure 5.4 Plot of Det 2 density and conditional mean values 124
Figure 5.5 Plot of CT 4 density and conditional mean values 124
Figure 5.6 Plot of Res2 density and conditional mean values 125
Figure 5.7 Plot of Q 4 density and conditional mean values 125
Figure 5.8 Plot of PhyT 3 and 4 density and conditional means 126

Figure 5.9 Scatterplot of utility values 139
Figure 6.1 Scatterplot of health state utilities by country* 168
Figure 6.2 Peru: kernel density plots of utilities 170
Figure 6.3 Denmark: kernel density plots of utilities 170

Acknowledgements

I would like to thank my three supervisors Professor Deborah Street, Professor Rosalie Viney and Professor Stephen Goodall for their guidance and advice throughout my PhD journey. This thesis would not have been possible without them. I am grateful to Debbie for her patience, effort and dedication to make sure I never felt at a loss during my PhD. I have truly been fortunate to have Debbie there to teach and advise me on generator developed designs. She has been a stellar main supervisor, I couldn't ask for more. Thank you to Rosalie for her insight and ability to see the bigger picture of my research, even when I couldn't. Her words are gold, and I cannot express how grateful I am to have her as part of my supervisory panel. Stephen, thank you for your steady advice throughout my PhD and for keeping the supervisor panel meetings from veering too far off topic. I am forever thankful that I had the opportunity to see Stephen present on DCEs at my very first full time workplace. This led me to discover the Centre for Health Economics Research and Evaluation (CHERE, UTS) and opened the door to the wonderful world of DCEs and health research with impact. It has been a dream to have Debbie, Rosalie and Stephen as my supervisors.

I would also like to acknowledge and thank the women with experience of breast cancer from the Breast Cancer Network Australia (BCNA) who volunteered to participate in my qualitative interviews. Also thank you to the INFOCUS Study for assisting in the recruitment of my cancer patient sample for one of my DCEs. Thank you to all the people with experience of cancer who completed my DCE. Special thanks to Associate Professor Susanna Park (University of Sydney) from the INFOCUS Study for her help with recruitment and clinical advice on my DCE. I would also like to thank Senior Lecturer Dr Alison Pearce (University of Sydney), for her advice on conducting qualitative interviews and letting me observe a few she conducted for her own DCE (with the permission of the participant).

There are also several people at CHERE, UTS who have supported me at various points during my PhD that I would like to thank. This includes Associate Professor Brendan Mulhern for all his valuable advice and knowledge with regards to EuroQol and valuation of the EQ-5D-5L. Thank you to Senior Research Fellow Patsy Kenny, for her advice and help with my ethics application. Research Fellows and PhD candidates

Kathleen Manipis and Elena Meshcheriakova, thank you both for your company and empathy especially during the last year of my PhD and COVID-19 lockdowns have been a tough time for everyone, but especially when you are writing a PhD. These two ladies have been there down in the trenches with me. The Zoom Hives have been immensely helpful, if only to share in some laughter and tears while we have been writing our theses.

And last but not least, I would like to thank my partner, Timothy Lou, for his enthusiasm and excitement for me when I decided to start my PhD and for his unwavering support during the journey.

Thesis Abstract

Discrete choice experiments (DCEs) are a popular stated preference technique used in health. A major challenge in health related DCEs is that they often involve terminology and concepts that may be unfamiliar to respondent. Therefore, it is important for DCEs to include explanations of the concepts and terminology used where needed. This raises questions about the different ways information in a DCE is presented to respondents. Most importantly, does the type of information and level of information provided in a DCE impact results?

In order to answer this question, this thesis included a scoping review and three empirical studies that add to the literature. The scoping review summarised DCE studies that investigated the impact of either the amount of information or the presentation style of the DCE or both on the choices of respondents. Studies were included if they allocated respondents to two or more arms to test the impact of differences in the information provided to respondents, either explicitly (e.g., providing more detailed information in one arm than in the other) or implicitly (e.g., by having more options in each choice set). The review showed that the impact of presentation differences on DCE findings was not consistent across studies and this begs the question of whether it is desirable or possible to design DCEs with conclusions that are robust to how the DCE is presented to respondents.

The first empirical study investigated patient preferences for features of an assessment tool for peripheral neuropathy, a possible side effect of cancer treatment. The second study investigated the same topic but this time in a general population sample. The aim was to investigate the differences in preferences between the two samples. Another aim was to understand whether increasing the level of information provided to the general population sample had an impact on results. This was investigated by splitting the general population sample into two arms, where one arm received more information compared to the other. The presentation format was also different between arms, with one arm receiving static images and the other a short video as well as GIFs i.e. moving images on loop. The study found that providing extra information and the use of different presentation formats did not lead to significant preference differences between the two general population arms. The patient and general population were

found to have similar preferences. Although differences were noted in terms of scale. The scale difference was found to be somewhat smaller between the general population arm that received more information and the patient sample compared to that between the general population arm that received only basic information and the patient sample.

The third empirical study investigated different anchoring methods and its impact on the valuation of the EQ-5D-5L, a widely used quality of life measurement instrument. Respondents were shown choice sets that included three options; two health states and a third option which was *immediate death* or *full health* i.e. the anchoring option. The study found that health state utilities were sensitive to the wording used in the third option of the choice sets; in particular, choice sets with *immediate death* as the third option consistently produced a narrower range of utilities than did those with *full health* as the third option.

Evidence from this thesis demonstrates that information in a DCE may impact the resulting preferences, but that the extent of this impact varies by context.

Chapter 1. Introduction

1.1 The case for using stated preference techniques in health services research

Due to the limited nature of most resources, it is not possible for suppliers – such as the government, private industry players or not-for-profit organisations – to produce every possible good or service. Instead, suppliers must determine the type of goods and services most desirable to and valued by a target population. This ensures that resources are allocated efficiently or, rather, allocated to where they are most valued and needed.

One way of determining what types of goods and services are most desirable is through revealed preference data; that is, by observing the behaviour or choices of consumers in the market or by looking at the response to changes in the market price in order to determine the value consumers place on a good or service (Mark & Swait, 2004; Mendelsohn, 2019; Vacca et al., 2019). For instance, travel card data can be used to understand user behaviour on public transport (Tirachini et al., 2016), GPS data can be used to capture route choices by drivers (Vacca et al., 2019) and supermarket scanning data can capture grocery purchase choices of consumers (Brooks & Lusk, 2010; Resano-Ezcaray et al., 2010).

Revealed preferences are ideal for understanding current, or even past, demand for goods and services in a competitive market. However, this only holds under the assumption that the market price is an accurate reflection of how consumers value the goods and services and, additionally, presumes that market information is readily available (Hall et al., 2004; Mendelsohn, 2019). In addition, the literature reports that a lack of variability or high collinearity among revealed preference data variables can make it challenging to estimate model coefficients (Mark & Swait, 2004). There are also instances where revealed preferences may not be available. For instance, in the case of a new product or service, there is no prior information about its market price. The availability of revealed preference data has also been a challenge for some public goods and services, as these goods and services are often not provided in market settings.

For instance, health care in many developed countries is heavily regulated, with public and private insurance, and includes heavy subsidisation from governments through public insurance schemes, such as Medicare in Australia (Services Australia, 2021). As a consequence, patients in these countries very rarely have to pay market price (Lancsar & Louviere, 2008). In addition, there is an issue of asymmetric information, with patients having little to no knowledge of health care products and services available to them and relying on information from health care providers to inform their choices. This means there is also a principal–agent problem, whereby choices patients make may not be due solely to their own preferences but could be influenced by recommendations from health care providers (Lancsar & Louviere, 2008). Thus, decisions made by patients may not reflect their own preferences, particularly as there is evidence health care professionals may have differing preferences to patients (Lancsar & Louviere, 2008; Mühlbacher & Juhnke, 2013). As such, revealed preference data, even if available, may not always be the best method to determine the value consumers place on health goods and services.

As noted earlier, revealed preference data can only provide information about preferences for goods and services currently available in the market. They cannot provide information about potential demand or preferences for new goods and services (Brooks & Lusk, 2010; Lancsar & Louviere, 2008). This is particularly important in areas such as health, where it would be useful to understand whether proposed changes or design of new products, policies or programs meet the demands of patients prior to their implementation.

As a result, stated preference techniques have come to play an important role in the measurement of preferences in health research. Stated preferences refer to choices that consumers say they would make when asked to imagine certain scenarios or in response to questions. This includes responses in surveys (Johnston et al., 2017; Lancsar & Louviere, 2008).

1.2 Type of stated preference techniques used in health services research

Stated preference techniques themselves are not unique to health and have been used in a wide range of different disciplines, including transport (Bansal & Daziano, 2018;

Kroes & Sheldon, 1988; McFadden, 2001), resource and environmental economics (Boxall et al., 1996; Hoyos, 2010; Rakotonarivo et al., 2016), political science (Belle & Cantarelli, 2021; Finkelstein et al., 2017; Youngkong et al., 2010) and marketing (Cantillo et al., 2020; Green & Srinivasan, 1990; Lockshin et al., 2017). Within health research, certain types of stated preference techniques have been more frequently used. These include contingent valuation (CV) studies, standard gamble (SG), time trade off (TTO) studies, discrete choice experiments (DCEs), conjoint analysis (CA), and best-worst scaling (BWS) DCEs.

1.2.1 Contingent valuation (CV) studies

CV studies have been used in the literature to derive the maximum (minimum) a respondent is willing to spend (accept) for a particular gain or loss, known as willingness to pay (WTP) or willingness to accept (WTA) (Bijlenga et al., 2011; Diener et al., 1998; Frew et al., 2003). A CV study uses a survey-based approach in order to elicit WTP/WTA estimates (Bijlenga et al., 2011; Diener et al., 1998). The survey approach includes open-ended questions in which respondents are provided no prompts but asked to provide their maximum WTP or WTA. Alternatively, studies could present respondents with price and ask them if they would be willing to accept or reject it (Diener et al., 1998).

1.2.2. Standard gamble (SG)

While CV studies are certainly useful in situations where monetary outcomes are appropriate, there are cases where it is difficult to ask people to provide a monetary value for willingness to pay or willingness to accept – for example, in relation to the value of life, or health outcomes/gains.

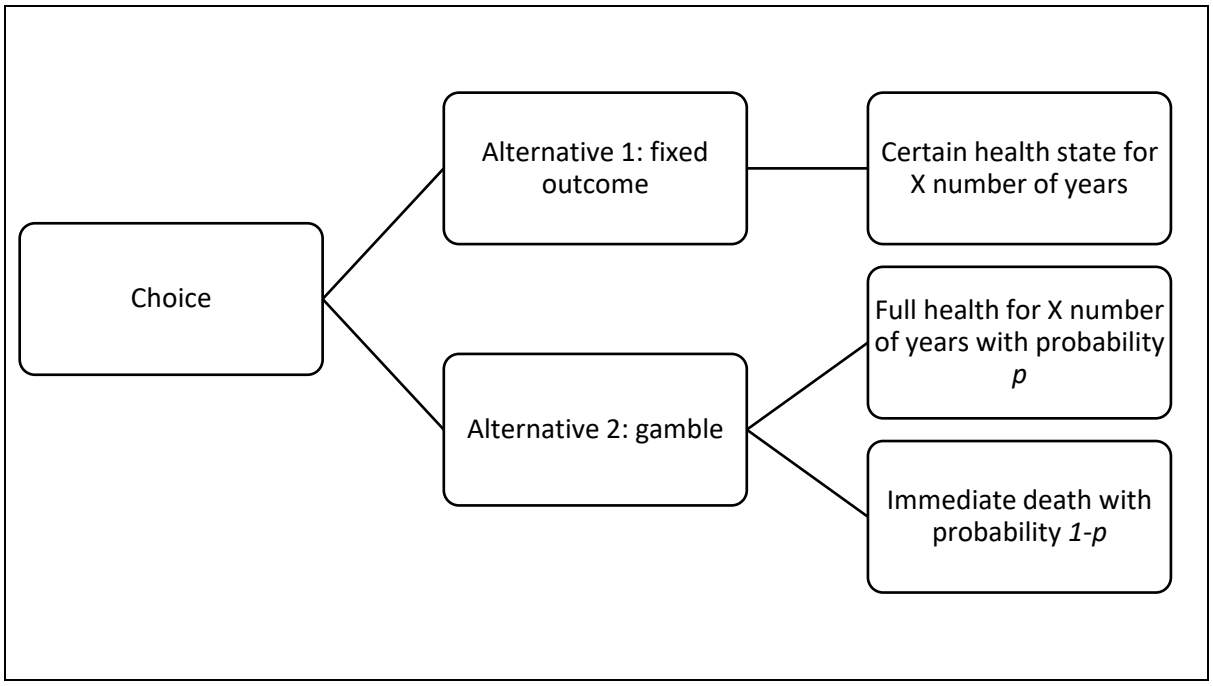
The standard gamble (SG) and the time trade-off (TTO) tasks are survey based methods to elicit valuation of health states on a cardinal scale. These values can then be used to calculate health gains in terms of quality-adjusted life years (QALYs) (Drummond et al., 2015). The QALY is a commonly used metric for health gains that accounts for the trade-off between quantity and quality of life (Drummond et al., 2015; Torrance, 1987).

The SG requires respondents to make trade-offs between different health states. The SG can be completed in a scenario where health states are compared to death. Von

Neumann–Morgenstern utility (Morgenstern & Von Neumann, 1953) forms the basis of SG and posits that respondents choose so as to maximise the expected utility of the possible outcomes under uncertainty, where uncertainty is measured by the probabilities (Wakker & Deneffe, 1996).

Respondents are presented with alternative scenarios. In one scenario, respondents are given a certain outcome (e.g., live with a chronic health condition for a set number of years). In the other scenario, they are faced with a 'gamble' – they have probability *p* of returning to full health but have probability 1-*p* of immediate death (Garza & Wyrwich, 2003; Drummond et al., 2005). Respondents are shown scenarios with different values for *p* until the respondent is indifferent between the alternative scenarios. Under the assumptions of expected utility, the value of *p* at point of indifference represents the utility of the health state. An example of an SG scenario is provided in Figure 1.1.

Figure 1.1 Example of SG scenario: value of p varies until respondent is indifferent between the two alternatives

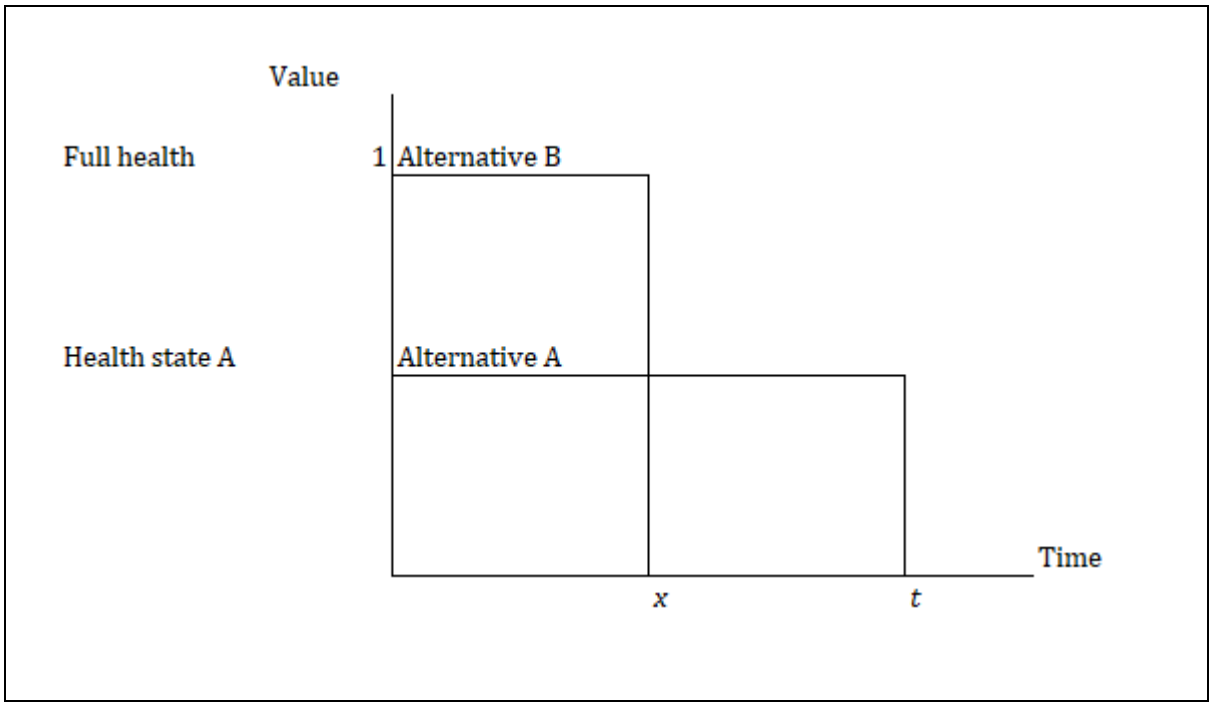


1.2.3 Time trade-off (TTO)

The TTO was developed specifically for use in health (Drummond et al., 2005; Torrance et al., 1972). The TTO is often regarded as having the advantage of being simpler than the SG (Ryan et al., 2001; Torrance, 1987). For the case where health states are considered relative to death, respondents are presented two scenarios in a TTO task. In

scenario A, respondents live with a specified health condition for *t* years followed by death. In alternative B, respondents are in full health for a period *x* (where $x < t$ years), followed by death (Drummond et al., 2005). Respondents are shown a number of scenarios where the value of *x* varies until the respondent is indifferent between the two scenarios. The value of *x* at the point of indifference becomes the utility of the health state. An example of results from a TTO is depcited in Figure 1.2. The TTO method has been popular and variations have been developed specifically for valuing quality-of-life instruments such as the EQ-5D-5L (Andrade et al., 2020; Rencz et al., 2020).

Figure 1.2 Visualisation of results from a TTO task: value of x varies until respondent is indifferent between the two alternatives



1.2.4 Discrete choice experiments (DCEs)

Discrete choice experiments (DCEs) are another stated preference technique that have been used in health research. DCEs have been used across many disciplines, including marketing, transport, resources and environment as well as health (Johnston et al., 2017; Lancsar & Louviere, 2008; Rakotonarivo et al., 2016). DCEs have also been reported to be easier for respondents to understand compared to the TTO (Waudby-Smith et al., 2020).

A DCE usually includes an introductory section in which the context of the DCE and relevant information are provided. Respondents are then directed to a choice scenario/vignette. A choice scenario can be useful as it provides a context for the choices respondents will be asked to make next. After the choice scenario, a choice set or series of choice sets are presented. Each choice set is made up of one or more options. In the case of a single option, respondents are asked to accept or reject the option given to them (e.g. in a DCE, each choice set might be a specific disease screening option with respondents are asked to accept or reject the option). If there are two or more options in a choice set, respondents are asked to choose the option they prefer. Each option is described in terms of a number of attributes, each of which usually takes one of a set number of plausible levels. (Lancsar & Louviere, 2008; Street & Burgess, 2007). An example is shown in Figure 1.3.

Figure 1.3 Example choice set: food delivery preferences

Imagine you don't feel like cooking for dinner and you are starting to get hungry. You are planning to get food delivered.			Choice scenario
Which food delivery service would you prefer?			
	Service A	Service B	Service C
Delivery fee	$4.50	$3	Neither Service A nor B
Standard delivery time	15 minutes	10 minutes	
Surcharge (i.e. peak times/weather)	No surcharge	$2	
Attributes which respondents are asked to consider in this choice set.	Attribute levels in Option A	Attribute levels in Option B	Opt out

In Figure 1.3, respondents are asked to indicate their preferred food delivery service. Services A and B vary in terms of the level of the attributes; delivery fee, standard delivery time and whether or not there is a delivery fee surcharge. In this choice set, respondents also have the option to 'opt out' – that is, to pick neither service A nor B.

The inclusion of an opt-out may be reasonable for some DCEs. However, it should be noted that if a respondent chooses an opt-out, it does not provide information about trade-offs respondents are making or about their preferences for the different levels of the attributes. It does provide information, however, in the sense that respondents did not find either service acceptable.

1.2.5 Related techniques to DCEs: conjoint analysis (CA)

In the literature, the term DCE is often used interchangeably with the term conjoint analysis (CA) and, in some instances, is considered a subset of CA (Reed Johnson et al., 2013). However, in this thesis a distinction is made between the two.

DCEs and conjoint analyses share common features in terms of the overall task format (Louviere et al., 2010). For instance, both may utilise fractional factorial arrays in order to determine the combination of attribute levels to present in choice sets. Both also use statistical methods to elicit part worth utilities. However, the format of the task is different in CA, with respondents are asked to rate or rank options in a choice set as opposed to choosing from a set of options. That said, variations of a DCE can include rankings (see below). DCEs and CAs also derive from different theoretical origins. Both utilise similar design and statistical methods. However, the analysis of CA rarely considers theories behind human decision-making processes (Louviere et al., 2010). This is in contrast to DCEs, which that are underpinned by utility theory.

1.2.6 Related techniques to DCEs: best–worst scaling (BWS) DCEs

It is useful to distinguish between the 'standard' DCE format and best–worst scaling (BWS) DCEs. In this thesis, the term DCE refers to the 'standard' format as described above. Although both the DCE and BWS DCE are both grounded in utility theory, the task format is different (Flynn et al., 2007; Louviere et al., 2015). In a 'standard' DCE format, respondents are asked to choose their preferred option from a choice set. In a BWS DCE, respondents are asked to rank options or attribute levels from 'best' to 'worst' that is, a respondent may be asked to indicate both the 'best' and 'worst' option from the standard DCE format of a choice set (Louviere et al., 2015). Given that the focus of this thesis is on understanding the presentation of information in a DCE, and not the format of the choice task, it was decided that the focus of this thesis would be the standard DCE format.

1.3 DCEs in health services research

The number of DCEs used in health services research, particularly in health economics, has increased over the years. A recent review finds that, on average, 60 DCEs were published per year in the period 2013–2017 (Soekhai et al., 2019) compare to 45 per year between 2009 and 2012 (Clark et al., 2014), suggesting that DCEs have gained popularity as a stated preference technique within health services research (Clark et al., 2014; de Bekker-Grob et al., 2012; Soekhai et al., 2019; Vass, Wright et al., 2018).

DCEs in health service research have been used to investigate a number of topics, including patient views regarding the important features or attributes of assessment, value or willingness to pay and treatment of health conditions (Flood et al., 2017; Marshall et al., 2016; Soekhai et al., 2019; Wright et al., 2017). DCEs have also been used to value health outcome measures (Gu et al., 2013; Mulhern et al., 2019; Norman et al., 2013; Soekhai et al., 2019) and have played a role in understanding the impact, or potential impact, of health programs (Brown et al., 2016; Franco et al., 2016; Salloum et al., 2015; Soekhai et al., 2019), as well as in the design of health delivery programs, systems and policies (Baji et al., 2016; Munger et al., 2017; Whitty et al., 2015). Overall, DCEs have been used to elicit important information about how health services can be designed, implemented and assessed, based on preferences by relevant stakeholders including patients, health providers and decision-makers.

DCEs can provide useful information to inform the design of health care products, programs and systems. However, this is based on the premise that DCE findings provide an accurate reflection of people's preferences. A key factor in ensuring that people's preferences are accurate is certainty that respondents actually understand the topic, concepts and terminology used in the DCE. In the previous section, choice of food delivery service was used to demonstrate what a choice set can look like in a DCE. The concept of a food delivery service is something most people would be familiar with. Hence, it is something that does not require much explanation, except perhaps the context for delivery. Similarly, for topics such as food preferences, design of new public parks, choice of holiday destination etc., it is very likely respondents are familiar with the topic and concepts or terms used. In contrast, this is often not the case for health related DCEs.

In health-related DCEs, the topic, concepts and terminology used may be unfamiliar to respondents. This is particularly the case if a DCE is focused on a particular health condition or health assessment process. The extent to which this is a problem will vary depending on whether patients or health providers who have experience or knowledge of the particular condition or assessment are used as respondents for the DCE. However, there may also be cases where the DCE is used to investigate a new assessment or treatment. In such a case, there may be no knowledgeable or experienced respondents. It may also be the case that knowledgeable or experienced respondents are not available for recruitment. For example, a DCE could be used to investigate two treatment options for cancer: Option A, the standard treatment, and option B, a new treatment not yet on the market that provides extra life expectancy but has more side effects. Respondents would need to understand what the potential new side effects are as well as their impact on their everyday lives in order to make a fully informed decision between these two treatments.

In addition, the purpose of the DCE may mean that a broader range of respondents, such as the general population, may be more appropriate. Indeed, Soekhai et al. (2019) note that a variety of respondents are used in health DCEs including patients, the general population, health care workers and students.

It is particularly crucial for health-related DCEs to provide information, either prior to presenting or within choice sets, explaining the concepts and terminology used to ensure respondents understand the context, attributes and/or levels used to describe the options presented in choice sets. Furthermore, researchers should consider how this information is presented in a DCE in order to maximise respondent understanding. This leads to an important question: are there ways of presenting DCEs so that the information presented is more easily understood by respondents?

1.4 Aims of thesis

In order to answer the above question, this thesis sought to investigate:

1. What are the main ways that information can be presented differently to respondents in a DCE?

2. Does changing how the DCE is presented to respondents lead to differences in results?

Research question 2 is particularly important as, for many health researchers, it is desirable for DCE results to be an accurate, reliable and generalisable reflection of respondent preferences and not a function of the amount or type of information provided or how the DCE was presented to respondents. These aims will be investigated through a scoping review and three empirical studies.

1.5 Overview of chapters

The current chapter has covered the context and the aims of this thesis. Chapter 2 provides an overview of the theoretical framework underpinning DCEs and introduces important terminology and concepts related to DCEs used in this thesis. The process of designing and analysing results from a DCE is also briefly reviewed. The third chapter covers the scoping review and formally introduces the framework used to examine the different ways a DCE can be presented to respondents. The empirical chapters cover two projects in total. Project 1 investigated the preferences of two different population types for the design of a health assessment tool. Chapter 4 covers the design and construction of the DCE as well as the patient population sample preferences. Chapter 5 details the general population sample preferences and compares the results from the two samples. Chapter 6 details Project 2, which investigated the use of two different anchoring methods of the choice set used in the literature and their impact on respondent preferences. This was in the context of eliciting QALY weights for the EQ-5D-5L using a Peruvian and Danish population sample. Implications of findings from the scoping review in Chapter 3 and the empirical chapters are discussed in Chapter 7.

Chapter 2. Methods

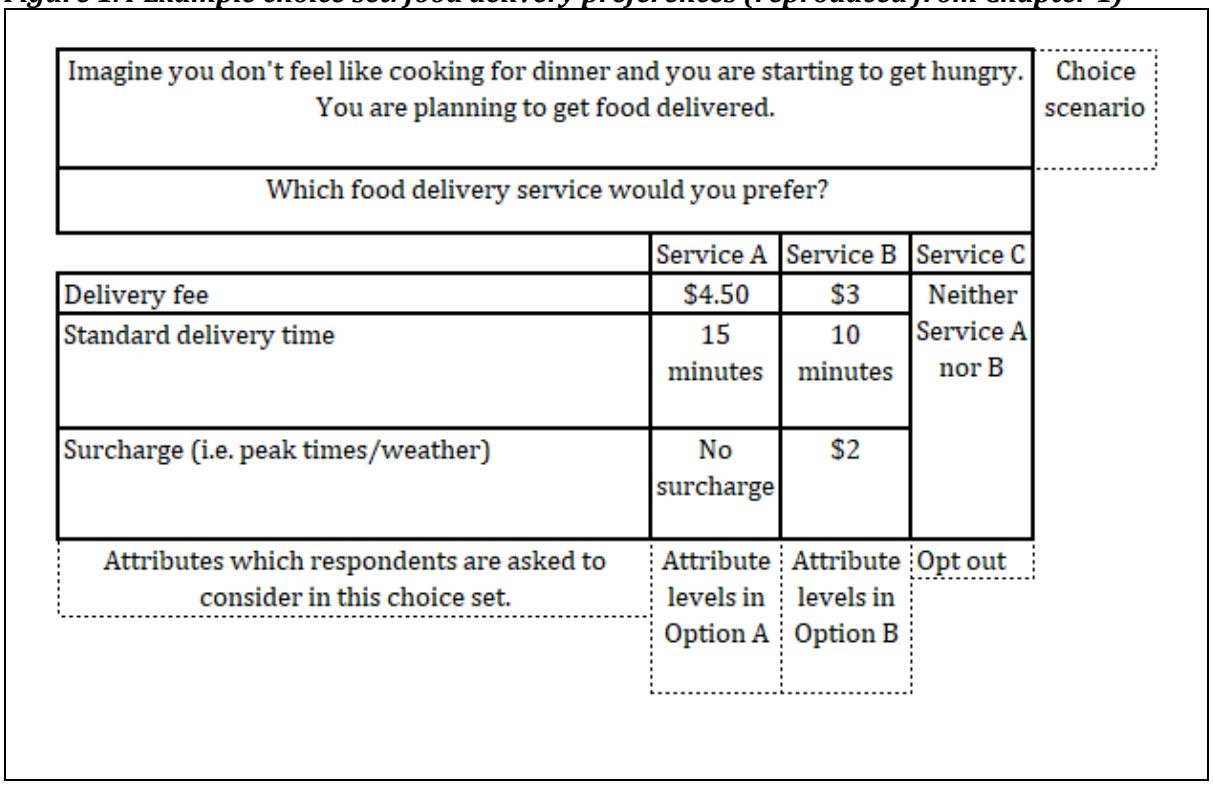
The DCE is the key method used in this thesis. In this chapter, key concepts and terminology related to DCEs are defined and the process of creating a DCE, including the construction of choice sets and development of attributes and levels, is explained. The different types of estimation models used to analyse the DCE data are also outlined.

2.1 Definition of a DCE

As discussed in Chapter 1, a DCE is a stated preference technique. A review of the use of DCEs in the health literature by Soekhai et al. (2019) finds they have been used to answer many questions. These includes understanding patient consumer experience, trade-offs between health outcomes and patient/consumer experience, and preferences of health professionals regarding treatment or screening. They have also been used to understand factors that contribute to job choice as well as factors contributing to priority-setting frameworks.

In a DCE, respondents are shown a series of hypothetical questions, known as choice sets (Street & Burgess, 2007). In each choice set, there one or more options for respondents to consider. In the case of a choice set with only one option, respondents are asked whether they would accept or reject the option. When there are two or more options, respondents are asked to choose their preferred option. Regardless of the number of options, each option is described by a number of features (attributes) that respondents must consider. Each attribute can appear at one of a fixed number of levels. The choice(s) respondents make provides information about the trade-offs they have made and their preferences (Lancaster, 1966; Lancsar & Louviere, 2008). The example used to describe a choice set in a DCE from Chapter 1 has been reproduced below as a brief refresher (Figure 1.4). This example choice set is in the context of food delivery options. Service A requires a longer wait time but the cost of delivery overall is less than Service B. Respondents may choose Service C if neither Service A nor Service B is satisfactory.

Figure 1.4 Example choice set: food delivery preferences (reproduced from Chapter 1)



2.2 How to create a choice set

In this section, the essential components required for the design and construction of choice sets are briefly reviewed. This section is intended to be a summary only, with references provided for readers seeking more detail.

2.2.1 Developing attributes and levels

Establishing which attributes to include, and what levels they can take, is essential to creating a DCE. In particular, attributes and levels should reflect important features to consider for the topic of interest and type of questions the DCE is being used to answer. For instance, if the aim is to calculate willingness to pay, the inclusion of an appropriate cost attribute is required. The exclusion of important attributes/levels can lead to omitted variable bias in DCE results (Clark et al., 2014). For instance, a DCE may be used to investigate different cancer treatment options. Suppose that risk of side effects is an important attribute people consider when choosing treatment options. However, in the DCE, the risk of side effects is not included as one of the attributes. Consequently, results from this DCE do not tell the whole story about what patients consider when choosing a treatment option.

Researchers have used various methods to develop attributes and levels of a DCE (Helter & Boehler, 2016). It is also not uncommon for researchers to utilise more than one method to develop attributes and levels. Qualitative methods have been growing in popularity as a means to identify important attributes and levels (Hollin et al., 2020; Soekhai et al., 2019). This has included the use of focus groups and interviews (Bridges et al., 2011; Coast et al., 2012). Consultation with relevant stakeholders has also been used (Klaiman et al., 2016; Zhao et al., 2013). It is also important to note that qualitative methods are not mutually exclusive, researchers can utilise more than one qualitative method in their DCE.

Other methods include referring to previous literature or conducting a literature review to investigate which attributes and levels have been included previously for the topic of interest (Knox et al., 2013; Nickel et al., 2018; Spinks & Mortimer, 2015). Researchers may purposely choose attributes and levels that reflect what is currently seen in the market (Arrua et al., 2017; De-Magistris et al., 2013; Oppewal et al., 2015). Pilot studies have also been used, including studies in which an initial draft form of the survey is completed by a smaller set of respondents with feedback and responses then used to finalise attributes and levels (Cole et al., 2018; Meyerhoff & Glenk, 2015; Scheufele & Bennett, 2013). Researchers have also used BWS DCEs in order to select the attributes to be used in a subsequent DCE (Webb et al., 2021).

2.2.2 Construction method of choice sets

Apart from the design of the survey, an important area of research has been in choosing the combinations of attribute levels that respondents see in each option within a choice set (i.e. components of x_{ij} (Street & Viney, 2019), explained further on page 15), referred to in this thesis as the construction method of the choice sets. In particular, level combinations in choice sets should be constructed so as to allow for parameters associated with levels of attributes to be independently estimated (Domínguez-Torreiro, 2014; Street & Burgess, 2007). Choice sets can be constructed to detect main effects only or can include two-way or higher-level interactions (Street & Burgess, 2007).

There are various methods by which choice sets can be constructed (see Street and Viney (2019) for further details). In this thesis, DCEs in the empirical chapters have

been constructed using a generator-developed approach (Street & Burgess, 2007). This method involves adding generators to an initial set of level combinations in order to create options in the choice sets (Street & Burgess, 2007). These were created in Mathematica (Wolfram Research, 2016) based on code initially written by Burgess (2007) and subsequently augmented by Street (2019). A simple example would be for a DCE that only has three attributes with two levels for each. Suppose that the choice sets each require two options. The first option might contain level 1 of the first attribute, level 2 of the second attribute and level 1 of the third attribute, represented by (0, 1, 0). In order to create the second option, a set of generators (1,1,1) is added modulo 2 to create (1, 0, 1) i.e. the second option of this choice set includes level 2 of the first attribute, level 1 of the second attribute and level 2 of the third attribute. Further information about generator-developed designs can be found in Street and Burgess (2007).

2.3 DCEs and utility theory

DCEs are underpinned by consumer theory (Lancaster, 1966; McFadden, 2001). Consumer theory assumes that respondents choose the bundle of goods and services that maximises their utility, subject to the resource constraints they face. In a DCE it is assumed that, for any given choice set, respondents will choose the option that gives them the most satisfaction or utility (McFadden, 2001). This is known as utility-maximising behaviour and is a key assumption behind the modelling of discrete choice data such as DCEs (Train, 2009).

In a DCE, utility is assumed to derive from the attribute levels in each option in a choice set (Lancaster, 1966; McFadden, 2001). In the context of choosing a food delivery service, it is assumed that the utility a respondent derives from each food delivery service option will depend on observed attributes, such as the delivery fee, time for delivery and any potential fee surcharges for peak delivery times. Respondents are assumed to choose the option that gives them the highest level of utility.

Formally, the respondent, *i*, faces a choice among *J* options within a choice set. The respondent will derive utility, *U*, from each option. As a utility maximiser, when faced with option *j* versus option *k*, the respondent will only choose alternative *j*, if $U_{ij} > U_{ik} \ \forall \ j \ \neq k$ (Train, 2009).

The researcher does not directly observe the utility of the respondent. As expressed by Train (2009), researchers can only observe some attributes of a particular option. For instance, for a respondent *i* facing option *j*, the observed attributes of option *j* as seen by respondent *i* are represented by x_{ij}, and some observable characteristics of respondent *i* are represented by c_i. These are the input into the function $V_{ij} = V(x_{ij}, c_i)$, which forms the systematic component of utility. There may also be factors that influence the utility of respondent *i* which are either unobserved or unobservable – for example, drivers of utility that cannot be observed, error in choices made by respondents, random respondent behaviour, errors in observation by the researcher etc. As a result, the utility for respondent *i* facing option *j* is represented by:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \qquad i = 1, \dots, I;\ j = 1, \dots, J, \tag{1}$$

where ε_{ij} represents the unobserved and/or unobservable component of the utility for respondent *i* and option *j*. As ε_{ij} is unknown to the researcher, it can be treated as random (Train, 2009). The probability that respondent *i* chooses option *j* over option *k* (Train, 2009) is denoted by P_{ij}:

$$P_{ij} = P(U_{ij} > U_{ik}, \forall\, k \neq j)$$

$$P_{ij} = P(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}, \forall\, k \neq j)$$

$$P_{ij} = P(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}, \forall\, k \neq j),$$
$$i = 1, \dots, I;\ j = 1, \dots, J; k = 1, \dots, J. \tag{2}$$

2.4 Analysis of DCE data

The data from the DCE are in the form of a discrete choice represented by a binary dependent variable, which is 1 if the option is selected in the choice set and 0 if the option is not selected. In this section, different techniques to model the discrete choices are described. This section begins with the simplest model, the multinomial logit model (MNL), which assumes that all respondents have the same preferences for each of the attribute levels; that is, the respondents have homogeneous preferences. Subsequent models relax this assumption by introducing different ways to model any heterogeneity that may be present in respondent preferences by using alternative assumptions about the error terms.

2.4.1 Multinomial logit (MNL) model

In the MNL model, for a respondent, *i,* who makes a choice among *J* alternatives in each choice set, each unobservable component of utility, ε_i , is assumed to be an independently, identically distributed (iid) extreme value type 1 (Train, 2009). In other words, the unobservable utility for one alternative is assumed to be uncorrelated with the unobservable utility of any other alternative. This gives the logit model independence from the irrelevant alternatives (IIA) property (see Section 3.3 of Train (2009) for a discussion).

In the MNL model, the observed portion of utility is treated as linear in parameters, where $V_i = X_i^T \beta$ and X_i^T represents the vector of observed variables (e.g. levels of the attributes) for any given alternative. The observed variables include the dummy coded attribute levels. Unless otherwise specified in this thesis, the dummy code uses the first level of each attribute as the omitted level or base case. This allows interpretation of the MNL model parameters as a comparison of each attribute parameter with the corresponding base case – for example, if the parameter was positive and significant ($p < 0.05$), this could be interpreted as the parameter being significantly preferred to the base case.

As mentioned above, the MNL model assumes that respondents have homogeneous preferences – that is, the vector of β is constant across respondents. Formally, in the MNL model, for a respondent *i,* the random utility for alternative *j* from *J* alternatives in a given choice set can then be expressed as:

$$U_{ij} = X_{ij}^T \beta + \varepsilon_{ij}, \quad i = 1, \dots, I; \; j = 1, \dots, J. \tag{3}$$

Predicted choice probabilities

A key advantage of the MNL model and the IIA property is that it allows for a closed form expression for the probability of a particular alternative being chosen. The derivation of this results can be found in Chapter 3.10 in Train (2009). Briefly, the logit probability of choosing alternative *j* among *J* alternatives for respondent *i* can be expressed as:

$$P_{ij} = \frac{e^{X_{ij}^T \beta}}{\sum_{k=1}^{J} e^{X_{ik}^T \beta}},$$
$$\forall\ i = 1, \dots, I;\ j = 1, \dots, J. \quad (4)$$

2.4.2 Latent class analysis (LCA)

The latent class analysis (LCA) relaxes the assumption of homogeneous preferences in the MNL model by introducing a discrete number of classes. In particular, each class has its own fixed parameter vector (Sarrias & Daziano, 2017; Shen, 2009). This model relies on the assumption that the chosen number of classes is representative of the preference heterogeneity present in the data (Shen, 2009). For analysis, it is assumed that the discrete mixing distribution follows the semi-parametric MNL format (Greene & Hensher, 2003; Sarrias & Daziano, 2017; Shen, 2009). The prior probability that a respondent, *i,* from class *s* chooses alternative *j* among *J* alternatives (Shen, 2009) can be expressed as:

$$P_{ij|s} = \frac{e^{X_{ij}^T \beta_s}}{\sum_{k=1}^{J} e^{X_{ik}^T \beta_s}},$$
$$\forall\ i = 1, \dots, I;\ j = 1, \dots, J;\ s = 1, \dots, S, \quad (5)$$

where β_s is the parameter vector associated with the observed variables X_{ij}^T in class *s* (Shen, 2009).

2.4.3 Mixed logit (MXL) model

The MXL model is similar to the LCA as it accommodates preference heterogeneity. However, unlike the LCA, which has a discrete mixing distribution, the MXL model allows for different taste preferences for each respondent (Sarrias & Daziano, 2017). For the observed component of utility, each respondent *i* is assumed to have their own coefficient, represented by β_i. It is assumed that β_i varies over respondents according to some density $f(\beta)$. The objective of the analysis is to estimates the parameters of the distribution of β_i. In this thesis, the density is assumed to be multivariate normal. The unobservable component of utility for alternative *k* is represented by ε_{ik} and is assumed to be iid extreme value (Sarrias & Daziano, 2017; Train, 2009).

Formally, in the MXL model the random utility for alternative *j* in a given choice set for a respondent *i* can then be expressed as:

$$U_{ij} = X_{ij}^{T}\beta_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \; j = 1, \dots, J, \tag{6}$$

where

$$\beta_i \sim MVN\ (\beta, \Sigma).$$

Interpretation of MXL model results: kernel density plots

For parameters with significant standard deviations ($p < 0.05$), results can be examined visually through kernel density plots. Kernel density plots can be thought of as smoothed histograms. The horizontal axis represents the range of values a parameter/s can take. The parts of the curve of the kernel density plot above 0 represent respondents who hold a preference for a parameter relative to the base case, while the parts of the curve that lie below 0 represent respondents who preferred the base case relative to the parameter being plotted. The height of the curve measured by the vertical axis represents how frequent or popular particular values were. If the density plot is very narrow and features one peak, this can be interpreted as indicating that, while heterogeneity may be present, there is a pattern of similar preferences among respondents. In contrast, if the density plot features a wide curve with more than one peak, it may indicate that respondents were more disparate in their preferences.

2.4.5 Scaled MNL (S-MNL) model

The scaled MNL (S-MNL) model can be used to detect whether the estimated vector of β coefficients, or parameters in the model, from two different populations are comparable. The estimated β parameters can only be compared directly between two samples if the scale parameter is not significantly different between them. Vass, Wright et al. (2018) discuss in detail the issue of scale in DCE data and the potential for confounding scale differences with estimated β parameter differences between samples.

When analysing DCE data, the utility difference is the focus, as utility itself has no defined unit (Train, 2009; Vass, Wright et al., 2018). As a consequence, multiplying all the estimated parameters by some constant does not change the relative size of the utility. It is therefore necessary to scale the utility. The usual way to do this is to normalise the variance of the error terms (Vass, Wright et al., 2018). The variance of the error terms may be different for different populations. Therefore, parameter estimates between two populations/sources can only be directly compared if the scale is the same.

The S-MNL model can detect the presence of scale differences through the inclusion of a scale factor (σ) in the utility function. Formally, the random utility for alternative *j* in a given choice set for a respondent *i* can then be expressed as:

$$U_{ij} = (\beta\sigma_i)X_{ij}^T + \varepsilon_{ij}, \quad i = 1, \dots, I; \; j = 1, \dots, J, \tag{7}$$

where

$$\sigma_i = e^{\bar{\sigma} + \tau\nu_i}, \nu_i \sim N(0,1).$$

The S-MNL model expresses a particular form of heterogeneity wherein β coefficients are scaled proportionally by σ_i (Fiebig et al., 2010; Sarrias & Daziano, 2017). Individual-specific differences, such as assignment to arm or sociodemographic variables, can also be accounted for in the mean of the scale (Sarrias & Daziano, 2017). In such a case, σ_i is modified to be:

$$\sigma_i = e^{\bar{\sigma} + \delta s_i + \tau\nu_i},$$

where s_i is the vector of individual specific characteristics for respondent *i.*

2.5 Summary

In this chapter, key concepts relating to the theory underpinning DCEs as well as the methods relating to the creation of a DCE and analysis of DCE data have been explained. The next chapter covers findings from the scoping review, which aimed to summarise the literature investigating alternative ways of presenting a DCE to respondents and the impact of presentation method on the results of the DCE.

Chapter 3. Different ways to present a DCE to respondents: A scoping review

3.1 Overview

In this chapter I describe a scoping review I conducted with the aim of summarising the existing literature on different ways a DCE can be presented to respondents and the impact of presentation on results. The review encompassed DCEs across all contexts and disciplines rather than health alone, as DCEs have been used across a broad range of disciplines and the way a DCE is presented is likely an area of interest to DCE researchers regardless of discipline. There is also the potential that different disciplines may utilise certain ways of presenting a DCE when compared to other disciplines. Thus, by broadening the scoping review beyond the health discipline valuable learnings from other disciplines could be considered. The objective of this scoping review was to draw learnings of ways to present a DCE, and to summarise these findings to assist DCE researchers in health as well as other disciplines.

This chapter begins with a summary of previous findings to provide a background to the current scoping review. A framework for investigating DCE presentation differences is established before the discussion moves on to details of the review search process. The results section is divided into two parts. The first part summarises the characteristics of the included studies, consistent with methods used in previous reviews of DCE literature. The second focuses on the details of DCE presentation differences investigated in the included studies.

3.2 Summary of previous reviews

A growing number of studies across disciplines have sought to investigate the impact of presentation differences in a DCE on the resulting findings. Presentation of a DCE refers to what respondents see in the DCE, including all the information presented both prior to and within the choice set/s. One previous review investigated the use of the 'risk' attribute in health care-related DCEs (Harrison et al., 2014). Within the health economics literature, risk – described in terms of the probability or likelihood of the

occurrence of an outcome – is an important component when it comes to decisions about health care (Harrison et al., 2014). It can be challenging to ensure that respondents understand the risk information presented in DCEs. The review included 117 papers and discusses how the risk attribute was communicated and presented in these papers, including whether the studies provided recommendations about which method would be the most effective form of risk communication.

While the systematic review by Harrison et al. (2014) provides useful information about the use of the risk attribute in the health care literature, its focus was on presentation of the risk attribute and did not cover presentation more generally. The authors also raised concerns about the lack of introductory or background information in DCE studies to educate respondents about the attributes and their levels prior to the presentation of the choice sets. This is of particular concern, as respondents are assumed to be making informed choices with a full understanding of the information provided. In other words, it is assumed that respondents fully understand the risk information provided to them in the choice sets. If this is not the case, it could lead to more variation in the data than is warranted and may hinder the ability of researchers to draw out accurate preference information from the data (Rakotonarivo et al., 2016).

In addition, how risk information is presented, and its impact on choice preferences, was not something that was considered in the majority of studies included in the review (Harrison et al., 2014). To explore how they information is presented in a DCE impacts on choice preferences requires some form of comparison built into the design of the DCE. Rakotonarivo et al. (2016) do this by comparing the consistency of findings across time and when respondents are split into two or more groups with each group presented with a slightly different DCE (Freeman, 2003; Rakotonarivo et al., 2016). In other words, their review explored findings from DCEs that used a within-subjects design (i.e. using the same individuals) or a between-subjects design (i.e. using different groups of individuals).

The authors identify several methods to compare DCE findings. These methods can be separated into two main approaches. The first involves investigating the temporal stability of results through a test-retest approach by repeating the same survey at different points in time using the same respondents (Liebe et al., 2012; Rakotonarivo et al., 2016; Schaafsma et al., 2014). The second approach involves the use of two slightly

differently presented DCEs, either simultaneously or one after the other. These differently presented DCEs can take one of four forms: showing more information to respondents (Rakotonarivo et al., 2016; Robinson et al., 2008); changes to the information presented in the choice scenario or vignette (Carlsson et al., 2010; Rakotonarivo et al., 2016); changes to the information presented in the attributes and levels of the choice sets (Bateman et al., 2009; Rakotonarivo et al., 2016); and use of different choice set structure characteristics – for example, different numbers of options, attributes and/or levels (Rakotonarivo et al., 2016; Rolfe & Bennett, 2009).

The second approach identified by Rakotonarivo et al. (2016) – presenting the DCE differently simultaneously in a between-subjects design – is of particular relevance to the aims of the current thesis. The comparison of respondent results from DCEs that are presented slightly differently contributes to understanding the impact of DCE presentation differences on the conclusions.

In total, Rakotonarivo et al. (2016) reviewed 50 studies, with 87 reliability outcome measures in total. Only 45% (39/87) of the outcome measures found a significant difference in results when respondents saw slightly differently presented DCE or when they were asked to complete the same DCE at two different points in time. For the majority (55% (48/87)) of outcome measures there was no significant difference.

It should be kept in mind that the reliability outcomes used in Rakotonarivo et al. (2016) included tests of intertemporal stability of results as well as impacts of presenting slightly different DCEs across groups of respondents. That said, the majority of the outcomes (72%, or 62/87) were from between-subject designs. Hence, the results do shed some light on the impact of presentation differences on DCE findings. It should also be noted that this review was restricted to the environmental economics literature.

3.2.1 Aims of scoping review

As far as this author is aware, there has been no review dedicated to investigating the the DCE literature as a whole in relation to the influence of differences in DCE presentation on results. Hence, the aim of this chapter is to conduct a scoping review (Munn et al., 2018) in order to identify and summarise existing literature that compares DCE presentation differences using between-subject designs.

In the following section, the framework used to explore presentation differences in a DCE is established. This is then followed by the methods section of the scoping review.

3.3 Framework used to investigate presentation differences in a DCE

Researchers may have many different purposes and aims when investigating presentation differences and their usage in DCEs. A distinction must be made between the *purpose* and the *method* or *type* of presentation difference used in a DCE and its *impact* on results, or rather whether it makes a difference to outcomes measured. The focus of this section, in particular, is on the *method* or *type* of presentation difference. In order to establish a framework and the terminology used to discuss presentation differences in DCEs within this thesis, this section begins by defining what 'presentation' refers to in a DCE. The term 'presentation differences' is defined and the main types of presentation differences are then described.

3.3.1 What is considered part of the presentation of the DCE?

In this thesis, a distinction is made between what is presented to the respondent and what is used to inform the design and construction of the DCE. Presentation in a DCE refers to the content that is presented to respondents within the introductory section, vignette, choice question and various aspects of the choice sets, including the options, attributes and levels. This is considered separate to other design aspects of the DCE including choice set construction design and quantitative/qualitative methods used to develop and refine the attributes, levels or DCE more generally.

3.3.2 Between-subjects design

In stated preference studies, researchers may choose to use a between-subjects design, where respondents are split into two or more groups and the results from these different groups are compared (Hampton, 2018). For instance, in clinical trials participants may be split into two or more groups. These groups are typically treated identically with the exception of the type of treatment they receive (e.g. control and two different active treatments). At the end of the trial, outcomes for these groups are then compared to understand the effectiveness of the treatment of interest. In this thesis, the groups into which respondents are split will be referred to as the arms of the study –

that is, each arm in the study will be a shown a slightly different DCE where the presentation of the DCE differs in some manner.

The focus is on between subjects designs in this thesis as opposed to within subjects designs. Between subject designs are particularly relevant to this scoping review as it allows results from arms to be compared in parallel. This also means that respondents can be randomly assigned to arms to complete DCEs, allowing for equivalent, or near equivalent demographics across arms. Within subjects designs would require the same respondent to complete different DCEs and the impact on results may be confounded by the effect of completing the DCEs at different points in time. In addition, there is also the issue of respondents dropping out and not completing the second or third DCE in within subject designs.

3.3.3 Presentation difference in a DCE

In a DCE, a set of respondents may be split into two or more arms. A presentation difference is present when the presentation of the DCE to respondents is dependent on arm assignment. Presentation differences between arms can include anything that respondents see prior to the choice sets (e.g. the introduction section or choice scenario) as well as differences within choice sets. There may be one or more types of presentation difference within a DCE.

An example of this is found in Knox et al. (2013), although their study uses the term 'information condition' to describe a particular type of presentation difference. Knox et al. (2013) describes how the introductory text was framed positively, negatively or neutrally (the control group) depending on the arm to which respondents were assigned. This thesis encapsulates this type of presentation difference but includes not only framing of text but also any instance where what is presented to respondents in a DCE depends on the arm to which they are assigned.

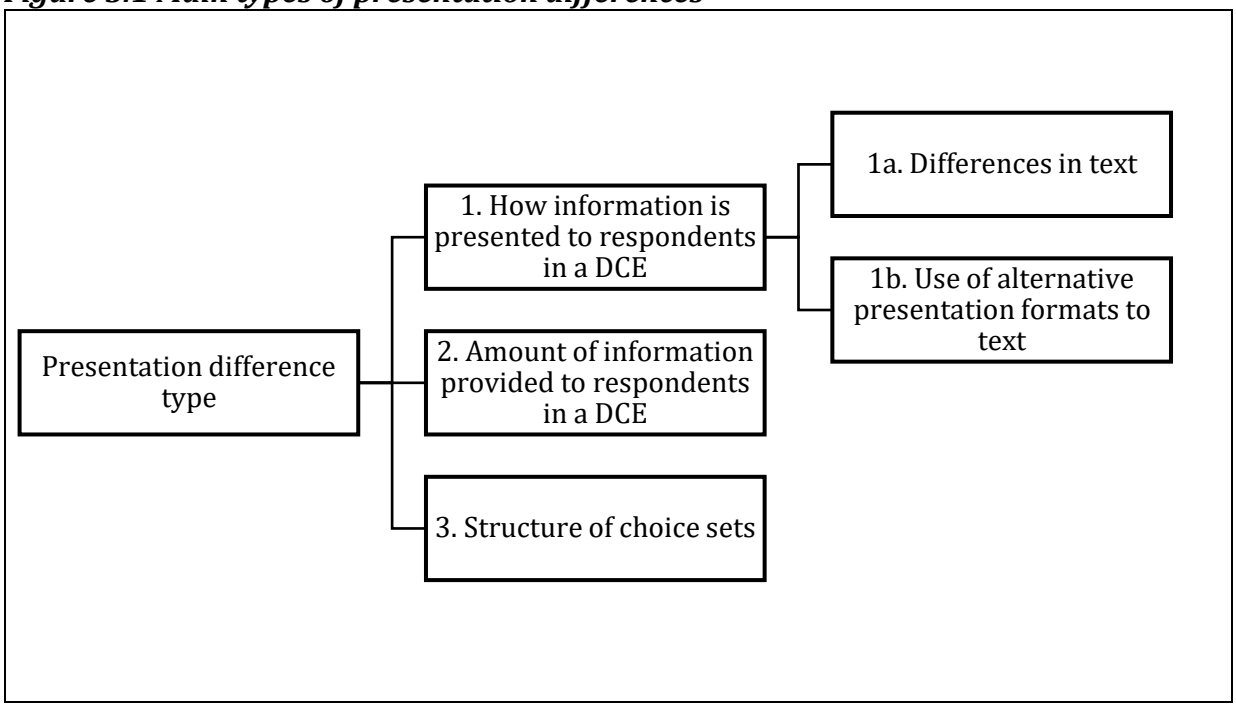
3.3.4 Main types of presentation differences used in DCEs

There are a number of ways a DCE can be presented differently between arms. The categorisation of ways a DCE can be presented used in this thesis is based on the framework used by Rakotonarivo et al. (2016), with differences in terms of major category types. In this thesis, there are three main types of presentation differences, and

presentation differences between arms is broadened to include anything that respondents see prior to the choice sets (e.g. the introduction section or choice scenario), as well as within choice sets. As one of the aims of thesis was to understand the impact of presentation differences on DCE results, the focus is on between-subjects designs. It should be noted that Rakotonarivo et al. (2016) also discuss within-subjects designs as the purpose of their review included evaluating intertemporal stability of results.

Each main type of presentation difference is described below. Researchers may use only one type of presentation difference, or they may use two or even more types of presentation difference in a DCE. Figure 3.1 outlines the main types of presentation differences.

Figure 3.1 Main types of presentation differences



1. How information is presented to respondents in a DCE

Researchers may intentionally choose to use certain words or way of phrasing the text. Researchers may also choose to use alternative formats or text to present information in a DCE. Differences in textual presentation and presentation format each count as a separate presentation difference when used in a DCE.

1a. How information is presented to respondents in a DCE: differences in text

Researchers may intentionally phrase text differently between arms in a DCE. There are many ways researchers can choose to describe a decision or choice problem, akin to using alternative perspectives to describe a visual scene (Tversky & Kahneman, 1981).

One way text can be presented differently is in terms of framing. In particular, information in a DCE can be framed in terms of a gain or loss (Gong et al., 2013; Kahneman & Tversky, 1979; Ruggeri et al., 2020). This is referred to as positive (gain) or negative (loss) framing in some studies. For instance, in one arm an attribute and its levels may be framed in terms of *lives saved* (gain/positive framing) and in another arm framed in terms of *lives lost* (loss/negative framing).

The text information presented (often a numerical value or statistic) in a DCE may be treated by respondents as an anchor or reference point when they are making choices (Ni et al., 2019; Tversky & Kahneman, 1974). Respondents may be more likely to make choices biased towards the anchor presented to them than the choices they would have made had this information or anchor had not been presented. This phenomenon is commonly referred to as the anchoring effect (Furnham & Boo, 2011; Ni et al., 2019; Tversky & Kahneman, 1974).

Researchers may also choose other ways to present text information that may not necessarily induce a framing or anchoring effect. For instance, text information may be presented slightly differently in each arm, meaning the choice scenario is different in each arm. Researchers may choose to change the choice scenario for consideration (e.g. hot or cold weather, different type of health issue, different locations). Similarly, an outcome or concept for consideration in a choice set may be described by similar but slightly different representations (e.g. a concept may be described using different terminology).

1b. How information is presented to respondents in a DCE: use of alternative presentation formats to text

Researchers can choose to present information in a DCE in a format other than text. This includes the use of numerical or statistical values, graphs, pictures, visual aids, moving images (e.g. GIFs, video, simulated environments) etc. This extends to factors such as the display format of choice sets, options and attributes.

It is possible for researchers to use both of these presentation differences in a DCE. For instance, researchers may choose to frame an attribute in terms of *lives saved* or *lives lost* depending on arm assignment. In addition, one arm may see the introductory information as text, and the other arm may see this information in the form of a video.

2. Amount of information provided

In a DCE, the amount of information provided can be different between arms. For instance, in one arm, respondents might only receive basic information about the DCE topic while in the other arm respondents receive the basic information plus extra information that may explain one or more attributes further. A distinction is to be made, however, between the amount of information provided and the number of options, attributes or levels presented to respondents in each arm. The options, attributes or levels relate to the structure of choice sets and are identified as a distinct presentation difference type.

3. Structure of choice sets

Researchers may systematically manipulate the number of options, attributes or levels presented to respondents in each arm. The presentation of different numbers of options, attributes or levels to respondents in an arm, influences the amount and presentation of information for consideration by respondents. This is consistent with how 'information' is defined in this thesis.

The number of choice sets is considered separately from the structure of choice sets. Increasing the number of choice sets shown to respondents is not, in effect, increasing the amount of information given to respondents, nor is it changing how information is presented. Rather, it is merely increasing the number of times respondents make choices. Previous studies that have explored the number of choice sets to use in a study have adopted a statistical efficiency perspective (Burgess et al., 2011) and trade-off with respondent fatigue (Hess et al., 2012). Hence, it was decided that number of choice sets is out of scope as a type of presentation difference for the purposes of this thesis.

3.4 Methods

In this section, the inclusion and exclusion criteria for studies are defined. The process to find relevant studies is then described. The screening process was conducted by the author of this thesis, with consultation and advice from the author's three supervisors.

3.4.1 Literature search strategy

Inclusion criteria

The inclusion criteria for the review were as follows:

- The focus of the study must be a DCE.
- The choice set/s must include at least a single attribute with at least two levels.
- The study must have at least two arms where the presentation of the DCE is different in some way between them; assignment to arm could be random or by some other method.
- The study must apear in a peer-reviewed journal article OR a relevant systematic review OR a relevant dissertation.
- The study must be in English.

Exclusion criteria

Studies not considered for this literature review included the following:

- Studies using a different form of stated preference experiment than those described in chapter 1 (e.g. best–worst scaling DCEs and similar, other types of stated preference techniques such as conjoint analysis and contingent valuation methods).
- Studies which present a single arm DCE and compare these results to another stated preference technique or revealed preferences. (this falls out of the scope of this review)
- Pilot studies where the main study has been included in the review and does not add any unique information.

Search process

The literature search was intended to be broad to capture as many studies as possible, so no restriction was placed on the discipline area for included studies. Both iterative and 'trial and error' approaches were taken to find relevant search terms relating to presentation differences. From this process, it was clear that a number of different terms were relevant to the search criteria and all were subsequently used. The search terms used are listed below in Table 3.1. Several terms for DCEs were also used, with some more popular than others over time. It should be clarified that studies using conjoint analysis were not included in this review; rather, the term was used to identify

studies that used DCEs, as defined in Chapter 2. The term 'conjoint analysis' was included as this term has been used interchangeably with DCEs in the past.

Table 3.1 Search terms

'DCE' Search Terms	'Presentation difference' Search Terms
discrete choice experiment discrete choice model stated choice experiment stated preference survey conjoint analysis	framing information AND condition information AND effect* information AND context format* communicat* presentation education priming

Each 'DCE' search term was combined with each of the 'presentation difference' search terms and used to search each of the research databases. Search by database are listed below in Table 3.2. The search included papers between January 2011 and November 2018. A total of 4176 publications were found through the search, with 3714 distinct studies identified. Table 3.2 provides a breakdown of search results by database.

Table 3.2 Searches by research database

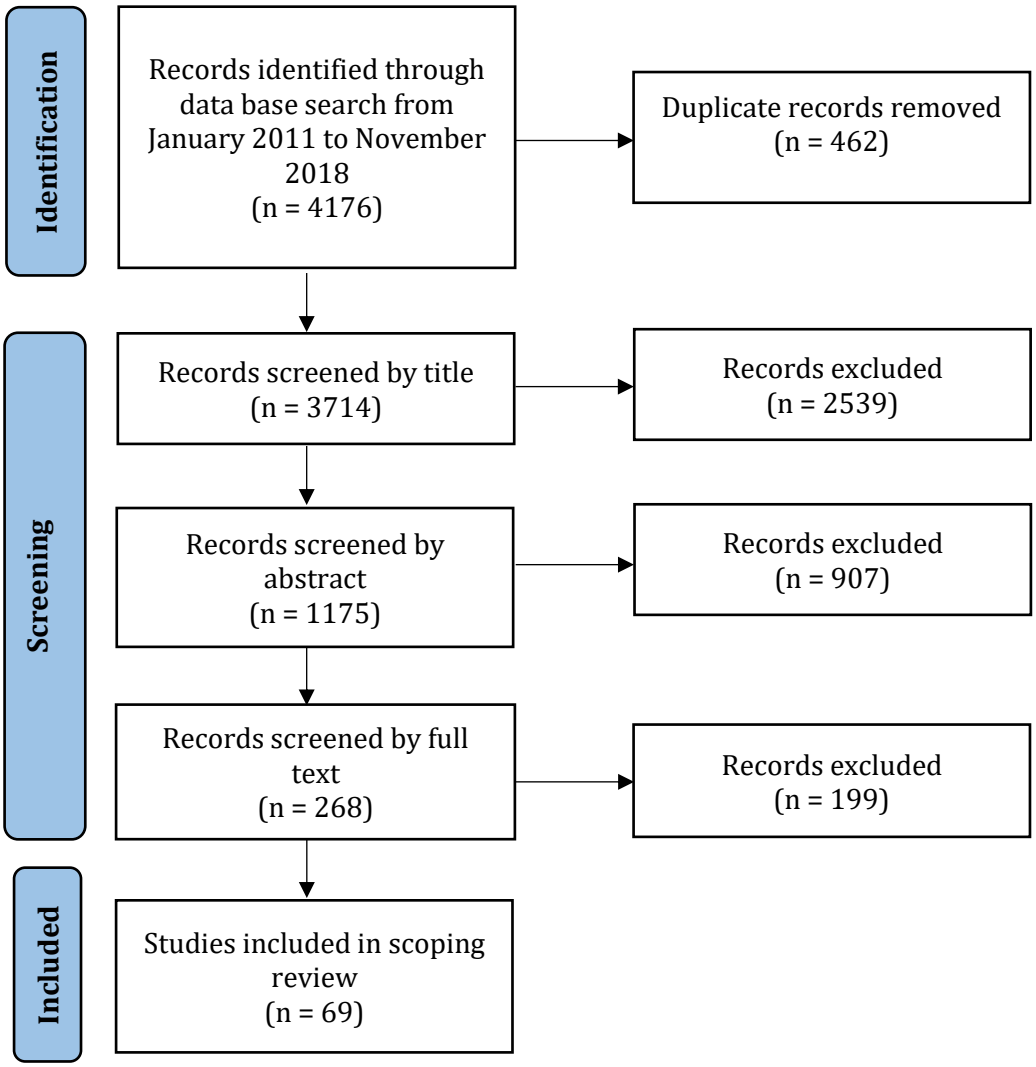
Database	No. of search results imported into Endnote
EBSCO Host*	2163
Pubmed	683
Scopus (Elsevier)	679
Web of Science Core Collection	651
Total of papers identified	4176
Total of distinct papers	3714

* Business Source Complete, CINAHL, EconLit, MEDLINE, PsycInfo, Psychology and Behavioural Sciences Collection

Records were first screened by title. At this stage, the goal was to ensure that the title of the study was broadly related to the literature search. For instance, if the title included terms such as 'preferences', 'choices', 'decision making', 'attributes', 'values', 'surveys', 'communication' or 'presentation', then they were included for further reading. If the title was ambiguous and its relevance could not be determined by the title alone, then the publication was included for further screening by abstract. From the 3714 records screened by title, 1175 were included for a further screen by abstract. The information provided in the abstract was sufficient, in most cases, to determine whether it was a

stated preference technique like a DCE and whether or not it included multiple arms. However, there were instances where this information was ambiguous and so 268 publications were screened by full text as well. At this stage, the focus was on the methods section and then the results section of studies. In particular, some studies that appeared to be DCEs from the abstract were identified as not being DCEs (as defined in Chapter 2) based on the description in the methods section of the publication. There were also some studies that included only a within-subjects design rather than a between-subjects design and, as a consequence, were excluded. Figure 3.2 provides a summary of the screening process. In total, 69 publications were included for the scoping review based on the inclusion criteria. A list of these publications is available in Appendix 3A.

Figure 3.2 Screening process flow diagram



3.4.2 Data extraction details

The data extracted from the included publications are summarised in the following section. Data to be extracted was developed based on what had been reported in previous literature reviews of DCEs (Rakotonarivo et al., 2016; Soekhai et al., 2019), and to ensure that the extracted data was comprehensive for the the aims of this review. Data extracted covered seven main components: general study details, survey development details, choice set construction details, choice set presentation details, details of arms, details of econometric models used and whether there were any differences in results detected between arms. Detailed descriptions and definitions of the scope of data extracted can be found in Appendices 3B to 3G.

3.5 Results: characteristics of studies

The purpose of this section is to provide an overview of the characteristics of the studies included in terms of criteria typically used to summarise DCEs.

3.5.1 General study details

Year of study

Table 3.3 provides as summary of when included studies were published. The majority of studies were from 2016-2018.

Table 3.3 Studies included by year

Year	No.
2011	7
2012	13
2013	5
2014	3
2015	5
2016	12
2017	14
2018	20

Country of study

Table 3.4 provides a summary of where studies were conducted.

Table 3.4 Country of study

Country of Study	No.	Country of Study	No.
USA	14	Switzerland	2
Australia	9	Croatia	1
Germany	7	Greece	1
Netherlands	7	Ireland	1
Canada	5	Italy	1
UK	4	Peru	1
China	3	Saudi Arabia	1
Denmark	2	Taiwan	1
France	2	Uruguay	1
Japan	2	Zambia	1
Spain	2	Not reported	1

The most common country of origin for studies was the USA, followed by Australia, Germany and The Netherlands, making up over 50% of the studies included in this scoping review. Other studies were conducted across a range of countries and regions. There was one case where the country of study was not explicitly mentioned.

Dissemination method of DCE

Data collection for 60% of studies was undertaken through an online survey. There were also several studies where respondents were required to visit a certain location to participate (e.g. laboratory) or only qualified for participation if they were at locations specific to study interest, such as shopping centres, fishing sites, residential areas etc. Computer-assisted personal interviews and face-to-face interviews were also used. A summary is provided in Table 3.5.

Table 3.5 Dissemination method of DCE

Dissemination method	No.
Online survey	41
Specified location	7
Face to face interview	6
Computer assisted personal interview	6
Mail survey	5
Not reported	4

Population of interest

Over half the studies included utilised a general population sample (see Table 3.6). There were also a substantial number of studies which required respondents to meet certain criteria to be eligible to participate (e.g. user of a particular good or service, or

required to have a specific diagnosis or condition). A small number of studies recruited through convenience methods (e.g. on campus at a university).

Table 3.6 Population of interest

Population of interest	No.
General population	35
Study specific population	29
Convenience samples	5

Study sample size

The number of respondents included in the study ranged widely, from 60 to 7200; summary statistics are provided in Table 3.7. There were two studies (De-Magistris et al., 2013; Kostyra et al., 2016) where the number of respondents used in the study was not clear; as a result, only 67 studies were included in this summary.

Table 3.7 Study sample size summary

Study sample size*	No. of respondents
Mean	1112
Median	743
SD	1170
Min	60
Max	7200
Interquartile range (IQR)	1001

*to the nearest whole number

3.5.2 Study development details

Details of the methods used to develop the study are summarised below. Most of the studies (87% (60/69)) reported the use of some type of qualitative or other (e.g. quantitative) method to develop the DCE in studies. This section is divided into qualitative methods and 'other' methods.

Use of qualitative methods

This section records whether there was any use of qualitative methods in the development of the study and, if so, what they were. Of the studies, 52% (36/69) reported the use of qualitative methods. This information is summarised in Tables 3.8 to 3.10.

Out of the 36 studies that used qualitative methods, a majority (66%) used only one type of qualitative method. There were a number of studies that used two types of

methods, with three studies each using three different methods (Lanz & Provins, 2015; Wright et al., 2017; Zhao et al., 2013).

Focus groups and some form of interview were the most popular qualitative methods. Consultation with experts and relevant stakeholders was also used to a lesser extent. Qualitative methods were most often used to develop the attributes and/or levels to be used in the survey. They were also used for the purpose of refining the survey or parts of the survey. There were only three instances (Grisolía et al., 2018; Spinks & Mortimer, 2015; Williamson et al., 2016) where qualitative methods were used to refine the presentation differences to be used in the study.

Table 3.8 No. of qualitative methods used

No. of qualitative methods used	No.
0	33
1	24
2	9
3	3
Total	69

Table 3.9 Type of qualitative methods used

Type of qualitative methods used*	No.
Focus groups	20
Interviews	20
Consultation with experts/stakeholders	11
Total	51

*studies could use more than one type of qualitative method

Table 3.10 Rationale for qualitative methods

Rationale for qualitative methods*	No.
Development of attributes/levels	29
Development of survey/parts of survey	13
Development of presentation difference/s	3
Total	45

*studies could have more than one rationale for use of qualitative methods

Other methods used to develop study

Although only about half the studies used qualitative methods, it was noted that the use of other means to assist in the development of the study was quite common. Indeed, 78% (54/69) of studies reported some means of refining the study prior to implementation. Tables 3.11 to 3.13 provide a summary.

Table 3.11 No. of methods used

No. of methods used	No.
1	35
2	17
3	2
Total	54

Among those studies that used other methods to refine the study, just over a third (35%) used more than one method. Pilot testing of the study or parts of the study prior to the main data collection was the most common method, followed closely by using evidence from the literature, including references to both previous studies and literature reviews. There were also several studies that sought to be as realistic as possible in terms of what respondents would see in the market, with many using market features as a point of reference for the study development, at least in part.

Table 3.12 Type of method used

Type of method used*	No.
Pilot testing	29
Evidence from published literature	26
Reflection of market	19
Respondent provided information	1
Total	75

*studies could use more than one type of method

Table 3.13 Rationale for method use

Rationale for method*	No.
Development of attributes/levels	43
Development of survey/parts of survey	21
Development of presentation difference/s	3
To obtain priors to inform choice set construction	2
Total	69

*studies could have more than one rationale for use of method

There was one study that utilised information provided by the respondent (Hyland et al., 2018). In this study, information for one of the attributes differed depending on the information a respondent provided about their home/work location.

In terms of the main purpose of using these methods, the refinement of attributes and/or levels was the most common reason. This was followed by developing the

survey or parts of the survey. In two studies, other methods were used to inform the construction of choice sets via collection of information for priors.

Use of validity checks

As part of quality checking DCE data, researchers can include internal validity checks as part of the survey (Johnson et al., 2019; Soekhai et al., 2019). Internal validity checks include repeated choice sets where researchers use the same choice set twice in a DCE to check whether the same response is given on both occasions. Dominance tests may also be used where researchers intentionally include a choice set where one option is clearly better than the other/s. A test for transitivity could also be included to test whether respondents satisfy the assumption of rational decision making in utility theory. Transitivity holds if option A is preferred to option B, option B is preferred to option C and option A is also preferred to option C (Johnson et al., 2019). Hold out tasks could also be included in the DCE to test for response reliability and also for how predictive the resulting model is (Janssen et al., 2017). One or more choice sets can be included in the DCE but, during analysis, are used to test the predictive ability of the model rather than as an input to the model estimation.

Most studies included in the review did not report the use of any pre-planned test for the validity of responses. Only 19% (13/69) of studies included some type of check. Details of the type of validity check are summarised in Table 3.14.

Table 3.14 Types of validity checks

Type of validity check	No.
Repeated choice sets	5
Dominance test	3
Hold out tasks	2
Transitivity test	1
Other	2
Total	13

There were two studies that utilised an alternative form of validity check. In a study by Logar and Brouwer (2018), respondents were asked for a reason if they consistently chose the opt-out option. Williamson et al. (2016) used an exercise to verify whether respondents absorbed the information given. In this study, respondents were randomly

assigned different advertorial style articles that highlighted different positive messages about Australia. Respondents were asked to associate a list of countries, including Australia, with a list of statements describing quality of produce.

3.5.3 Choice set presentation details

The focus of this section is on how choice sets were presented to respondents. This section was broken down by arm, so if the number of options, attributes or levels was different for each arm, they were assigned a separate observation; hence, the total number of observations is larger than the number of studies.

Choice sets

Tables 3.14 to 3.16 summarise information about choice sets seen by respondents. In terms of the total number of choice sets included in a study, this varied widely, ranging from 4 to 560 choice sets, although there were 44 instances in which this information was not provided. Table 3.15 provides a summary.

Table 3.15 Total number of choice sets

Total no. of choice sets	No.
4 to 20	16
21 to 50	19
51 to 100	5
101 to 200	7
201 and over	4
Not reported	44
Total	95*

*studies could use more one set of choice sets e.g. by arm, hence exceeds number of studies

The majority of respondents completed 20 or fewer choice sets. In terms of popularity, completion of 12 choice sets was most common (18/95), followed by 8 choice sets (10/95) and 16 choice sets (9/95) per respondent. There were three studies where this information was not provided. There was one study which was unique in that each respondent was given six choice sets but was also asked afterwards if they would like to complete 0, 6 or 12 more (Hyland et al., 2018). Table 3.16 provides a summary.

Table 3.16 No. of choice sets per respondent

No. of choice sets per respondent	No.
1 to 10	38
11 to 20	45
21 to 30	7
32	1
Other	1*
Not reported	3
Total	95

* Hyland et al (2018)

Choice sets were frequently shown in blocks, meaning respondents saw only a proportion of the total number of choice sets included in a study (see Table 3.17). Choice sets in blocks could be presented in random order or in the same order for each respondent. It was fairly common for all choice sets included in the study to be shown to each respondent but in random order. In 29 instances, information about choice set assignment was not provided or was not clear.

Table 3.17 How were choice sets assigned?

How were choice sets assigned?	No.
Blocks	39
Random	27
Not reported	29
Total	95

Number of options and using opt out

The number of options that respondents saw in choice sets varied from one to seven; Table 3.18 provides a summary. The inclusion of two options in a choice set was by far the most popular format, followed by having 3–4 options. There were four studies that included one option – that is, a binary choice set, where respondents could accept or reject the option shown. In one study, respondents were given a set sequence of choice sets with varying number of options and were also allowed to choose the number of options seen in each of the last few choice sets (Burton & Rigby, 2012).

Table 3.18 No. of options

No. of options	No.
1	4
2	53
3	20
4	14
5	2
7	1
Other	1*
Total	95**

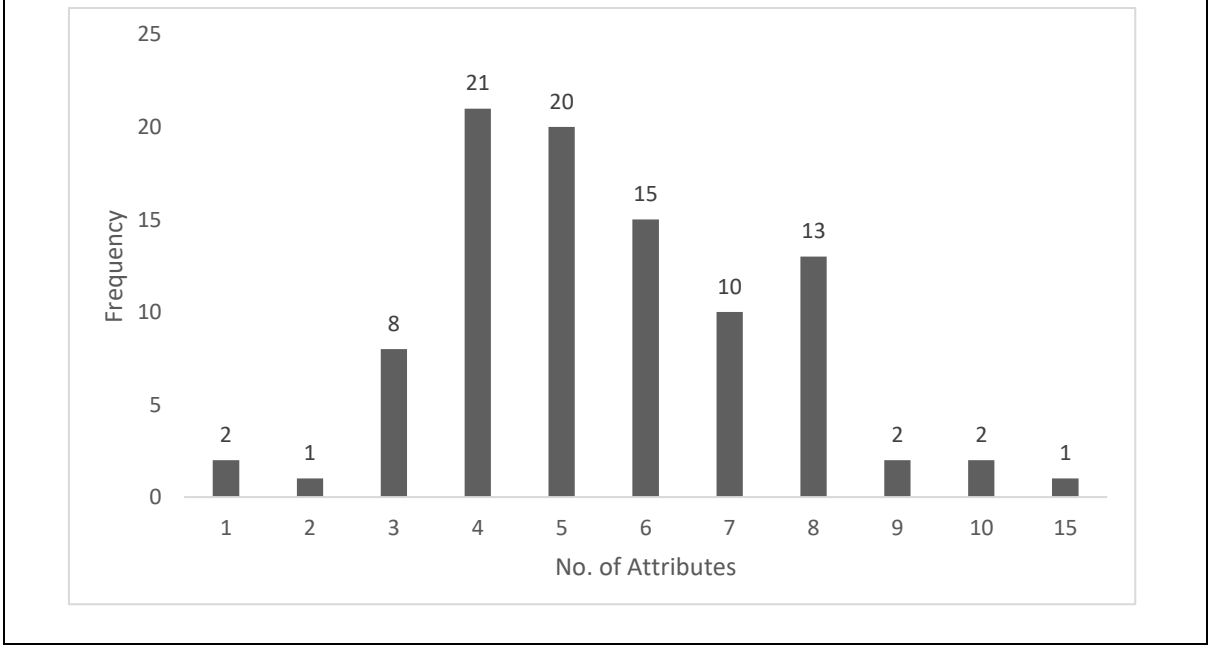
*Burton & Rigby (2012) ** studies could use different number of options by arm, hence exceeds number of studies

An opt-out option was included 72% (68/95) of the time. When an opt-out option was included, it was often part of the choice set. There were only 5/68 instances where it was a separate question after the choice set (Baba et al., 2016; Harris et al., 2015; Heinzle, 2012; Veldwijk et al., 2016; Williamson et al., 2016).

Number of attributes

Figure 3.3 summarises the number of attributes used to describe the options seen by respondents. This did vary by arm in some studies, and where it did this was recorded as a separate observation. Hence, the total number of observations is larger than the number of studies (95 versus 70 respectively).

Figure 3.3 No. of attributes presented to respondents



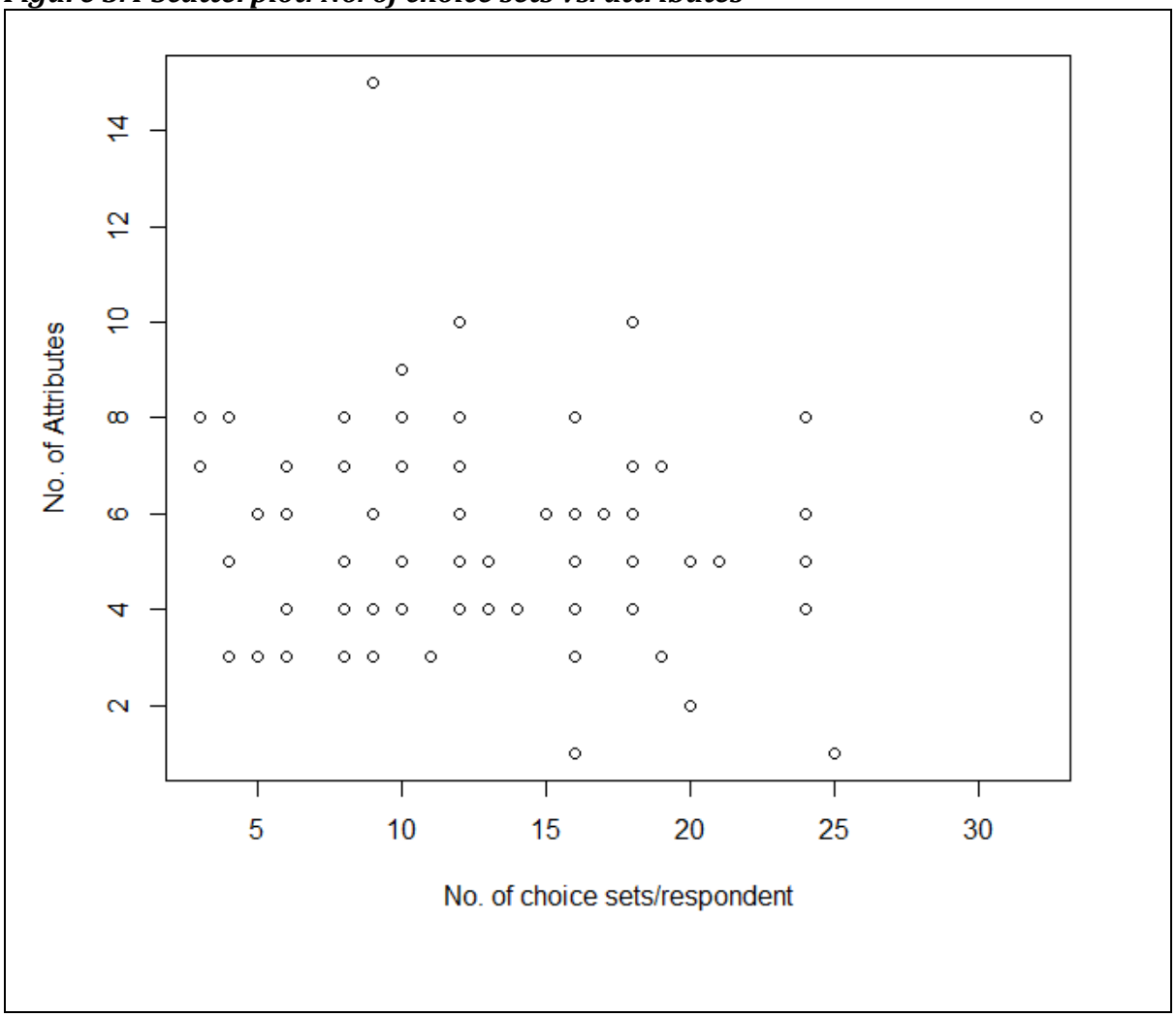
In particular, it should be noted that 16 of these observations came from one study alone (Oehlmann et al., 2017). This study systematically varied the number of options (2–4), attributes (4–7) and levels (2–4) seen by respondents.

The number of attributes seen varied from one to 15, although respondents most frequently saw between 4–6 attributes, with four attributes being most frequent.

Number of attributes vs number of choice sets per respondent

The relationship between the number of attributes and the number of choice sets completed by each respondent was examined through a scatterplot (see Figure 3.4 below). Only 91 observations from 65 studies were included for this analysis as three of the studies did not report the number of choice sets completed per respondent (Jin et al., 2017; Oppewal et al., 2015; Patterson et al., 2017) and one study had a unique way of presenting choice sets (Hyland et al., 2018). Pearson's correlation formula, performed in R Studio (RStudio Team, 2020), was used to examine the relationship between the number of attributes and the number of choice sets completed by each respondent in an arm. A slight negative correlation was noted at −0.0579; however this was not significantly different from 0 ($p = 0.6$).

Figure 3.4 Scatterplot: No. of choice sets vs. attributes*



*includes 65 studies only as 4 studies did not provide relevant information about number of choice sets per respondent

3.5.4 Choice set construction details

Table 3.19 provides a summary of whether studies provided details of what aspects the choice sets were designed to model. It should be noted that studies can have multiple arms where construction methods differ. As a result, the total number of observations is more than the number of included studies.

The majority of studies did not provide details of what aspects the construction method was designed to model. Among those that did, modelling of main effects only and of main effects and two-way interactions were most common. There were two studies where the construction method was designed to detect other features. Gajic et al. (2012) used a construction method designed to detect main effects, all two-way interaction effects and some higher-order interaction effects. The DCE study by Antoniou and

Kostovasilis (2017) was designed to detect main effects, interactions and quadratic effects in differences.

Table 3.19 What is the construction method designed to model?

What is the construction method designed to model?*	No.
Main effects	13
Main effects and two way interactions	10
Main effects, 2 way and higher order interactions	2
Main effects and interactions	3
Not reported	47
Total	75

*studies could use more than one construction method
** Antoniou and Kostovasilis (2017); Gajic et al. (2012)

The majority of construction methods in the included studies were optimised for D-efficiency, although many studies did not report this information. For the studies utilising a Bayesian construction method, a Bayesian D-efficiency criterion was used, although there was one study by Oehlmann et al. (2017) which used C-efficiency to minimise the sum of the variances of the marginal WTP estimates reported in the study. Table 3.20 summarises this information.

Table 3.20 What was the construction method optimised for?

What was the construction method optimised for?*	No.
D-efficiency	28
Bayesian D-efficiency	7
C-efficiency	1**
Not reported	39
Total	75

*studies could use more than one construction methods hence total larger than number of studies **Oehlmann et al. (2017)

In terms of software used for construction, Ngene was quite common followed by SAS and Sawtooth. In several studies authors used custom code in programs including epiGenesys, R, Julia, SSI Web 8.3.2 and VBA in Microsoft Excel.

Table 3.21 What software was used for construction?

What software was used?	No.
Ngene	19
SAS	10
Sawtooth	7
Custom code	5
SPSS	3
JMP	2
Not reported	29
Total	75*

* studies could use more than one construction method hence total larger than number of studies

In the majority of cases (63%), information regarding the actual method used to construct choice sets was not reported. Table 3.21 summarises the variety of methods that were reported, while Table 3.22 provides information about the studies that utilised other methods.

Table 3.22 Method to construct choice sets

Method used to construct choice set (can be more than one)	No.
Bayesian design methods	8
Generator developed design	3
Methods requiring attribute/level changes	2
Methods requiring option changes	5
Random design method	3
Sawtooth: Complete enumeration strategy	3
Sawtooth: Balance overlap method	2
Other	2*
Not reported	47
Total	75**

*see Table 3.22 ** studies could use more than one construction methods hence total larger than number of studies

Table 3.23 Method of construction: other

Authors	Year	Description
Mitrani	2016	boundary value based approach
Arrua et al	2017	catalogue based design: designed following the mix and match procedure (Johnson, Kanninen, Bingham & Ozdemir, 2007)

There were 18/75 instances where studies reported using priors of some kind. This included point priors and prior distributions for Bayesian design methods. Priors used

were based on several methods, although details about where priors came from were not always provided. Among studies that did, this included the use of pilot data (Grisolía et al., 2018), previous studies (Johnston et al., 2016) or simulations (Devarasetty et al., 2012). There were eight instances where Bayesian design methods were used, of which only one study reported on the priors used. Devarasetty, Burris and Shaw (2012) provided a tabular summary of the mean and standard deviations of attribute priors.

3.6 Results: Presentation differences and DCE outcomes

This section provides details of the differences in information between the arms in the studies.

3.6.1 Arms in study

The number of arms in each study has been summarised in Table 3.24. Just over half of the included studies used a split sample approach wherein respondents were divided into two arms. There were two studies that used more than 10 arms; Amano et al. (2016), who used 12 arms, and Oehlmann et al. (2017), who used 16 arms. Oehlmann et al. (2017), in particular, were concerned with the effect of changing the number of options, attributes and levels used in a DCE on the results obtained.

Table 3.24 No. of arms in study

No. of Arms in Study	No.
2	36
3	13
4	10
5	4
6	1
7	2
8	1
12	1
16	1
Total	69

Over 60% of studies indicated that respondents were randomly assigned to arms (see Table 3.24). There were three studies conducted by mail survey where the different arms involved creating different versions of the survey that were then distributed randomly. It was not possible to adaptively randomise respondents as surveys were

completed, so the responses per arm could not be controlled for. It was also noted that 25% of studies did not report how arms were assigned.

Table 3.25 How were arms assigned?

How were arms assigned?	No.
Random	43
Random by mail	3
By convenience	3*
By specific criteria	3**
Not reported	17
Total	69

*see Table 3.26 ** see Table 3.27

There were also three studies where convenience was used for arm assignment, details of which are provided in Table 3.26. These included a study by Rid et al. (2018) in which respondents were assigned to an arm based on their internet connection speed.

Table 3.26 By convenience: details

De-Magistris & Gracia, Nayga Jr	2013	Participants in each session completed the same treatment
Zhou & Bukenya	2016	People were approached for interviews at appliance retailing supermarkets in different districts of Nanjing. States in the paper that efforts were made to keep the sample random but doesn't specify how
Rid et al	2018	Respondents directed to still images if have only standard internet connection, directed to 3D film sequence if had broadband internet access

In three studies, respondents were not just randomly assigned to one arm but also assigned based on different criteria (see Table 3.27). This included studies where respondents were asked to complete more than one arm.

Table 3.27 By criteria: details

Kenny et al	2017	Each respondent assigned to see 1 numeric format (styles 1 or 4) and 1 star rating format (style 2 or 3).
Dellaert, Donkers, van Soest	2012	Each respondent assigned to two arms
Bech, Kyaer, Lauridsen	2011	Group 1: assigned to 1 of 8 blocks of 4 choice sets. Group 2: assigned to 1 of blocks 1+2, 3+4, 5+6 or 7+8. Group 3: assigned to 1 of blocks 1-4 or blocks 5-8

3.6.2 Which aspect/s of the DCE were different between arms?

In terms of describing the presentation differences used in studies, the parts of the survey or DCE that were different between arms were explored (summarised in Tables 3.28 and Table 3.29). The vast majority of studies only had one aspect of the survey that was different between arms. There were eight studies where more than one aspect of the survey was different between arms. In two studies, it was not clear which part of the DCE was different between arms.

Table 3.28 No. of aspects of the DCE different between arms

No. of aspects of DCE different	No.
1	59
2	8
Not reported	2
Total	69

The choice sets were a common aspect of the DCE that differed between arms. This included differences in choice set presentation between arms as well as differences in terms of the number of choice sets, options or attributes seen by respondents. Differences in terms of what is presented or provided in the introductory section were also common. Manipulating the vignette or choice scenario was another aspect that was common. There were two studies in which experience with product or service was different between arms. This included Meenakshi et al. (2012), where respondents in one arm were allowed to take a product home to test it for a period of time prior to answering the DCE. In Torquati et al. (2018), respondents in one arm had a sensory test then participated in the DCE while in other arm the order was reversed.

Table 3.29 List of parts different between arms

List of parts different*	No.
Introductory section	24
Vignette	12
Choice sets	35
Other: Experience with topic	2**
Other: Incentives to participation	1***
Not reported	2
Total	76

*could be multiple parts different between arms in a study**Meenakshi et al. (2012); Torquati et al. (2018) *** Gajic et al. (2012)

3.6.3 How have presentation difference been used?

After considering which part/s of the DCE were different between arms in the included studies, the number of ways information was different between arms in studies was then explored (summarised in Table 3.30). The majority of studies only had one way in which information was different between arms, although there were also quite a few studies where information differed in two or more ways.

Table 3.30 No. of presentation differences used per study

No. of presentation differences used per study	No.
1	52
2	13
3	4
Total	69

Table 3.31 summarises the different ways information was different between arms or, rather, the presentation difference types used in studies. In following sections, some of the studies are summarised by presentation difference type. This is for the purpose of providing the reader with a better idea of presentation difference types used.

Table 3.31 Frequency of different presentation difference types

Presentation difference type	No.
Differences in text	40
Amount of information provided	22
Differences in presentation format	19
Structure of choice sets: Number of options	5
Structure of choice sets: Number of attributes	3
Structure of choice sets: Number of levels	1
Other: Experience with topic	2
Total	92*

*DCE can be presented differently in more than one way in a study

Summary of studies: differences in text

The most common difference between arms was in terms of the text shown to respondents i.e. framing of text. A classic example of framing as a gain (positive frame) or loss (negative frame) can be seen in Veldwijk et al. (2016). In this study, the survival attribute was described in terms of the probability of surviving colorectal cancer (positive frame) or in terms of mortality, probability of dying from colorectal cancer

(negative frame). There are also studies that investigate particular ways to phrase potential gains or losses. For instance, Grisolía et al. (2018) had three arms in their study, with the potential gain in health outcome phrased differently. The attribute for health outcome was framed in terms of reduced risk of a heart attack in the next 10 years, extra months of life expectancy or as an increase in the probability of reaching your full life span.

Studies also made use of the anchoring effect. Jahn et al. (2018) provides an example in the context of providing information about the average amount of sugar in muesli and investigated its potential impact on sugary food choice. In the introductory text to this DCE, respondents were given either a low or high average amount of sugar in muesli to consider. Alternatively, Meyerhoff and Glenk (2015) investigated the impact of a practice choice set on stated WTP for water quality changes. In the practice choice set, depending on arm assignment, respondents were given either a low or high price to consider.

There were also studies where each arm in the DCE presented slightly different text information that meant the choice scenario or choice to consider was different. For instance, in a study by Sicsic et al. (2016), respondents were assigned to one of three arms, with the choice scenario framed in terms of breast, cervical or colorectal cancer screening. Similarly, researchers have asked respondents to imagine different scenarios depending on arm assignment including good or bad weather (Hyland et al., 2018), business or leisure context for travel (Kim & Park, 2017) as well as different health conditions when making choices (Mühlbacher et al., 2016; Spinks & Mortimer, 2015).

Studies also framed text by using different terminology to describe an outcome or concept. For instance, Zhao et al. (2013) had two arms in their study, investigating ecological indicator choices. One of the attributes was framed in terms of two similar ecological indicators: the probability that migratory species will still be migrating in the river in 50 years, or the expected number of fish that will swim upstream each year. Alternatively, Nickel et al. (2018) used different terminology to describe the same condition, papillary thyroid cancer and papillary thyroid lesion.

Summary of studies: amount of information provided

The next most common presentation difference type seen in the included studies was the amount of information provided. For instance, Kostyra et al. (2016) provided one arm with extra information in the form of a summary rating as an extra attribute to the choice sets, which the other arm did not see. Benning et al. (2012) presented extra information in one arm about an invasive follow-up test, including its details in the introductory information, whereas the other arm was not privy to this information.

Summary of studies: differences in presentation formats

Studies also presented information in different formats between arms. This included the use of pictures, graphs, videos and simulated environments. For instance, Arrua et al. (2017) had two arms, with one shown a traffic light labelling system and the other shown the Chilean nutrition warning system. Johnston et al. (2016) also had two arms, with one receiving a generic policy area map and the other being provided a slightly more personalised map with a yellow dot indicating their local area. Rid et al. (2018) had one arm view still images while the other arm was shown a 3D film sequence.

Summary of studies: structure of choice sets

Although less frequent, the scoping review also found studies that chose different number of options, attributes or levels to present to respondents depending on arm assignment. The most prominent is Oehlmann et al. (2017), where the researchers had a total of 16 arms with each varying in terms of the number of options, attributes and levels presented to respondents. There were also a number of studies where the number of options and the inclusion of an opt-out was dependent on arm assignment (Jarvis, 2012; Parker & Schrift, 2011; Zhang & Adamowicz, 2011).

Summary of studies: multiple presentation difference types

There were also a number of studies that utilised more than one presentation difference type. Knox et al. (2013) provides an example of a combination of presentation difference types used, varying in both framing of text and the amount of information provided. This study investigated contraceptive choices. Respondents were assigned to one of three arms. The control arm received only basic information about contraceptive products. Another arm received basic information and also information about the risks of the combined oral pills (negative/loss frame). The third arm received basic

information and information about the benefits of a new contraceptive product (positive/gain frame).

In the study by Dellaert et al. (2012) the researchers used a combination of structure of choice sets and framing of text. This study had two arms, with respondents seeing a higher (lower) number of options and attributes depending on arm assignment. In addition, attribute levels could either be more similar or less similar depending on arm assignment. Alternatively, Amano et al. (2016) used a combination of different presentation formats and amount of information provided. The amount of introductory information provided to respondents varied by arm assignment. In addition, respondents saw the cancer risk attribute expressed as either a fraction or as a multiple of the World Health Organization standard.

3.6.4 Analysis of the impact of presentation differences

Models used to analysis DCE data

In this section, the models used to analyse the DCE data are reported. Most studies only used one type of model (see Table 3.32); however, it was also common to use two models. For one study, the authors did not examine the actual DCE results; instead, the focus was on the impact of incentives (Gajic et al., 2012).

Table 3.32 No. of models used

No. of estimation procedures used	No.
1	41
2	22
3	3
4	2
Other	1*
Total	69

*Gajic et al. (2012)

The mixed logit model (MXL) was by far the most frequently used model for the DCE data (see Table 3.33). The multinomial logit (MNL) or conditional logit (clogit) model and its variations was the second most frequent. There were also a few studies that used less common methods; these have been detailed in Table 3.34.

Table 3.33 List of models used

List of estimation procedures used*	No.
Conditional logit(clogit)/ Multinomial logit (MNL)	20
Variation of clogit	7
Logit/probit model and variations	6
Hierarchical Bayes (HB) estimation	9
Mixed Logit (MXL)	40
Generalised MNL (gmnl)	6
Latent class analysis (LCA)	2
WTP space model (Train & Weeks, 2005)	2
Other	6**
Total	98

*More than one model may have been estimated for each study
**see Table 3.34

Table 3.34 Other modelling procedures used

Author	Year	Description
Hyland et al	2018	Monte Carlo simulation methods used to estimate main and interaction effects with presentation difference
Sandorf, Sourd, Mahieu	2018	Use of DCE estimation procedures but also used an estimation procedure according to the random regret minimisation model (as opposed to using the traditional random utility model)
Lanz, Provins	2015	Regression used to explore Status Quo choices specifically
Dellaert, Donkers, van Soest	2012	Modified logit model which parameterises complexity and individual decision time (see page 430 and following in paper).
Gajic, Cameron, Hurley	2012	This paper did not look at the actual DCE results, instead the focus was on the impact of incentives. The focus of analysis was on other data including contact rates, response rate. cost effectiveness of incentives (ICER, dollar amounts), response completeness, speed of response and consistency of response
Vossler, Doyon, Rondeau	2012	WTP regression based on maximum likelihood estimator of Cameron and James (1987). WTP treated as a censored dependent variable, choice sets involve only 1 option (vote yes or no to maintain status quo)

Comparison of results by arm

In examining results for the included studies, the methods used to test for differences by arm were extremely varied. This meant that it was difficult to find standardised criteria to define when there is a difference in results by arm. Hence, the question has been defined in terms whether there are any detected statistical differences by arm as reported from the methods used by the authors (summarised in Table 3.35). The vast

majority (94%) of the studies detected some difference in results by arm. There were only four studies that reported no differences by arm. These results should be viewed with caution, as statistical difference does not necessarily translate to an economically significant difference (Rakotonarivo et al., 2016).

Table 3.35 Any differences in DCE results by arm?

Were there any significant differences by arm?	No.
yes	65
no	4*
Total	69

*Devarasetty et al. (2012); Spinks and Mortimer (2015); Vass, Rigby et al. (2018); Zhao et al. (2013)

Considering there were four studies that reported no difference by arms, a closer look was taken at the differences in arms in these studies. Devarasetty et al. (2012) chose to present an attribute as text or in a graphical format. Similarly, in Vass, Rigby et al. (2018) the difference between arms was in the chosen format for risk presentation. In this case, it was presentation of risk as a percentage only or as a graphical array. In Spinks and Mortimer (2015), respondents were asked to imagine different mild ailment conditions, mild joint pain or insomnia. In Zhao et al. (2013), similarly to the other three studies, only two arms were used. In this study, one attribute was framed by one of two indicators of a similar concept.

3.7 Discussion

3.7.1 Summary of findings

The goal of this scoping review was to gain a picture of the existing literature where a DCE is presented differently across one or more arms within a single choice experiment. There were 69 studies in total between 2011 and 2018 that met the inclusion criteria, indicating that there are many studies in the literature in this area of research. As far as this author is aware, this is the first literature review to investigate presentation differences in DCEs, as a between-subjects design, across multiple disciplines. Studies included covered a broad range of topics including health, consumer goods and services and environmental topics, including land use, housing and national parks.

There was significant variation in terms of study location and sample size. The majority of studies were conducted online. The majority of studies also sought to improve the

quality of the DCE through qualitative methods and through other means such as referring to previous literature or current market conditions. There were only a few studies that had pre-planned tests for the validity and consistency of responses. This is consistent with what has been observed in the health economics literature (Soekhai et al., 2019).

Most studies also reported the number of choice sets completed by each respondent and the number of options within each choice set. There was less consistency when it came to details about choice set assignment and the total number of choice sets included in the study. Many studies did not report the method for choice set construction and when it was reported, it was often very brief with limited information given.

In this scoping review, 95% of studies reported some difference in results by arm. This figure is substantially higher than the findings of Rakotonarivo et al. (2016), who report no difference in the majority. It should be noted that Rakotonarivo et al. (2016) investigated studies that could use either between-subjects or within-subjects designs, although the majority of studies included were between-subjects designs. In addition, outcomes included not only the impact of differences in presented information but also temporal stability of results. Findings were also specific to the environmental economics literature. As a consequence, findings from Rakotonarivo et al. (2016) are only partially relevant to the current scoping review. However, the evidence from Rakotonarivo et al. (2016) suggests the potential presence of under-reporting of the finding of no difference between arms in the DCE literature as a whole.

3.7.2 Limitations of findings and directions for future research

In this scoping review, the intention was to understand what has been investigated in the existing literature. For this reason, key details of each DCE have been reported without assessing the quality of each study. In addition, the current review only reports the conclusions reported by the authors, rather than making a separate judgement about the suitability of data analysis techniques used in studies. This review also does not define what constitutes a significant or economically significant difference in results between arms.

During the screening process, care was taken by the author to be as inclusive as possible. For instance, the term conjoint analysis was used as a search term, to ensure

that articles that were in fact DCEs, but were labelled as conjoint analysis, were still included for consideration. However, some studies may have potentially been missed at the screening by abstract stage. Studies that were DCEs but did not appear to be so in the abstract, could have potentially been excluded. As such, the studies included in the scoping review may not be exhaustive.

Future work could assess whether different conclusions would emerge were a quality assessment applied such that only 'high-quality' DCEs were included. This was considered beyond the scope or aims of the current review. Inconsistencies in terms of what was reported in studies were noted, and one quality criterion could be to exclude studies that did not report key information. Studies also varied in terms of whether respondents were randomly assigned to arms or whether other criteria were used, though it should be noted that practical concerns may limit randomisation – for instance, some of the included studies delivered the survey by mail. There may also be difficulties with dissemination of the survey that make randomisation more challenging, such as targeting older populations or respondents in remote regions. In addition, a subsequent review could examine in more detail the estimation procedures and outcome measures used. In this scoping review, three main types of presentation differences were investigated. It may be of interest to some researchers to focus on one type of presentation difference type when conducting future reviews.

Future studies in this area are strongly recommended to indicate through keywords and abstract terms that the study involves a comparison of different ways a DCE can be presented between two or more arms. The author acknowledges that different terminology may be used across different disciplines. However, some indication that respondents were split into two or more arms would be very helpful. For instance, studies could include general words such as 'arms', 'split sample' or 'between-subjects design'. Even more helpful would be indications that a presentation difference type was used. The author has chosen to use the term 'presentation difference' and acknowledges that this may not be a familiar term to many researchers. However, researchers could include terms that have been used in the literature to indicate the specific type of presentation difference used, such as, 'presentation', 'framing effect', 'anchoring effect', 'gain/loss framing' , ' design of design' or ' amount of information'. The search for studies in this area was made more difficult by the lack of relevant keywords. The

search for studies in this area of research often required reading the methods section, rather than just the abstract or keywords, to determine an article's suitability for further consideration. In particular, often there was no indication whether or not the study was a between-subjects design until the methods section is reached. Authors of future DCE studies are also encouraged to provide more details about DCE construction and design. This would benefit readers by providing useful information about the current methods being used and improve the chance of being able to replicate the study.

3.7.3 Implications of findings

Based on the current scoping review and related reviews such as Rakotonarivo et al. (2016), it appears there are many studies in the DCE literature that have sought to compare presentation differences in DCEs and their impact on DCE findings. The current scoping review had an overwhelming majority of studies that found presentation differences did lead to statistically significant differences in results. In contrast, Rakotonarivo et al. (2016) found that the majority of articles reviewed did not lead to significant differences in results. Based on the current scoping review and the review by Rakotonarivo et al. (2016), the impact of presentation differences in DCEs on DCE findings are mixed. There are instances where small differences in how a DCE is presented have led to significant differences in results; conversely, and there are instances where they have not. This could, perhaps, be due to the variation in many aspects of DCEs across studies.

3.7.4 Next steps

At first glance, it may not seem clear what is to be done with this information However, from the perspective of DCE researchers in health economics, this conclusion raises promising possibilities. If DCE presentation choice does not necessarily lead to differences in results, the queston arises of whether it is possible to intentionally design a DCE with two or more arms that employs different presentation choices but does not lead to differences in results. More specifically, to what extent can the amount of information provided in one arm differ from/be greater than that provided in another arm without impacting on results? This is of particular importance in health economics where extra information is often provided in a DCE with the aim of ensuring respondents understand the task but without any intention to influence respondent

choices (see Chapter 1). In the next two chapters, these questions will be explored in the context of designing an assessment tool for chemotherapy-induced peripheral neuropathy.

Chapter 4. Preferences for the assessment of chemotherapy induced peripheral neuropathy (CIPN): The patient perspective

4.1 Introduction

There is debate in the health literature about the appropriateness of using patients versus the general population in preference studies. This can be an issue if the preference studies inform decision-making processes such as values used in cost utility analyses (Ogorevc et al., 2019; Rand-Hendriksen et al., 2012; Shafrin et al., 2021). It has been argued that when it comes to decisions about public resource allocation a general population sample is most appropriate (European Network for Health Technology Assessment, 2015; NICE, 2013; Ogorevc et al., 2019). For instance, decisions about public health resource allocation does not affect just current patients but future patients as well i.e. general population. However, arguments have also been made that patient preferences should be considered as patients may be better able to imagine specific health states or outcomes due to their personal relevance or even experience with them (Najafzadeh et al., 2019).

An assumption underlying this debate is the existence of preference differences between patients and general population samples. The evidence for preference differences between patients and general population has been mixed. A meta-analysis by Dolders et al. (2006) found no difference in preferences. However, more recent evidence suggests there are preference differences. This includes another meta-analysis by Peeters and Stiggelbout (2010), who find evidence for patients providing higher valuations of health states using a range of stated preference techniques. In particular, there are findings in the stated preference literature that prior knowledge or experience can affect the relative weightings of different health states by respondents (Froberg & Kane, 1989; Mann et al., 2009). Within the DCE literature more specifically, studies report that patients showed much more concern for dimensions that may be less easy to imagine, such as anxiety/depression and pain/discomfort. In contrast, the general

population sample placed more importance on the ability to take care of themselves and mobility (Ogorevc et al., 2019).

It has also been argued that patients adapt to chronic health conditions and, as a consequence, will have different valuations for health outcomes (Loewenstein & Ubel, 2008; Ogorevc et al., 2019). For instance, in a DCE examining WTP for insurance coverage of novel lung cancer treatments, it was found that healthy respondents demonstrated a higher WTP compared to patients with lung cancer (Shafrin et al., 2021). This theory of adaption as an explanation for differences in patient and general population preferences was directly assessed in another DCE study. Although evidence for preference differences between population types was found, the evidence of it being due to adaptation was weak (Ogorevc et al., 2019).

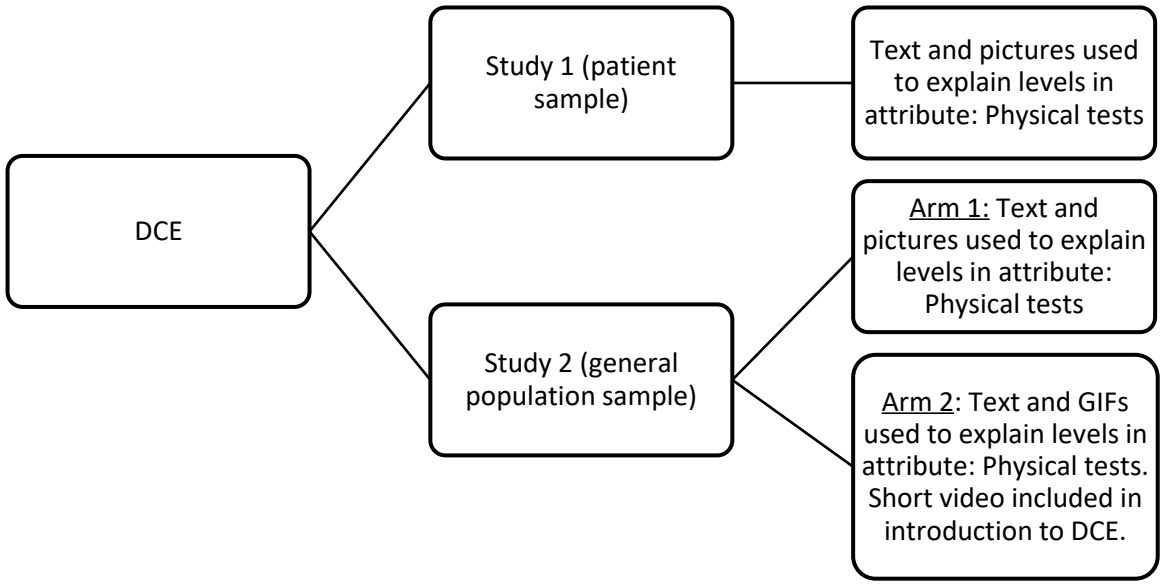
Overall, the available literature suggests there may be preference differences between different population types, including patient and general population samples. This is of particular concern if DCE researchers are interested in the preferences of a particular population sub-group to which they do not have access, or if the specified population sub-group is too small to allow for preference elicitation. For example, in the case of preferences for rare health outcomes the small number of patients who have experienced this outcome and are willing to participate in research may result in inadequate sample sizes for analysis. In addition, DCE researchers may want to explore preferences regarding a new assessment or treatment, in which case a relevant experienced patient sample will not exist. In these situations, there is a knowledge or experience gap for respondents that needs to be addressed in the overall design of the research.

Therefore, an important issue for investigation is whether it is possible to reduce any potential preference differences between patient and general population samples. Studies mentioned thus far have compared preferences from these two samples, but none have sought to intentionally manipulate what is seen by one sample versus another. As far as this author is aware this study is one of the first to investigate whether it is possible to reduce differences between a patient versus general population sample. In particular, can the provision of extra information to a general population reduce the knowledge/experience differences from a patient sample? And, more

importantly, can this thereby reduce any preference differences between these two populations that arise from direct experience or knowledge?

These questions will be explored in Chapters 4 and 5. This research project has been divided into two studies. Study 1, reported in Chapter 4, uses a patient sample and Study 2, reported in Chapter 5, uses a general population sample. The Study 2 sample is further divided into two arms. In Arm 1, respondents receive the same information given to the patient sample, while in Arm 2 respondents receive additional information to explain the attributes and their levels in more detail. This has been summarised in Figure 4.1.

Figure 4.1 Project outline



This chapter introduces the context in which this question will be investigated. The construction and design of the DCE are summarised and the results from Study 1, the patient sample, are then be reported and discussed. Chapter 5 will describe the results from Study 2, the general population sample, and will also compare and discuss the results from Study 1 and Study 2.

4.2 Background

Cancer continues to have an impact worldwide. According to the 2014 World Cancer Report, approximately 14 million people were newly diagnosed with cancer in 2012 (Stewart & Wild, 2014). Advances in the treatment of cancer have meant that there have

also been an increasing number of cancer survivors. There were an estimated 33 million people in 2012 still alive five years after an initial diagnosis of cancer[1] (Stewart & Wild, 2014), an increase from 28 million in 2008 (Boyle & Levin, 2008; Park et al., 2013).

As a result, while treatment of cancer continues to be an important area of research, there has been a movement to increase research into improving the quality of life of cancer survivors (McCrary et al., 2017; Park et al., 2013). One area that is receiving increasing attention is the impact of side effects from chemotherapy treatment on patient quality of life during and after chemotherapy treatment. One important side-effect that has been investigated is peripheral neuropathy caused by chemotherapy treatment, or chemotherapy-induced peripheral neuropathy (CIPN).

CIPN refers to damage to the peripheral nerves and can lead to impairment in the detection of touch, vibration and/or proprioception (awareness of position and movement of the body) (Park et al., 2013). Symptoms can include, but are not limited to, paraesthesia or 'pins and needles', numbness, loss of balance and/or difficulty rising from bed or from a seated position (Kolb et al., 2016; McCrary et al., 2017; Park et al., 2013).

CIPN may develop in cancer patients who are receiving treatment with certain chemotherapy agents, including taxanes, platinum compounds, vinca alkaloids, thalidomide and bortezomid (McCrary et al., 2017; Park et al., 2013; Staff et al., 2017). CIPN has been estimated to affect 30–40% of patients who are treated with chemotherapy agents, and in some cases CIPN symptoms have been reported to persist years after the treatment concludes (Argyriou et al., 2012; Kerckhove et al., 2017; Staff et al., 2017).

CIPN symptoms can have a significant negative impact on the quality of life of cancer patients and survivors (Winters-Stone et al., 2017). For example, if CIPN symptoms manifest as numbness in the hands, daily activities such as cooking or cleaning become difficult and even hazardous. Losing balance or falling can also be a real danger if a patient experiences symptoms as problems with balance or numbness in the feet. As a

[1] A more recent report is available however the type of statistics reported were different. Hence, the 2012 figures were reported.

result, assessing for CIPN while patients are undergoing chemotherapy treatment is important as part of survivorship planning.

However, cancer patients are not currently routinely assessed for CIPN, nor are there any guidelines for how cancer patients should be assessed for CIPN. It has been suggested that one reason for this is that there is currently no agreed 'gold' standard or 'best practice' CIPN assessment tool or package (McCrary et al., 2017; Park et al., 2013). A recent systematic review by McCrary et al. (2017) identifies 117 unique CIPN assessment tools available in the literature. The review included a two-stage Delphi survey conducted with a multidisciplinary group of experts with expertise relevant to the assessment of CIPN. This resulted in the identification of six assessments that were considered to be the 'best' based on different assessment criteria. However, no consensus was reached in the review with regards to one CIPN assessment that could be considered the 'gold standard'.

In the debate on the 'best' CIPN assessment, it is important to consider what is considered relevant and meaningful to both the clinician and the patient. Little attention has been given to understanding what patients consider as relevant and important when they are assessed for CIPN. The aim of this study is to address this gap. This chapter reports the findings from a cancer patient sample. In the next chapter, these results will be compared with results from a general population sample.

4.2.1 IN FOCUS Study

The author conducted this DCE as part of a larger project, the IN FOCUS Study (InFocus, 2019). The IN FOCUS Study is a research project with the goal of finding effective methods to assess and treat CIPN. This project has involved collaboration with a number of researchers, clinicians and hospitals across New South Wales, Australia. The IN FOCUS Study has sought to understand CIPN from various angles. This DCE fits into one of the aims, which is to improve the assessment of CIPN. Study 1 of this DCE, with the patient sample, will contribute by providing information about preferences for the assessment of CIPN from the patient perspective.

4.2.2 DCE theoretical framework

To the best of the author's knowledge, this is the first study to investigate patient preferences for the assessment of CIPN. The stated preference technique of a DCE was

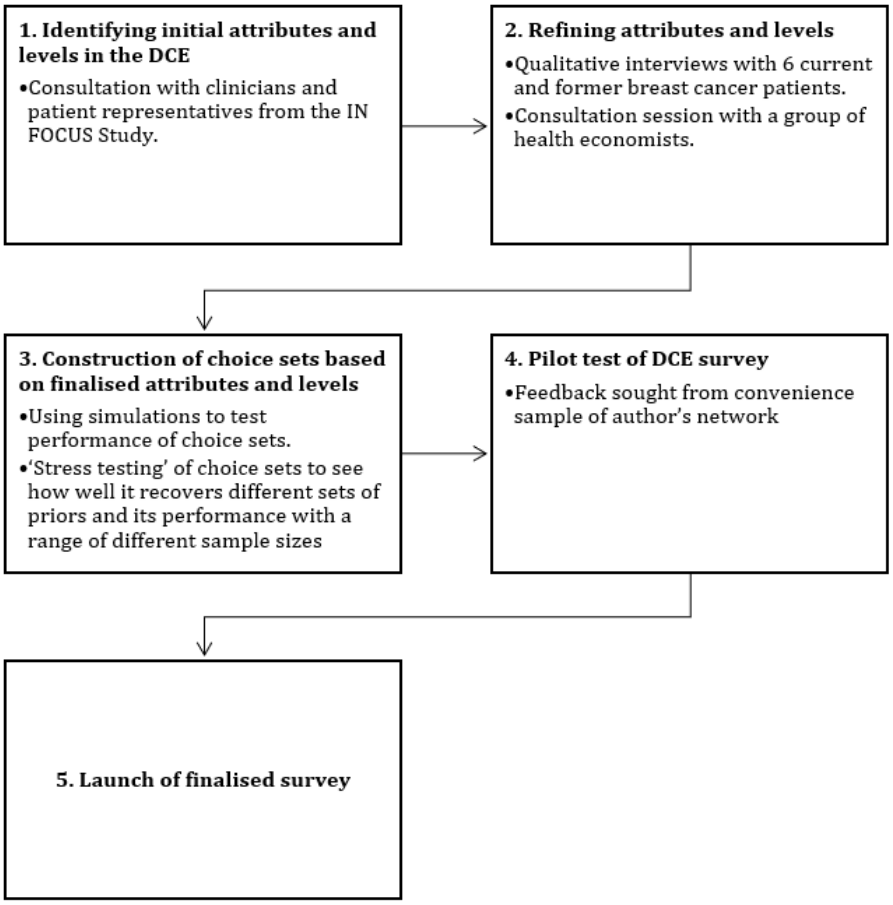
chosen as the method of investigation. The use of a DCE would allow the estimation of the value that patients place on various potential features of a CIPN assessment tool and the identification of features they consider to be important.

Within this DCE, respondents were asked to complete a series of choice sets. Within each choice set, respondents were asked to imagine they have been given a chemotherapy drug for which CIPN is a known side effect and that they will be assessed for CIPN while they undergo chemotherapy treatment. Respondents were then asked to choose between two options. Each option described a potential CIPN assessment tool made up of different features or attributes. These attributes could take on a number of different levels, which could differ or remain the same between the two options in each choice set. Preference information is generated from respondents picking an option within each of the series of choice sets. Implicit in this preference information is that when respondents choose one option over another, they have made a 'trade-off' between the different levels of each attribute between the two options in a choice set (Vass, Wright et al., 2018).

4.3 Methods

Details of the development process of the DCE are presented in Figure 4.2.

Figure 4.2 Developing the DCE flow chart



4.3.1 Initial attributes and levels

The initial identification of potential attributes and levels was based on the systematic review and Delphi survey by McCrary et al. (2017). This was further refined through consultation with and feedback from members of the IN FOCUS Study, including both clinicians and patient representatives. These initial attributes and levels appear in Table 4.1.

Table 4.1 CIPN DCE: initial attributes and levels

Attribute	Levels
Symptoms and Quality of Life	The assessment asks about your **symptoms** The assessment asks about how your symptoms impact on your **usual activities**
Level of Detail	The assessment detects small changes, defined as worsening within a particular grade The assessment detects large changes, defined as a movement to a more severe grading
Mode of administration	You fill in a survey You are asked questions &/or examined **during** your appointment

Attribute	Levels
	You do some physical tests **during** your appointment You conduct a range of assessments **by yourself** which may involve a combination of survey questions and physical tests
Frequency of administration	Once a week Once a fortnight Once a month Once every 3 months
Time for assessment	During usual clinic time Usual clinic time plus 10 minutes extra Usual clinic time plus 30 minutes extra You require a separate appointment, which can take up to 60 minutes
How will results influence care/treatment	The results may lead to **modifications** in your **treatment** The results may lead to **practical interventions** in your lifestyle The results may lead to modifications in your treatment **and** practical interventions in your lifestyle The results **will not change** your care/treatment, but could be used for research to help patients in the future

4.3.2 Refining attributes and levels: qualitative Interviews

To refine attributes and levels, qualitative interviews were conducted with current and former cancer patients. A specific method, termed cognitive interviewing or 'think aloud' interviewing, was used (Drennan, 2003; Willis, 1999). In the cognitive interview, the participant sits with the author in the room while completing the survey.

The survey included a draft DCE based on the initial attributes and levels shown above. Background information to explain the attributes and levels was included, as were questions related to cancer and general demographics. This is explained in more detail later in this chapter.

Participants were asked to verbalise their thinking process as they went through the survey. The purpose of these cognitive interviews was to gain an understanding of how participants process and interpret attributes and levels in the DCE and the survey more generally. Respondents were asked to comment on the included attributes and levels, including offering suggestions about improvements to wording or additional attributes or levels that were important. Feedback and suggestions were also sought about what attributes and levels to include, change or exclude. Cognitive interviews were recorded with the consent of participants.

Prior to conducting the cognitive interviews, the author undertook training with an experienced qualitative interviewer (Dr Alison Pearce). This included reviewing relevant literature and observing two sessions in which Dr Pearce conducted cognitive interviews on her own project (with the permission of the participants).

Ethics approval for the cognitive interviews was granted under the CHERE's Program Ethics Approval from the University of Technology Sydney (UTS) Human Research Ethics Committee (UTS HREC REF No. ETH18-2507). A copy is attached at Appendix 4A.

Sample for cognitive interviews

Participants were recruited from the Breast Cancer Network Australia (BCNA). The BCNA is Australia's leading organisation for women with breast cancer. Women associated with the BCNA can volunteer their time to participate in research through the BCNA Review and Survey Group. Participants for cognitive interviews were recruited from this group.

BCNA was chosen as its members provided a suitable sample. A disproportionate number of breast cancer patients develop CIPN symptoms, as they often receive taxane-based chemotherapy, including Docetaxel and Paclitaxel, which are commonly associated with the development of CIPN symptoms (Bandos et al., 2017; Bao et al., 2016). As a result, there was a high likelihood that participants recruited would have suffered from CIPN.

BCNA sent out a general email to its members requesting volunteers for this study, and interested participants were asked to contact the author of this thesis directly. Eligibility criteria for participants included having had a diagnosis of breast cancer and received therapeutic chemotherapy. The interviews were conducted in a suitable meeting room at the university research centre. Recruitment continued until February 2019 when the author judged that saturation had been reached (see the section below ' cognitive interview process' for further details). In total, six participants were interviewed. Three of the participants requested a summary of the results from the interviews. This was sent by the author in July 2019.

Creation of the DCE for cognitive interviews

Once the initial attributes and levels had been determined (see Table 4.1), the choice sets were constructed. Choice sets were designed to estimate main effects only and the

choice design chosen was based on D-optimality for the MNL model with zero priors. A generator-developed method was employed to create the choice sets (Street & Burgess, 2007). A 4^5 orthogonal array from the SAS website (Kuhfeld, 2006) was used as the starting design. The starting design was transformed from a 4^5 to 2^2 4^4 orthogonal array. Transformations to the starting design were based on principles from Chapter 2 of Street and Burgess (2007). Transformations and creation of choice sets were made using Mathematica based on code initially written by Burgess (2007) and subsequently augmented by Street (2019). From this, a total number of 16 choice sets of size two were created. To minimise respondent burden, participants were randomly assigned to complete eight choice sets from the total 16 choice sets.

Information provided to participants in cognitive interviews

Prior to being asked to complete the choice sets, participants were shown some information explaining key terms used in the attributes and levels. This information included examples of physical tests of CIPN, including text description and pictures of physical tests, and the respondents were told they could refer back to this information as needed.

Creation of the survey for cognitive interviews

The online survey to complete the DCE was created within the online platform provided by SurveyEngine (SurveyEngine GmbH, 2021). SurveyEngine is a survey creation and distribution company. An anonymous survey link was provided to all participants so that they could access and complete the survey.

Cognitive interview process

Cognitive interviews were conducted using an iterative process of improvement. The cognitive interview protocol for the first four participants can be found in Appendix 4B. The cognitive interview protocol for the last two participants was very similar except that more targeted questions were asked; these can be found in Appendix 4C. The participant was left to lead the 'conversation', with prompting from the author only when the participant was silent for a prolonged period. To minimise the influence of the author, if the participant asked a question they were usually advised to do what they would do if they were at home completing the survey. Some probing questions were used for specific situations. For example, where there was a long silence from the participant, the author asked questions such as 'what were you thinking just then?' If

the participant had difficulty choosing an option, they would be asked 'What is it about these options that makes it difficult for you to choose between them?' Details of other prompts can be found in Appendix 4D. More directed questions by the author were left until after the participant completed the survey.

Participants were asked to read through an information sheet prior to completing the survey. The survey included eight choice sets, followed by questions about their cancer and some socio-demographic questions. After the participant completed the survey, a discussion was had about the choice sets and the survey more generally. The participant was also shown and asked to comment on the full set of attributes and levels and the presentation format of the choice sets. Some participants were shown the physical tests without pictures in the information sheet. These participants were shown the pictures at the end and asked to comment on whether they would have been useful.

After the first three cognitive interviews, the survey was modified based on the feedback and suggestions received. The last three participants completed a modified survey. In addition to a modified survey, the last two participants also received more directed questioning as outlined in Appendix 4C. A summary of the feedback and modifications made to the survey can be found in Appendix 4E.

Summary of results from cognitive interviews

Changes were made in terms of the information sheet shown to participants prior to the choice sets. In particular, an explanation of CIPN was added and revisions around the explanation of the attribute 'Level of Detail'. Further information was provided about what was meant by 'small changes' or 'large changes' in CIPN. While modifications were made after the third cognitive interview, participants still reported difficulty understanding what these terms meant. The feedback on the use of pictures to explain physical tests showed that they were well received by participants, with most saying that it was helpful to have the pictures. The use of bolding and underlining of key terms was also viewed as helpful by most participants.

The attribute 'Time for Assessment' was not considered to be important by the majority of participants. The attribute 'Frequency of Administration' caused some confusion in the first three interviews. Participants commented that the desired frequency of being assessed for CIPN would depend on whether it was during chemotherapy treatment or

post-treatment. If it was post-treatment, the desired frequency would be less. It was decided to set the context as during chemotherapy treatment and this appeared to clarify the issue for the final three interviews. All participants drew attention to the last attribute, 'How will results influence care/treatment?'. Participants were particularly concerned with the level 'The results will not change your care/treatment, but could be used for research to help patients in the future'. There was confusion about the reason for doing the CIPN assessment if it did not actually contribute to their general care or help their treatment. Participants were positive about contributing to research to help other cancer patients but, as several put it, 'What about me?'

In the last two interviews more targeted questions were asked while the participants were completing the choice sets. When these participants were prompted about their interpretation of what 'modifications in their treatment meant', interpretations were consistent with the intention of the author. Specifically, one participant interpreted this to mean a reduction in dosage, while the other interpreted it as a change to a different chemotherapy agent. It is important to note that based on other feedback, these two participants viewed modifications to treatment positively rather than negatively. This was supported by unprompted comments made by earlier participants. This is in contrast to initial discussions with clinicians and patient representatives, where the assumption was that any modification to treatment was likely to be viewed quite negatively by patients.

4.3.3 Refining attributes and levels: consultation with health economists and DCE experts

Introduction

In addition to the cognitive interviews with patients, a consultation session was also conducted by the author with a number of health economists and DCE experts. The purpose of this session was to receive feedback and comments on specific issues relating to the attributes and levels from a group who were familiar with DCEs.

The author was the main facilitator and a dedicated note-taker was also present to capture any comments; the author also took notes. The session started with a brief presentation in which the project was introduced in the wider context of the IN FOCUS Study. The presentation also acted as a guide to discussion with the information sheet

presented in stages followed by the attributes and levels presented in stages. This ensured that comments were focused and that each section of the DCE was covered during the session. A summary of the notes taken during the discussion session can be found in Appendix 4F.

Key takeaways from consultation session

The participants proposed the need for a choice scenario or vignette in order to put the choice sets in context. They also suggested revision of the attribute 'Mode of Administration'. 'Physical test' was noted to be a very vague term. In addition, questions were raised about whether it was realistic that a CIPN assessment would only require a patient to do a survey or a type of physical test. It was suggested that combination of the two would be better. The attribute 'Frequency of administration' was also discussed in terms of the suitability of the range of levels chosen, and the evidence base for the levels chosen. Discussion with some members of the IN FOCUS Study also indicated that this attribute was not of importance from a clinical perspective. Instead, the other time-related attribute, 'Impact on clinic time', was considered of greater importance to clinicians.

4.3.4 Finalised attributes and levels

Based on the feedback received from the cognitive interviews and the discussion session, the attributes and levels were further revised substantially and a vignette was also created (see Tables 4.2 and 4.3 below).

For the attribute 'Level of Detail', the description of what was meant by small and large changes was modified. The wording that was finally adopted was based on suggested wording by a patient representative from the IN FOCUS Study, on the basis that this was clearer and simpler than the wording used for the cognitive interviews. 'Mode of Administration' was split into two attributes, 'Questionnaire' and 'Physical Test/s'. Choice sets now described a combination of survey questions and physical tests. This ensured a more realistic reflection of what could potentially be involved in a screening tool as opposed to only completing a survey or only completing a physical test.

The attribute 'Frequency of administration' was removed as it was not considered to be of sufficient value to justify the extra information respondents would have to process. The levels in the last attribute, 'How will results influence care/treatment', were

modified substantially. The final wording was based on a suggestion by a patient representative from the IN FOCUS Study, as it was considered to better reflect the intentions of this attribute. In particular, there is now a focus on the doctor being able to make the decision of the impact of results versus the doctor and the patient making the decision together.

4.3.5 Construction of choice sets

Once the final set of attributes and levels were decided, choice sets were also finalised. Choice sets were constructed in Mathematica based on code initially written by Burgess (2007) and subsequently augmented by Street (2019).

A 4^5 orthogonal array from the SAS website (Kuhfeld, 2006) was used after one 4-level attribute was converted to two binary attributes using expansive replacement and another 4-level factor was converted to a 3-level factor by collapsing one of the levels (Street & Burgess, 2007). To ensure that the two binary attributes would not dominate the choice made by respondents, each of these was presented at the same level in both options in one half of the choice sets. To do this, two generators were chosen, one with a 0 corresponding to the first binary attribute and the other with a 0 corresponding to the second binary attribute. This resulted in 32 choice sets.

To ensure realism of the choice sets, a restriction was also imposed so that the levels 'no questionnaire' and 'no physical test/s' never appeared together. The final set of choice sets was 87.2% efficient relative to the best possible using the D-optimality criterion for the MNL model assuming zero priors.

Table 4.2 Project 1: vignette for choice sets

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some tests during your appointment. If these were your only options, which assessment would you prefer? A or B?

Table 4.3 Project 1: finalised attributes and levels

Attribute	Levels
Symptoms and Usual Activities	The assessment asks about your symptoms The assessment asks about how your symptoms impact on your usual activities
Level of Detail	The assessment will pick up minor nerve damage and small changes in your condition, whether it is important or not The assessment will only pick up major nerve damage and large changes in your condition
Questionnaire	3 questions to answer 12 questions to answer 20 questions to answer No questionnaire
Physical Test/s	Clinician administered test e.g. sharp and dull test, tuning fork test Patient activity based test e.g. peg board test, sway test Technical test e.g. nerve conduction studies No physical test
Impact on Clinic Time	During usual clinic time Usual clinic time plus 10 minutes extra Usual clinic time plus 30 minutes extra You require a separate appointment, which can take up to 60 minutes
How will results influence care/treatment	The doctor will discuss the results with you, and together you can decide what they mean for you and your care/treatment The doctor may change your chemotherapy/cancer treatment if there are significant changes in your condition over time The doctor may change your general care (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time

The 32 choice sets were then divided into groups of eight choice sets, henceforth referred to as blocks. Choice sets were assigned with the objective of level balance across the four blocks (Huber & Zwerina, 1996; Street & Burgess, 2007). In addition, choice sets in each block were assigned such that each of the binary attributes was presented at the same level in one half the choice sets. Each respondent was randomly assigned to one of the four blocks.

4.3.6 Using simulations to assess the performance of choice sets constructed

Purpose of simulations

Once the finalised choice sets were constructed, simulations were conducted to test how well the chosen design recovers assumed sets of priors. In addition, the simulations provide information about how well the chosen design performs for different sample sizes. These simulations are based on the multinomial logit model, similar to the process in Street et al. (2017). The statistics program RStudio (RStudio Team, 2020) was used to carry out the simulations. Code for the simulations was written in R by Street (2019).

Determining sets of priors

Four sets of priors were tested, including a set of zero priors and three sets of priors based on 'logical' reasoning. The priors and the logic for the positive or negative sign for each entry are given in Appendix 4G.

Determining number of responses for simulations

The total number of responses in the simulations was varied among 800, 1600, 3200 and 6400. The range of total responses chosen for the simulations was based on the potential pool of respondents, given that there were around 1000 potential respondents. For the actual DCE, it was planned that each respondent would be randomly assigned to complete a block of eight choice sets from the total 32 choice sets. A range of response numbers were tested as it was not certain what the response rate would be. At the conservative end, the author tested for a minimum of 800 responses or, equivalently, 100 respondents. Based on the optimistic assumption of an 80% participation rate, this corresponded to 800 respondents, giving a maximum of 6400 responses.

Table 4.4 Summary of respondent numbers used

Number of responses/ choice set	Number of subjects required if only competing 8 choice sets each	Total number of responses over the 32 choice sets
25	100	800
50	200	1600
100	400	3200
200	800	6400

Assessment of simulations

Based on a review by Burton et al. (2006), 1000 simulations was the most frequent number of simulations used in the literature. Simulations were performed in RStudio (RStudio Team, 2020) for each of the total number of responses using 1000 simulated samples for each. Results for the simulations were very similar across all of the total number of responses tested. Boxplots of the parameter values and their standard errors for the case of 800 responses are shown in Figure 4.3.

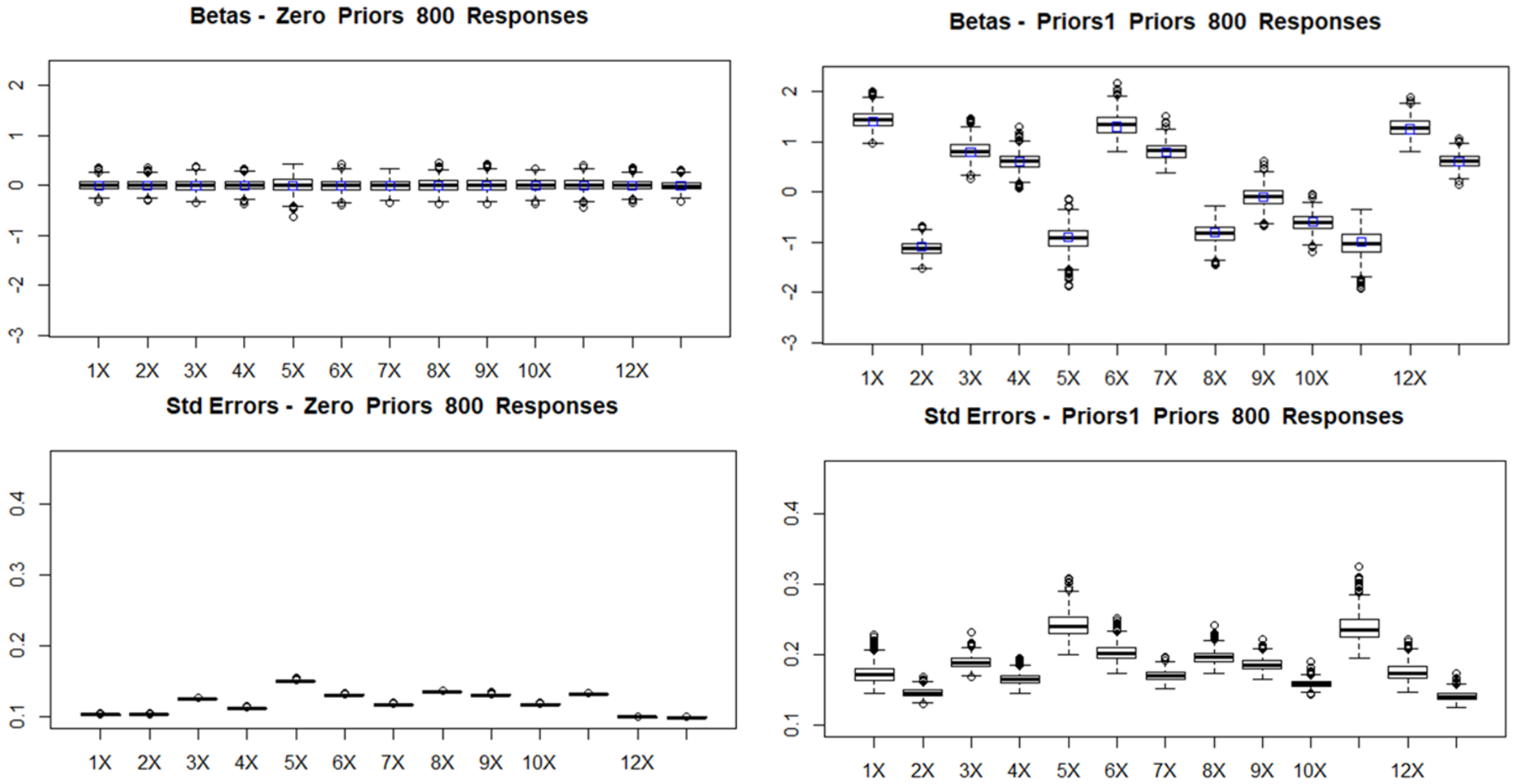
These simulations were then assessed using performance measures recommended by Burton et al. (2006). In particular, simulations were assessed based on standardised bias and coverage (the proportion of times the confidence intervals contains the assumed priors).

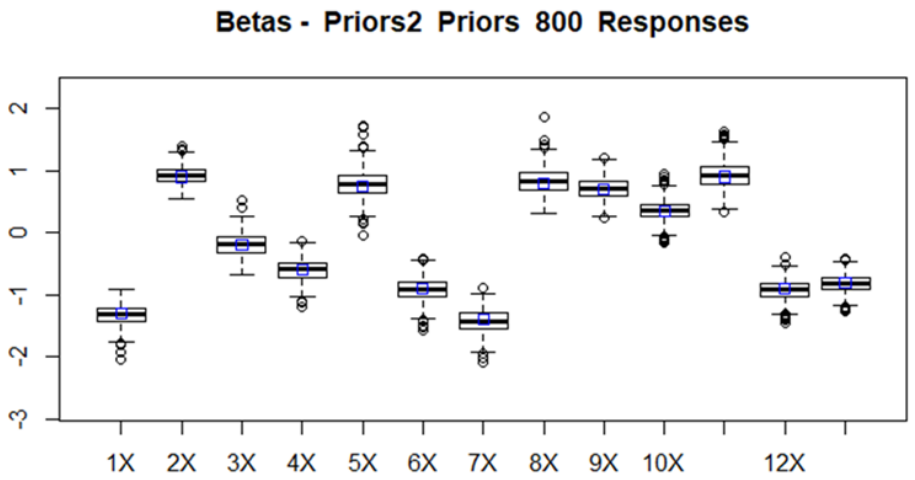
Standardised bias

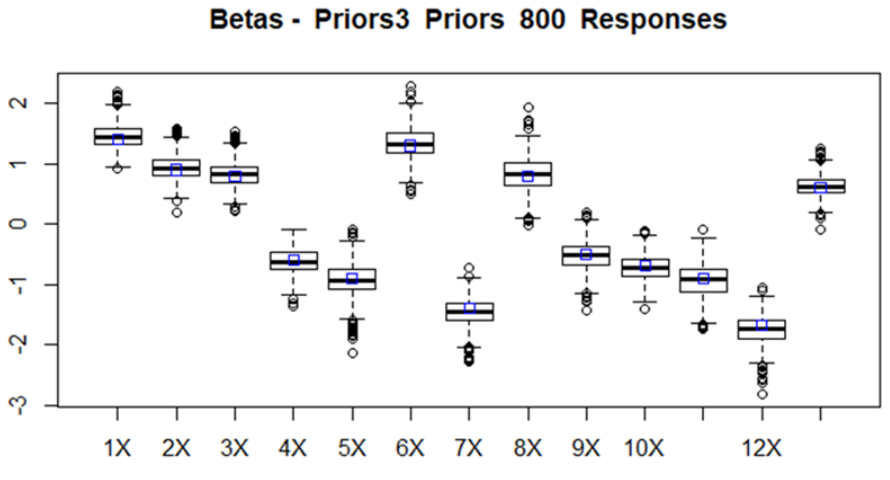
Standardised bias was calculated using both the mean and median of the parameter values and their associated standard errors. Table 4.5 provides a summary of the minimum and maximum values observed in the simulations. Standardised bias should be below the absolute value of 40% in order to avoid adverse impacts on standard errors, coverage and accuracy (Burton et al., 2006; Collins et al., 2001). It was noted that across the range of responses tested and the four sets of priors used, standardised bias was consistently below the absolute value of 40%. In general, as the number of responses increased, the range of standardised bias values became narrower and the values themselves became smaller.

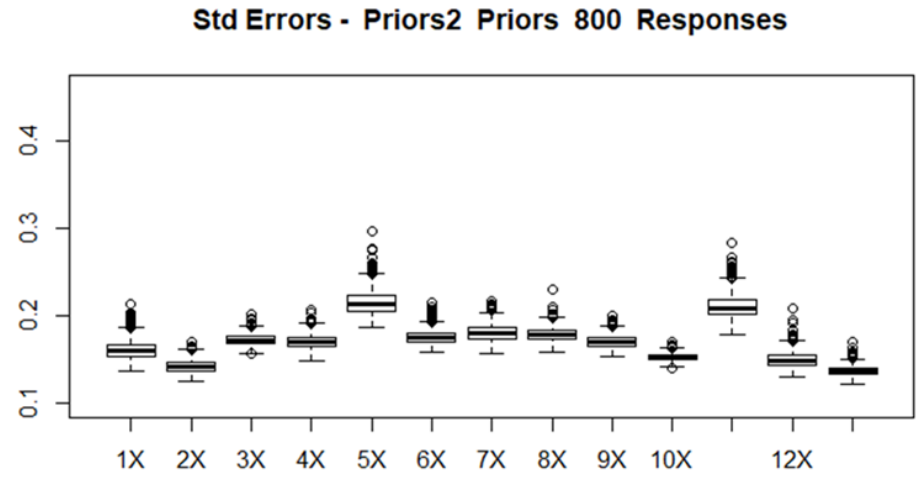
Figure 4.3 Boxplots of betas and standard errors

1000 Simulations assuming 25 responses per choice set, 800 responses in total




Betas - Priors2 Priors 800 Responses
2
1
0
-1
-2
-3
1X
2X
3X
4X
5X
6X
7X
8X
9X
10X
12X


Betas - Priors3 Priors 800 Responses
2
1
0
-1
-2
-3
1X
2X
3X
4X
5X
6X
7X
8X
9X
10X
12X


Std Errors - Priors2 Priors 800 Responses
0.4
0.3
0.2
0.1
1X
2X
3X
4X
5X
6X
7X
8X
9X
10X
12X

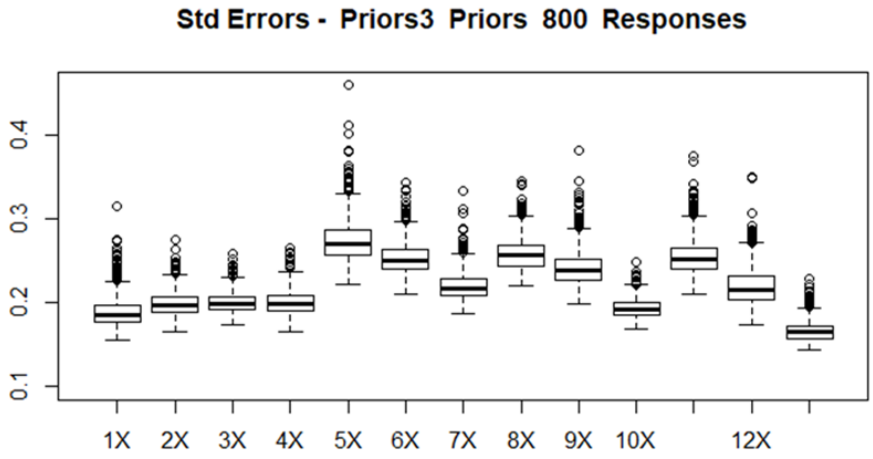

Std Errors - Priors3 Priors 800 Responses
0.4
0.3
0.2
0.1
1X
2X
3X
4X
5X
6X
7X
8X
9X
10X
12X

Table 4.5 Standardised bias summary

1000 Simulations assuming:		25 responses per choice set, 800 responses in total		50 responses per choice set, 1600 responses in total		100 responses per choice set, 3200 responses in total		200 responses per choice set, 6400 responses in total	
Standardised Bias calculated with:		Min	Max	Min	Max	Min	Max	Min	Max
Zero Priors	Mean	−5%	5%	−2%	3%	−7%	6%	−5%	6%
	Median	−10%	9%	−4%	5%	−10%	6%	−9%	9%
Priors1	Mean	−25%	25%	−13%	16%	−11%	17%	−9%	8%
	Median	−24%	23%	−8%	10%	−11%	15%	−9%	9%
Priors2	Mean	−22%	18%	−20%	15%	−12%	13%	−9%	8%
	Median	−16%	16%	−19%	16%	−11%	11%	−8%	10%
Priors3	Mean	−34%	31%	−20%	20%	−19%	20%	−10%	13%
	Median	−27%	28%	−23%	19%	−20%	18%	−11%	18%

Coverage

Coverage was assessed according to the criteria suggested by Burton et al. (2006) – namely, coverage is considered acceptable only if it does not fall outside an interval of approximately two standard errors from the nominal coverage probability. In this case, a nominal coverage probability of 95% was used. As such, coverage was judged to be acceptable if it fell between 93.6% and 96.4%.

In general, coverage was considered acceptable across the range of total responses tested and across the four sets of priors, with the exception of prior set 1, where it was noted that parameter 1 had a coverage of 97% for 1600 responses, and prior set 2, where parameter 3 was noted to have a coverage of 97% for 3200 responses. Over-coverage in these instances is a potential indication of a higher than expected type II error (Burton et al., 2006).

4.3.7 Study 1: patient Sample

Ethics approval

A separate full ethics application was submitted for the final survey. Ethics approval was granted on 9 of September 2019 under UTS HREC REF NO. ETH19-3464 (Appendix 4H).

Piloting

Prior to the launch of the survey it was informally reviewed by a convenience sample drawn from the author's networks, including colleagues with experience in DCEs. This was done to obtain feedback and comments on wording and flow of the survey and to check for any grammatical/spelling errors.

Survey flow

The final survey consisted of four main sections. When respondents clicked on the online survey link, they were taken to an information page explaining the study. The survey link automatically generated a pseudo IP address, meaning surveys cannot be linked back to any individual, ensuring the anonymity of respondents. Respondents were told that continuing to the rest of the survey would be taken as consent to be part of the current study. Respondents were then filtered through two screening questions to determine eligibility. In order to participate in the survey, respondents had to have been diagnosed with cancer and have had chemotherapy as a treatment for their cancer.

Eligible respondents were then asked to read through some background information that explained peripheral neuropathy (see Figure 4.4). The background information also introduced attributes for consideration in the choice set (Figure 5 and Figure 6). For the attribute 'Types of physical tests', text and pictures were used to explain the levels (see Figures 4.7 to 4.10). An instructions page describing the choice sets was also included. Respondents were then randomly assigned to a block of eight choice sets to complete. An example choice set, with the choice scenario above, appears in Figure 4.11.

Figure 4.4 Screenshot: explanation of peripheral neuropathy

On this page and the next few pages, some background information is provided to define what is meant by peripheral neuropathy and to explain some key definitions and terms.

Please read through them.

Please note: the following information *is not* intended to reflect your personal circumstances. They *are not* intended to have relevance for your personal decisions

What is peripheral neuropathy and how is it related to chemotherapy?

Peripheral neuropathy is a major side effect experienced as a result of chemotherapy treatment, affecting up to 40% of cancer survivors.

Peripheral neuropathy occurs when drugs used to treat cancer cause damage to the peripheral nerves (i.e. nerves in the hands and feet, sometimes extending to the arms and legs).

Some common symptoms of peripheral neuropathy can include:
- Numbness
- Tingling, 'pins and needles' or electric shock-like sensations
- Burning sensations
- Balance problems
- Muscle weakness
- Constipation
- Decreased reflexes

These symptoms can lead to problems with completing everyday activities, such as:
- Trouble using your fingers to pick up or hold things; dropping things
- Trouble with buttoning clothes
- Tripping or stumbling while walking

Figure 4.5 Screenshot: information about attributes Part 1

One of the ways in which a peripheral neuropathy assessment can be different, is in terms of the level of detail of the assessment.

What do we mean by this?

The peripheral neuropathy assessment could pick up:

- *minor and major* nerve damage, including *small changes* in your condition, whether it is important or not OR
- only *major* nerve damage and *large changes* in your condition

What is the difference between an assessment that picks up small versus a large change in peripheral neuropathy?

Potential Symptom	Examples of Assessment Detecting Small Change	Examples of Assessment Detecting Large Change
Numbness in hands	The assessment detects when the numbness becomes worse, although you are still able to perform the same tasks e.g. using knives or dressing yourself.	The assessment detects when the numbness that did not interfere with daily activities worsens such that chopping vegetables and carrying pots becomes difficult.
Numbness in feet	The assessment detects when the numbness becomes worse, but it does not affect your ability to walk or maintain balance.	The assessment detects when the numbness in the feet progresses to being unsteady on your feet, especially at night.
Pain/tingling in hands or feet	The assessment detects when increasing pain or tingling occurs.	The assessment detects when pain or tingling that makes cooking difficult worsens to being unable to button clothing or use keys.

Figure 4.6 Screenshot: information about attributes: Part 2

Other ways in which a peripheral neuropathy assessment can be different include:

- Whether the assessment focuses on symptoms or how symptoms impact on usual activities
- The length of questionnaire that the patient may have to complete
- The impact of assessment on clinic time
- How will results from the assessment influence care/treatment

Figure 4.7 Screenshot: types of physical tests Part 1

Another way in which a peripheral neuropathy assessment can be different is in terms of the type of physical tests involved.

What do we mean by this?

The peripheral neuropathy assessment may involve different types of physical tests on your body.

There are three main types of physical tests:

- Clinician administered tests
- Patient activity based tests
- Technical tests

Examples of clinician administered tests

Sharp and dull test
This is a pain perception test where you will be asked to perceive the difference between a sharp or dull stimulus.
An example of a sharp and dull test is given below.

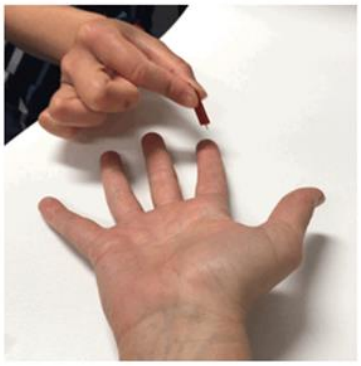


Figure 4.8 Screenshot: types of physical tests Part 2

Tuning fork test

This is a vibration sensation test. A tuning fork is placed against different parts of your body and you are asked to say when you feel the vibration stop.
An example of a tuning fork test is given below.

Examples of patient activity based tests

Peg Board Test

This is to test the manual dexterity of the hands.
An example of a peg board test is the grooved pegboard test, which consists of holes that are randomly positioned.
The pegs need to be rotated to match the hole before they can be inserted.
An example of a grooved pegboard test is given below.

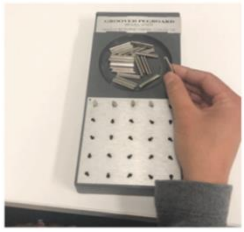


Figure 4.9 Screenshot: types of physical tests Part 3

Sway Test

An example of a sway test is the Romberg's test, which is a test of the body's sense of positioning.
You will be asked to close your eyes and you will be assessed on your sense of balance.
An example of a sway test is given below.

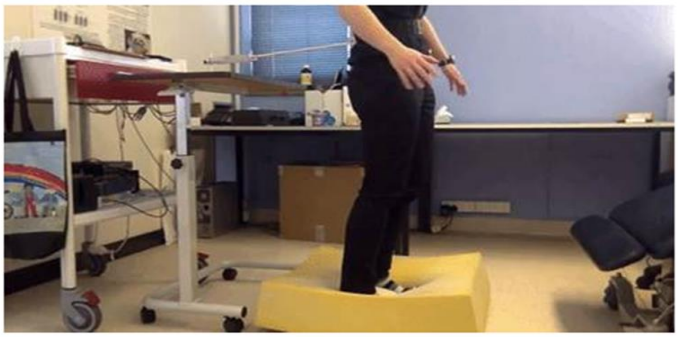


Figure 4.10 Screenshot: types of physical tests Part 4

Example of a technical test

Nerve conduction studies

These tests record the properties of electrical impulses travelling along your nerves using stick-on-electrodes and impulses which feel a little like tapping. An example of a nerve conduction study is given below.

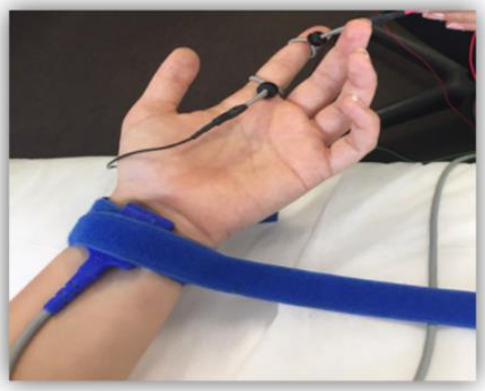


Figure 4.11 Example choice set

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about your *symptoms*
Level of detail	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	12 questions to answer	20 questions to answer
Physical test/s	No physical test	Patient activity based test e.g. peg board test, sway test
Impact on clinic time	Usual clinic time plus 30 minutes extra	Usual clinic time plus 10 minutes extra
How will results influence care/treatment	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

After completing the assigned block of eight choice sets, respondents were asked some evaluation questions in relation to the choice sets. The final section consisted of questions related to their cancer and general socio-demographic characteristics. A full copy of the survey is provided at Appendix 4I.

Sample

The pool of potential respondents consisted of current or former cancer patients who had participated previously in research by the IN FOCUS Study. From this pool of over 1000 respondents, approximately 80–90% had agreed to be contacted for future research. The link to the survey was distributed through a newsletter type email sent by IN FOCUS (see Appendix 4J for a copy). Interested respondents could then click on the link to participate and access the survey.

Survey launch and issues

The survey was launched on 9 October 2019 and closed on 11 December 2019.

A problem was noticed after the first few days of the initial launch. Specifically, an unusually high number of system-initiated time-outs were recorded on the survey platform, SurveyEngine (SurveyEngine GmbH, 2021), for the current study. It appeared that these respondents started the survey but then left the survey inactive for 30 minutes or more.

Test runs were performed on the link sent out. It was noticed that there was a problem with the display of the survey that was not detected during previous test runs prior to launch. There was a problem with the bottom of the survey overlapping with the buttons to press 'previous' and 'next'. On devices with smaller screens this made it extremely difficult to press 'next' when completing the choice sets, as the option for Assessment B almost completely overlapped with the 'next 'button.

After this issue was detected, an email was sent to the newsletter editor at IN FOCUS, who was in charge of distributing the survey link, in order to put the survey on hold. Data gathered prior to the detection of this issue were analysed to detect any potential bias towards Assessment A being chosen. A total of 49 respondents had started the survey, with 39 completing all assigned choice sets. Of those who did not complete all choice sets, the majority (5/9) only completed one choice set before discontinuing.

A chi-squared test was performed on the responses by respondents who had completed all assigned choice sets. It would be expected that there would a higher frequency of Assessment A being chosen if the display had, in fact, been an issue. The chi-squared test was not significant ($p = 0.42$), suggesting that the display issue had not affected results.

The survey was relaunched on 1 November 2019. An email was sent out by IN FOCUS to newsletter members (a copy of the wording used in the email can be found at Appendix 4K). The relaunch was positively received, with 65 completed responses after only four days. The survey was set up so the same IP address could not access the survey more than once. If the same IP address was detected, these respondents were screened out.

Data analysis

Models were estimated in R Studio (RStudio Team, 2020) using the gmnl package (Sarrias & Daziano, 2017). Initially, a MNL model was estimated. MNL parameter estimates were used to calculate the predicted choice probabilities. Next, several variations of the MXL model were estimated until the model of best fit to the data was found. Heterogeneous preferences identified in the MXL model via significant standard deviations ($p < 0.05$) were examined further via choice probability distributions and by kernel density plots. The kernel density plots were created in R Studio (RStudio Team, 2020) using R base graphics. The choice probability distributions were simulated in Mathematica (Wolfram Research, 2016) with 10,000 draws using custom code written by Professor Deborah J. Street (Street, 2021).

4.4 Study 1: Patient sample results

4.4.1 General descriptive statistics of respondents

In total, 117 respondents completed the whole survey. Of the 131 respondents who started the survey, two were screened out as they either have never been diagnosed with cancer or have never received chemotherapy as a treatment for cancer. 12 discontinued the survey at some point before completion and were also excluded from data analysis.

Table 4.6 provides a summary of the education level of respondents. Almost half of the respondents had completed a university degree, and one-third had completed TAFE or an apprenticeship.

Table 4.6 Education level of respondents

Education level	No. (%)
No school certificate/ other qualifications	2 (2%)
Secondary school	19 (16%)
Trade or apprenticeship	5 (4%)
TAFE or vocational college	35 (30%)
Bachelor's degree	33 (28%)
Postgraduate degree	23 (20%)

Each respondent was randomly allocated to one of four blocks of eight choice sets. Table 4.7 provides a summary of the assignment to blocks. There were close to 30 respondents per choice set, higher than the lowest simulation estimate of 25 respondents per choice set. Notably, having close to 30 respondents per choice set will lead to smaller standard errors (Burgess et al., 2011).

Table 4.7 Summary of assignment to block

Block	No. (%)
1	28 (24%)
2	32 (27%)
3	30 (26%)
4	27 (23%)

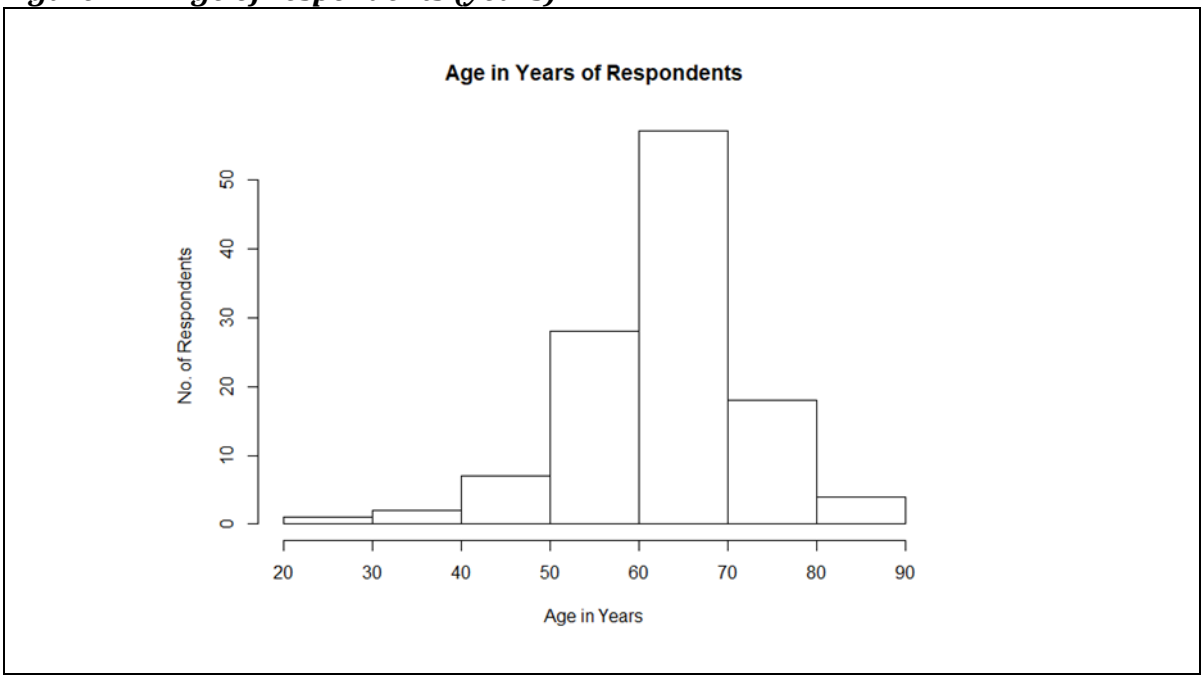
A high proportion of respondents were female, with 78% (91) of respondents identifying as such. The skew towards females was also reflected in the proportion of females in each block, as summarised in Table 4.8.

Table 4.8 Number of females and males by assigned block

Block	Females (%)	Males (%)	Total (%)
1	22 (79%)	6 (21%)	28 (24%)
2	24 (75%)	8 (25%)	32 (27%)
3	23 (77%)	7 (23%)	30 (26%)
4	22 (81%)	5 (19%)	27 (23%)
Total	91 (78%)	26 (22%)	117 (100%)

The median age of respondents was 64, although respondent ages did vary widely (27–87 years). A graphical summary is provided in Figure 4.12.

Figure 4.12 Age of respondents (years)



The mean and median time taken to complete the survey in minutes by block was also examined and is summarised in Table 4.9. Overall, the median completion time for the whole survey was 17.5 minutes. Respondents assigned to block 3 had the lowest mean and median completion time. A one-way ANOVA test was used to test whether mean completion time was different between assigned blocks. However, no significant difference was found ($p = 0.45$).

Table 4.9 Mean and median completion times by block to the nearest minute (mins)

Block	Mean (mins)	Median (mins)
1	23	17
2	22	20
3	18	15
4	22	18

Cancer-related descriptive statistics

Respondents had a varying range of experience with cancer. For the majority of respondents, it had been 10 years or less since they were first diagnosed with cancer. This has been summarised in Table 4.10 below.

Table 4.10 Years since first cancer diagnosis

Years	No. (%)
≤ 5	46 (39%)
6–10	45 (38%)
>10	26 (22%)

Among respondents, diagnoses of breast cancer, bowel cancer or myeloma were the most common. However, as can be seen in Table 4.11, there were multiple cancer types that respondents were diagnosed with.

Table 4.11 Cancer type (can select more than one)

Cancer Type	No.
Breast cancer	56
Bowel cancer	28
Myeloma	14
Other type of cancer (unspecified)	10
Ovarian cancer	9
Lymphoma	6
Leukaemia	3
Skin cancer	3
Head and neck cancer	2
Kidney cancer	2
Liver cancer	2
Thyroid cancer	2
Bladder cancer	1
Brain cancer	1
Lung cancer	1
Pancreatic cancer	1
Uterine cancer	1
Total	142

At the time of the survey, 15% of respondents (18/117) were receiving chemotherapy treatment. Among those who had received chemotherapy treatment in the past, 57% (56/99) had completed their treatment within the last five years.

Respondents were also asked about the type of chemotherapy drugs they received. The list of included drugs for selection were all commonly used drugs that have been associated with CIPN (Staff et al., 2017). This is summarised in Table 4.12.

The majority of respondents selected at least one drug that was commonly associated with CIPN. There were only six respondents who stated that they hadn't received any of the listed drugs.

Respondents were also asked if they had ever been assessed for CIPN. Some 39% (46/117) of respondents stated they had been assessed for CIPN previously, which suggests that a large minority of respondents have had at least some knowledge or experience related to CIPN. The author, in consultation with clinicians from the IN FOCUS Study, made the decision to not include a question about whether respondents had CIPN. In particular, this reflected an assessment that not enough information could be elicited from responses to one or two survey questions to determine whether a respondent actually had CIPN.

Table 4.12 Type of chemotherapy drugs received (can select more than one)

Type of chemo used	No.
Docetaxel (Taxotere, Dotax, Oncotaxel)	28
Oxaliplatin (Eloxatin, Oxalatin, Oxaliccord, Xalox, FOLFOX, XELOX)	26
Paclitaxel (Taxol, Anzatax, Plaxel, Abraxane)	24
Carboplatin (Carbaccord)	12
Bortezomib (Velcade)	10
Cisplatin (cisplatinum, Platinol)	9
Thalidomide (Thalomid)	9
Lenalidomide (Revlimid)	9
Pomalidomide (Pomalyst)	6
Vincristine	3
Vinblastine	2
None of above	6
I don't know	22
Total	166

Summary of attributes and levels with abbreviations

Table 4.13 provides a summary of attributes and levels in the order used for analysis. The first level listed in each attribute was used as the reference level during analysis – that is, it is the omitted level.

Table 4.13 List of attributes with abbreviations

Attribute		Levels
Sy&UA	Symptoms and Usual Activities	The assessment asks about your symptoms The assessment asks about how your symptoms impact on your usual activities
Det	Level of Detail	The assessment will only pick up major nerve damage and large changes in your condition The assessment will pick up minor and major nerve damage, including small changes in your condition whether it is important or not
Q	Questionnaire	No questionnaire 3 questions to answer 12 questions to answer 20 questions to answer
PhyT	Physical Test/s	No physical test Clinician administered test e.g. sharp and dull test, tuning fork test Patient activity based test e.g. peg board test, sway test Technical test e.g. nerve conduction studies
CT	Impact on Clinic Time	During usual clinic time Usual clinic time plus 10 minutes extra Usual clinic time plus 30 minutes extra You require a separate appointment, which can take up to 60 minutes
Res	How will results influence care/treatment	The doctor will discuss the results with you, and together you can decide what they mean for you and your care/treatment. The doctor may change your general care (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time. The doctor may change your chemotherapy/cancer treatment if there are significant changes in your condition over time.

4.4.2 MNL model results

Initially, a MNL model was estimated, the omitted level being the base for each attribute. The results are summarised in Table 4.14.

Table 4.14 Summary of MNL model results

Log Likelihood	−558
Obs	936
Iterations	4
AIC	1142
BIC	1205

MNL model results		Estimate	S.E.	P-value
S&Q 2	Symptoms & usual activities	0.206	0.1	0.039*
Det 2	Minor and major changes	0.853	0.11	0.000***
Q 2	3 questions to answer	0.368	0.15	0.011*
Q 3	12 questions to answer	0.403	0.14	0.004**
Q 4	20 questions to answer	0.369	0.15	0.015*
PhyT 2	Clinician-administered test	0.758	0.14	0.000***
PhyT 3	Patient activity-based test	0.911	0.13	0.000***
PhyT 4	Technical test	0.679	0.14	0.000***
CT 2	Usual clinic time + 10 mins	−0.087	0.13	0.802
CT 3	Usual clinic time + 30 mins	−0.479	0.13	0.369
CT 4	Separate appointment, takes up to 60 mins	−0.600	0.13	0.084
Res 2	Doctor may change your general care	−0.480	0.1	0.000***
Res 3	Doctor may change your chemo/cancer treatment	−0.600	0.1	0.000***

*** p < 0.001; **p < 0.01; *p < 0.05

Based on results of the MNL model, which assumes preferences are homogeneous across the sample, most attributes were significant. The results indicate that respondents preferred an assessment that asks about the symptoms impact on usual activities. Respondents also preferred that the assessment pick up minor as well as major nerve damage, whether it is important or not. In general, respondents liked having a questionnaire and a physical test as opposed to not having one. Respondents also significantly preferred an assessment where the patient makes a decision together with their doctor about how it impacts their care or treatment, as opposed to the doctor making the decision for them. Respondents appeared to be indifferent to the length of time needed for the assessment and whether or not it would require a separate appointment.

MNL results were examined further by looking at choice probabilities based on MNL parameter estimates; these are summarised in Table 4.15. As the levels were dummy coded, the parameter estimate for the reference level in each attribute was set at zero. This means that the probability of the reference level being chosen in each attribute is 1 minus the sum of the probabilities of the other predicted probabilities within it. For the attribute *Level of Detail* (Det), holding all other attributes constant, there was a 70% probability that respondents would choose an assessment that picks up both minor and major changes rather than one that detects major changes only. In contrast, examining Impact on *Clinic Time* (CT), holding all other attributes constant, the probability of any level being chosen was relatively similar across the four levels. This reflects the non-significant parameter estimates related to *Impact on Clinic Time* in the MNL model.

Table 4.15 Predicted choice probabilities (MNL)

Attributes and Levels		Pred. Prob.
Symptoms and Usual Activities		
Sy&UA base	Symptoms only	0.448
Sy&UA 2	Symptoms impact on usual activities	0.552
Level of Detail		
Det base	Major changes only	0.298
Det 2	Minor and major changes	0.702
Questionnaire		
Q base	No questionnaire	0.185
Q 2	3 questions to answer	0.27
Q 3	12 questions to answer	0.277
Q 4	20 questions to answer	0.268
Physical Test/s		
PhyT base	No physical test	0.132
PhyT 2	Clinician-administered test	0.283
PhyT 3	Patient-activity based test	0.326
PhyT 4	Technical test	0.26
Impact on Clinic Time		
CT base	Usual clinic time	0.269
CT 2	Usual clinic time + 10 mins	0.278
CT 3	Usual clinic time + 30 mins	0.24
CT 4	Separate appointment, takes up to 60 mins	0.214
How will results influence care/treatment		
Res base	Doctor and patient decide together	0.46
Res 2	Doctor may change your general care	0.286
Res 3	Doctor may change your chemo/cancer treatment	0.253

4.4.3 MXL model results

Uncorrelated MXL models were estimated as the sample size was not large enough to obtain good estimates of a correlated MXL model.

Initially, a MXL model in which all parameter estimates were allowed to be random was estimated. Henceforth, this is known as Model 1 (see Table 4.16).

Table 4.16 MXL model 1

Log likelihood	−518
Obs	936
Iterations	88
Draws	2000

MXL Model with all parameters entered as random		Estimate	S.E.	P-value
S&Q 2 Mean	Symptoms & usual activities	0.401	0.178	0.025*
Det 2 Mean	Minor and major changes	1.812	0.391	0.000***
Q 2 Mean	3 questions to answer	0.436	0.268	0.105
Q 3 Mean	12 questions to answer	0.389	0.253	0.123
Q 4 Mean	20 questions to answer	0.431	0.262	0.1
PhyT 2 Mean	Clinician administered test	1.103	0.279	0.000***
PhyT 3 Mean	Patient activity based test	1.478	0.284	0.000***
PhyT 4 Mean	Technical test	1.113	0.278	0.000***
CT 2 Mean	Usual clinic time + 10 mins	0.09	0.201	0.653
CT 3 Mean	Usual clinic time + 30 mins	−0.137	0.222	0.535
CT 4 Mean	Separate appointment, takes up to 60 mins	−0.475	0.294	0.106
Res 2 Mean	Doctor may change your general care	−0.898	0.252	0.000***
Res 3 Mean	Doctor may change your chemo/cancer treatment	−1.037	0.24	0.000***
S&Q 2 SD	Symptoms & usual activities	0.084	0.597	0.889
Det 2 SD	Minor and major changes	2.134	0.473	0.000***
Q 2 SD	3 questions to answer	0.541	0.553	0.328
Q 3 SD	12 questions to answer	0.122	0.615	0.843
Q 4 SD	20 questions to answer	0.033	0.63	0.959
PhyT 2 SD	Clinician administered test	0.343	0.635	0.588
PhyT 3 SD	Patient activity based test	0.117	0.627	0.851
PhyT 4 SD	Technical test	0.203	0.959	0.833
CT 2 SD	Usual clinic time + 10 mins	0.262	0.55	0.634
CT 3 SD	Usual clinic time + 30 mins	0.684	0.42	0.103
CT 4 SD	Separate appointment, takes up to 60 mins	2.138	0.44	0.000***
Res 2 SD	Doctor may change your general care	1.549	0.397	0.000***
Res 3 SD	Doctor may change your chemo/cancer treatment	1.248	0.358	0.000***

*** p < 0.001; **p < 0.01; *p < 0.05

In R Studio (RStudio Team, 2020) there is the option of specifying some features of the Halton sequences. To reduce the risk of strong collinearity among the Halton sequences the first 500 sequence elements were dropped for each of the primes (Andersen, 2014; Czajkowski & Budziński, 2019).

In Model 1, some of the standard deviations associated with the attributes *Level of Detail*, *Impact on Clinic Time* and *Influence of results on care or treatment* were significant ($p < 0.001$). This indicated that respondents had heterogeneous preferences for these attributes.

Other models were also explored in order to find the model of best fit based on AIC and BIC, with a smaller number indicating a better fit to the data. Further MXL models with each of the attributes in turn allowed to be random were estimated. These are summarised in Table 4.17. In each instance, Model 1 was still considered to be a better fit to the data based on the AIC. However, models where parameter estimates associated with *Impact on Clinic Time* (CT_rand) or *Influence of results on care or treatment* (Res_rand) were the only attributes allowed to be random were considered to have a slightly better fit using the BIC.

Table 4.17 MXL model variations: AIC and BIC summary

MXL Model Variations Estimated	AIC	BIC
Model 1	1088	1214
Model with only levels of Sy&UA allowed to be random	1144	1212
Model with only levels of Det allowed to be random	1118	1185
Model with only levels of Q allowed to be random	1147	1224
Model with only levels of PhyT allowed to be random	1147	1224
Model with only levels of CT allowed to be random	1128	1205
Model with only levels of Res allowed to be random	1132	1204
Model with Q collapsed to a two level attribute	1078	1184
Model with PhyT collapsed to a two level attribute	1076	1182
Final model	1073	1165

Other model specifications considered included collapsing the questionnaire attribute to a two level attribute (Q_2lvl), in other words no questionnaire or questionnaire in the assessment. Similarly, a MXL model with physical tests collapsed to two levels (PhyT_2lvl) was also estimated.

Both these models were a better fit to the data than Model 1 according to both the AIC and BIC. However, the better fit to data comes at the expense of a loss of efficiency in choice sets. Originally, the choice sets were 87.2% efficient relative to the best possible using the D-optimality criterion for the MNL model with zero priors. However, after converting a 4-level attribute to a 2-level attribute, the efficiency decreases to 78.7%.

After considering these models, the final model of 'best' fit was similar to Model 1, but where attributes that had a significant standard deviation in at least one of their parameter estimates in Model 1 were allowed to be random. As a result, the final model estimated contained three attributes (six parameters) that were allowed to be random (see Table 4.18).

Similarly to the MNL model, respondents significantly preferred an assessment that asks about symptoms' impact on usual activities as opposed to an assessment that asks only about symptoms. Unlike the MNL model, parameters associated with the *Questionnaire* attribute were no longer significant. However, parameters associated with *Types of Physical Tests* attribute were still significant, indicating that respondents prefer a CIPN assessment to include some type of physical test. In particular, a *patient activity test* was most preferred.

Table 4.18 Final MXL model

Log likelihood	−518
Obs	936
Iterations	77
Draws	2000

Final MXL model		Estimate	S.E.	P-value
S&Q 2 Mean	Symptoms & usual activities	0.385	0.167	0.021*
Det 2 Mean	Minor and major changes	1.738	0.335	0.000***
Q 2 Mean	3 questions to answer	0.419	0.248	0.091
Q 3 Mean	12 questions to answer	0.39	0.237	0.101
Q 4 Mean	20 questions to answer	0.419	0.249	0.092
PhyT 2 Mean	Clinician-administered test	1.054	0.244	0.000***
PhyT 3 Mean	Patient activity-based test	1.425	0.246	0.000***
PhyT 4 Mean	Technical test	1.063	0.247	0.000***
CT 2 Mean	Usual clinic time + 10 mins	0.103	0.186	0.58
CT 3 Mean	Usual clinic time + 30 mins	−0.127	0.211	0.548
CT 4 Mean	Separate appointment, takes up to 60 mins	−0.439	0.279	0.115
Res 2 Mean	Doctor may change your general care	−0.86	0.229	0.000***
Res 3 Mean	Doctor may change your chemo/cancer treatment	−0.985	0.206	0.000***
Det 2 SD	Minor and major changes	2.033	0.403	0.000***
CT 2 SD	Usual clinic time + 10 mins	0.019	0.729	0.979
CT 3 SD	Usual clinic time + 30 mins	0.711	0.385	0.065
CT 4 SD	Separate appointment, takes up to 60 mins	2.062	0.377	0.000***
Res 2 SD	Doctor may change your general care	1.456	0.326	0.000***
Res 3 SD	Doctor may change your chemo/cancer treatment	1.146	0.292	0.000***

4.4.4 Examining MXL model results: choice probabilities

To examine the MXL model results further, choice probabilities were calculated. These are summarised in Tables 4.19 and 4.20.

Table 4.19 Choice probabilities: attributes with no heterogeneity

Attributes & Levels		Probability
Sy&UA	Symptoms and usual activities	
Sy&UA base	Symptoms only	40%
Sy&UA 2	Symptoms impact on usual activities	60%
Q	Questionnaire	
Q base	No questionnaire	18%
Q 2	3 questions to answer	28%
Q 3	12 questions to answer	27%
Q 4	20 questions	28%
PhyT	Physical Tests	
PhyT base	No physical test	9%
PhyT 2	Clinician administered test	26%
PhyT 3	Patient activity based test	38%
PhyT 4	Technical Test	27%

Table 4.19 shows the predicted choice probabilities for different levels of the attributes where respondents had homogeneous preferences. Using the attribute *Symptoms and Usual Activities* as an example and holding all other attributes constant, the modelling estimates that there is a 60% probability that respondents would choose an assessment which asks about how symptoms impact their usual activities and a 40% probability that respondents would choose an assessment which asks only about their symptoms. In other words, respondents preferred an assessment that asks about the impact of symptoms rather than one that asks about symptoms alone.

Similarly, respondents were slightly more likely to choose an assessment that contains a questionnaire of any length as opposed to one that contains no questionnaire. From the MXL model results, this slight preference for an assessment with a questionnaire was not significant ($p > 0.05$). Respondents were also more likely to choose an assessment that contained a physical test, with a *patient activity-based test* most likely to be chosen.

Table 4.20 summarises the mean probability and probability distribution for the attributes with at least one level that the MXL modelling found to be heterogeneous across respondents.

Table 4.20 Choice probabilities: attributes with heterogeneity

Attributes & Levels		Mean probability	Quantiles		
			25%	50%	75%
Det	Level of Detail				
Det base	Major changes only	15%	41%	15%	4%
Det 2	Minor and major changes	85%	59%	85%	96%
CT	Impact on Clinic Time				
CT base	Usual clinic time	28%	32%	28%	18%
CT 2	Usual clinic time + 10 mins	31%	35%	31%	20%
CT 3	Usual clinic time + 30 mins	24%	28%	24%	16%
CT 4	Separate appointment, takes up to 60 mins	18%	5%	18%	46%
Res	How will results influence care/treatment				
Res base	Doctor and patient decide together	46%	75%	56%	34%
Res 2	Doctor may change your general care	24%	12%	24%	39%
Res 3	Doctor may change your chemo/cancer treatment	21%	13%	21%	27%

For the attribute *Level of Detail*, on average respondents were strongly in favour of an *assessment that picks up small changes* in their condition whether it was important or not (85%). The choice probability distribution emphasises this preference with 75% of respondents having a 59% or stronger probability of choosing this level. In other words, although respondents had varying preferences, this was in relation to the strength of positive preference for this level rather than indicating a mix of positive and negative preferences.

Examining *Impact on Clinic Time*, on average respondents were least likely to choose an assessment that required a separate appointment; with a probability of 18%. However, there was a small proportion of respondents (25%) who actually preferred having a separate appointment altogether, with a 46% chance of choosing an assessment with such a feature.

Finally, for the attribute *How will results influence care/treatment*, holding all other attributes constant the modelling estimates there was a much greater probability, on average, of respondents choosing the situation where the doctor and patient decide together what assessment results mean compared to the situations where the doctor alone decides on changes to care or treatment. Respondents were particularly sensitive to the situation where the doctor may decide to change the treatment without

consulting the patient. Although respondents had varying preferences for this particular level, the choice probability distribution had a relatively narrow range, with 75% of respondents having a 27% or lower probability of choosing an assessment with such a feature.

4.4.5 Examining MXL model results: kernel density plots

To examine the levels with heterogeneity further, kernel density plots were used, as summarised in Figures 4.13 to 4.15.

Figure 4.13 Kernel density plot of the conditional mean distribution for Det 2

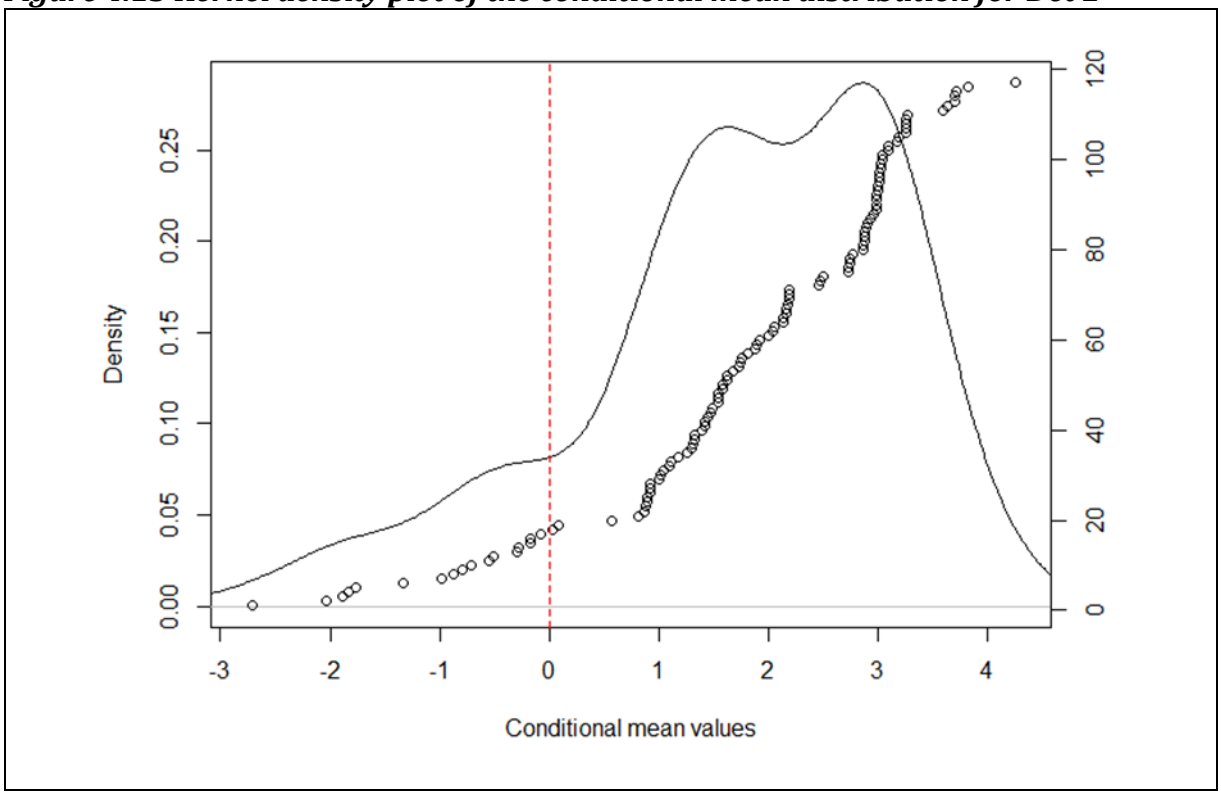


Figure 4.14 Kernel density plot of the conditional mean distribution for CT 4

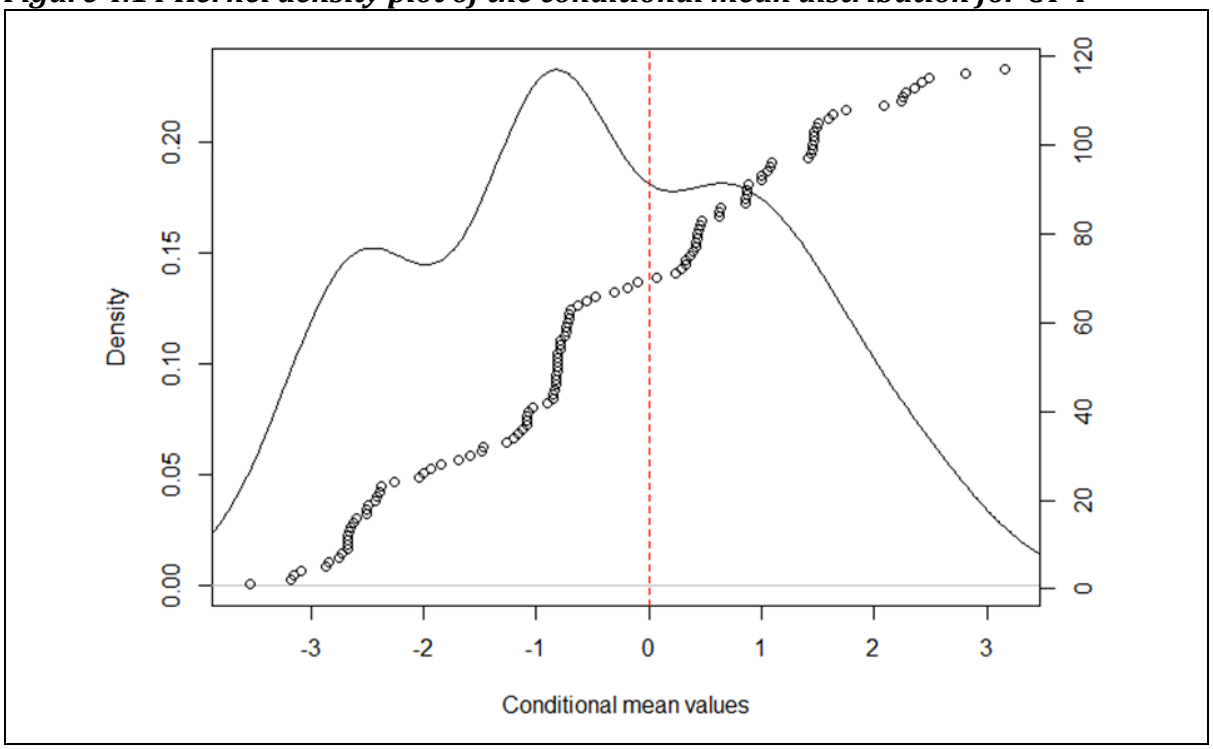
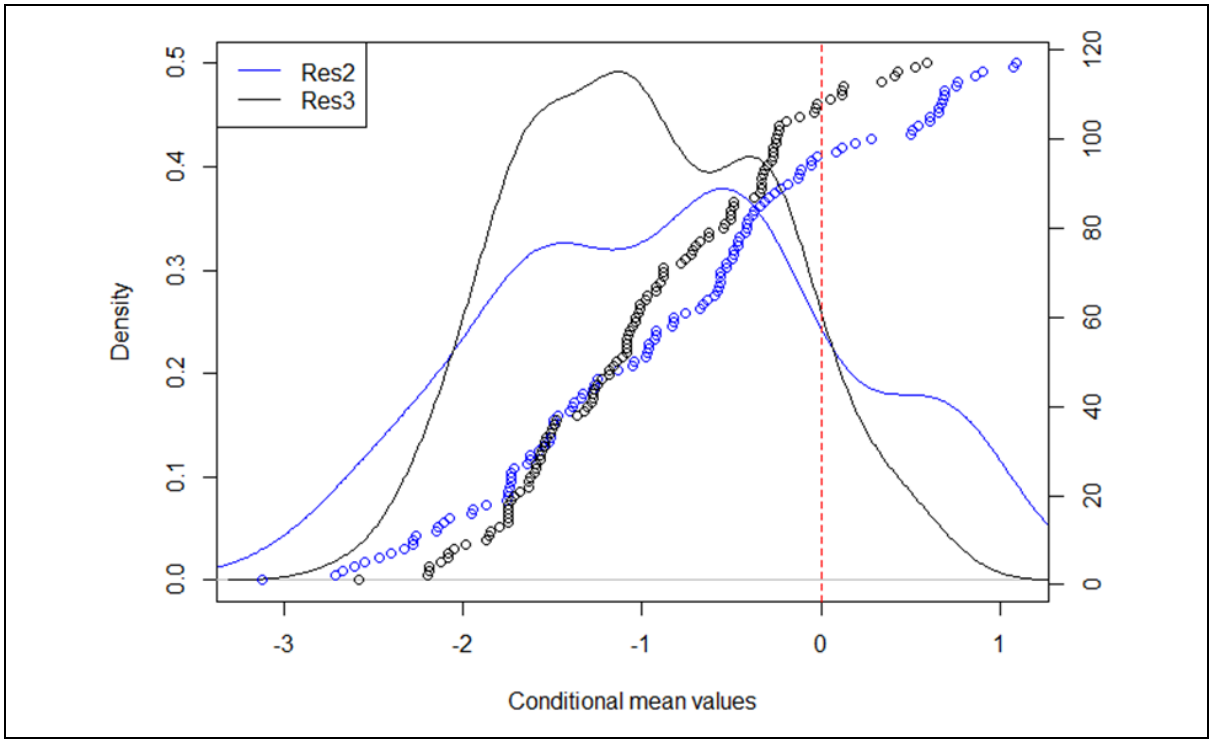


Figure 4.15 Overlaid kernel density plot of the distribution of conditional means for Res 2 and Res 3



For the attribute *Level of Detail* in Figure 4.13, the conditional means for each of the 117 respondents were overlaid on top of the kernel density plot in ascending order. The

majority of conditional means were skewed upwards and above zero, indicating that the majority of respondents preferred an assessment that picks up both minor and major changes. The kernel density plot features two peaks, which are both positive. This indicates that there were a group of respondents who had a strong positive preference for an assessment that detects minor and major changes and another group that had the same preference but did not hold it as strongly.

The mean parameter estimates for the attribute levels of *Impact of clinic time* (CT) were not significant. However, there was significant heterogeneity observed in CT 4, which was the parameter estimate for the level requiring a separate appointment up to 60 minutes long. Figure 4.14 provides a kernel density plot of CT 4; in addition, the conditional mean values in ascending order were overlaid on top.

In Figure 4.14 the largest peak in conditional mean values was between 0 and −2 with a smaller peak between −2 and −4. This indicates that there were two groups of respondents who found having to attend a separate appointment to be significantly worse than completing the assessment during usual clinic time, with one group holding a stronger preference compared to the other. It was noted that there was also another 'peak' around the positive value of 1, indicating there was also a minority group of respondents who perceived a separate appointment to be better than having an assessment in the usual clinic time.

Significant heterogeneity was also observed for the parameter estimates associated with the attribute *How will results influence care/treatment* (Res). Figure 4.15 provides an overlaid kernel density plot of the parameter estimates Res 2 and 3. Again, the corresponding conditional mean values in ascending order were overlaid. Conditional means were skewed towards negative values, indicating that respondents preferred an assessment where the impact of results is decided together between the patient and the doctor. In particular, the major peak in the kernel density plot for Res 3, between the values of −1 and −2, was much higher than the highest peak for Res 2. In other words, while the majority of respondents were averse to a situation where the doctor may change their general care, they were even more averse to the situation where the doctor makes the decision alone to change their treatment.

4.4.6 Summary of responses to evaluation questions

Respondents were asked a number of questions about the choice sets. Over half (57%) found the information to define peripheral neuropathy and to explain some key definitions and terms prior to the choice sets helpful in terms of understanding the choice sets. A third (37%) of respondents did not find the information helpful, since they already knew this information prior to the survey.

A large majority of respondents (79%) agreed or strongly agreed with the statement that it was easy to identify differences between assessment options most of the time. A majority (73%) also agreed or strongly agreed with the statement that it was easy to choose between assessment options most of the time.

In terms of the strategy for choosing whether Assessment A or B was better in each choice set, there was a mix of responses, summarised in Table 4.21. A large minority of respondents only considered attributes/levels that were most important to them in each option.

Table 4.21 Responses to question 'How did you decide whether Assessment A or B was better in each choice question?'

Options	No. (%)
I considered all features for each assessment option	25 (21%)
I considered all features which were different between the two assessment options	31 (26%)
I considered only the features which were most important to me	52 (44%)
Other strategy	9 (8%)

Respondents were also asked if there were attributes not included in the choice sets that they thought were important. Eighty-three per cent of respondents said no, which indicates that for the large majority of respondents the attributes that were included were sufficient. Of the 20 respondents who indicated yes, some mentioned mental or emotional health. Many of the responses were not directly relevant to the question, but rather were general comments. For instance, many comments were about encouraging discussion of CIPN during treatment or about being made aware of the long-term impacts of symptoms. Actual responses from respondents are available in Appendix 4L.

4.5 Discussion of Study 1: patient sample results

4.5.1 Statement of principal findings

As far as the author is aware, this is the first study that has formally examined patient preferences for the design of a CIPN assessment tool. As such, this study adds a unique and important perspective to the design of a CIPN assessment tool.

Overall, respondents preferred a CIPN assessment that asks about symptoms' impact on usual activities rather than asking about symptoms alone. This suggests that respondents were more concerned about the impact of CIPN on their quality of life as opposed to the presence of CIPN symptoms in their lives. A CIPN assessment that contained some form of physical test was of particular importance to respondents, with a preference for a physical test that involved an active role for the patient (e.g. a peg board test or a sway test). This suggests that respondents preferred a test that is objective but also involves input from the patient.

Significant variation in preferences was observed when it came to the level of detail of the CIPN assessment, impact of the assessment on clinic time and impact of the results on care or treatment. Variation in preference was mostly in relation to the strength of preference rather than a polarisation in preferences. That is, from the kernel density plots it was observed that most of the peaks were clustered around positive values or negative values. In other words, the significant standard deviations that were seen in the MXL model were indicating that respondents had different strengths of preference rather than having opposing preferences for a particular level.

It was of also of note that respondents had a strong preference for the use of assessment results to be decided together between the doctor and patient rather than the doctor making the decision alone. This highlights a desire by respondents for control in their own general care and/or chemotherapy treatment.

In comparing the current findings to the literature, it is interesting to note that McCrary et al. (2017) report that three out of the six assessments which were considered the 'best' based on different assessment criteria were patient-reported outcomes (PRO). The other three included a clinical grading scale and a composite of clinical grading and objective measures. Findings from the current study indicate that patients would

respond positively to the use of physical tests as a core part of the CIPN assessment process. On the other hand, it is not certain whether patients would value the use of clinical grading scales based on the current findings. In particular, a key theme that emerged from the results of the current study was a preference for both patient and doctor input. This is demonstrated by the preference for a physical test – that is, an objective measure – where patients were actively involved and also by a strong preference for patient and doctor involvement in decisions on care or treatment.

4.5.2 Strengths and weaknesses of the study

A potential limitation of the study is the small sample size (n = 117), noting this was a survey involving participants from a volunteer sample. This is a result of the recruitment source of the IN FOCUS Study. These factors possibly limit the generalisability of findings. However, the sample consisted of respondents who had direct experience or interest in the subject. In particular, the majority of respondents were currently undergoing chemotherapy or had completed their chemotherapy treatment within the last five years at the time of the survey.

4.5.3 Study implications for clinicians or policy makers

Much of the literature on CIPN assessment tools has focused on the psychometric properties, comparison with other tools and validation in cancer populations (Calhoun et al., 2003; Cavaletti et al., 2007; Haryani et al., 2017; McCrary et al., 2017; Shimozuma et al., 2009). While these are certainly important areas of research, the current study can be considered as the start of the development of a complementary literature that aims to incorporate patient views into the process of evaluating the validity and significance of different CIPN assessment tools. In particular, findings from the current study can assist clinicians in selecting a CIPN assessment tool to use in their routine clinical practice that takes account of patient preferences.

4.6 Next steps

In this chapter, the research project was introduced and results from Study 1, the patient sample, were discussed. In the next chapter, results from Study 2, the general population sample, will be discussed. This includes a comparison of results from Studies 1 and 2.

Chapter 5. Preferences for the assessment of chemotherapy induced peripheral neuropathy (CIPN): Contrasting the general population and patient perspectives

5.1 Background

As discussed in the previous chapter, there are many reasons for using a general population sample, even when investigating preferences for a specific health intervention. For example, there may be no ready access to an appropriate patient sample. Even where a patient sample is available, in many cases it may not be possible to access a sufficiently large sample of patients in a reasonable timeframe, such as when considering rare conditions. Recruitment of patients may also be prohibitively expensive, since dedicated patient panels can be costly to access. It must also be considered that an appropriately experienced patient sample may not even exist, especially if the study is investigating new treatment options that have not been tried in patients before.

Consequently, there are many situations where the best option may be to use a general population sample, which makes it possible to recruit large numbers of participants relatively quickly and at low cost. The use of a sample drawn from the general population also allows for consideration of the preferences of those who have not experienced a particular treatment but may do so in the future.

The main limitation of this approach, however, is the potential differences in knowledge, experience and, therefore, comprehension between a patient sample that is familiar with the health condition and a general population sample with limited knowledge or experience of the health condition and associated treatments.

It is also important to understand whether potential differences in preferences between patients and the general population can be reduced by providing adequate information about the health problem. For example, it would be expected that the vast majority of the general population would have no knowledge of CIPN, its related assessment tools or how it can impact daily life. Providing detailed information to educate respondents may assist them in understanding the context behind the DCE and in making informed

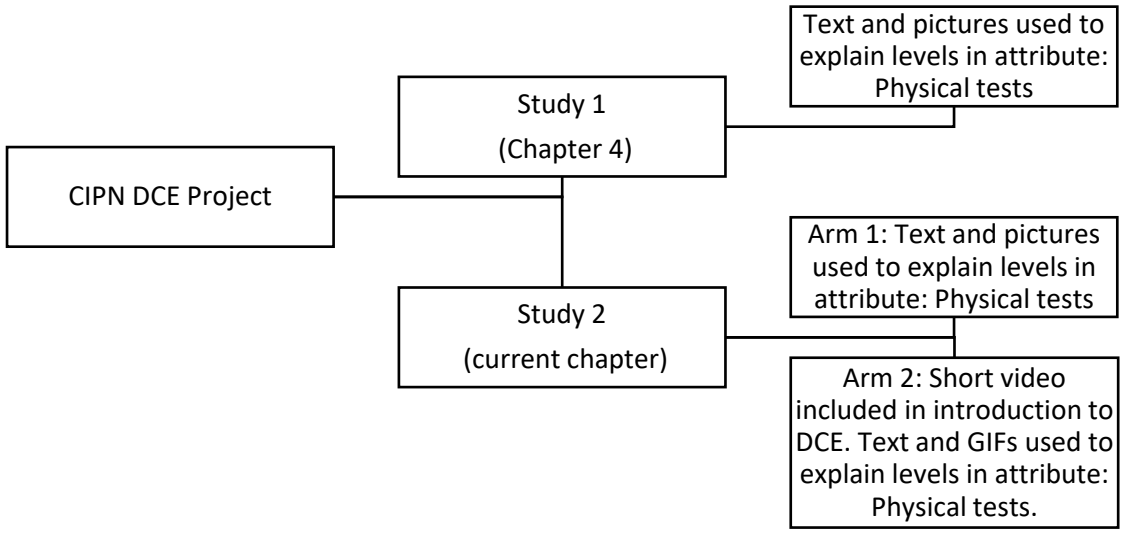
decisions when completing choice sets, and these responses may be more similar to those of experienced patients.

The first part of this chapter describes Study 2, which was used to investigate general population preferences for how CIPN is assessed. The aim of Study 2 was to investigate the impact of providing different levels of information on elicited preferences, using a general population sample. To investigate this aim, as discussed in the previous chapter, Study 2 was split into two arms. In Arm 1, respondents receive the same level of information provided in Study 1 (that is, the patient sample). In Arm 2, respondents receive additional information in the form of moving pictures or GIFs, and a short video. The preferences from Arm 1 and Arm 2 of the general population sample are then analysed and compared. The purpose of the additional information was to provide respondents with increased understanding of peripheral neuropathy and of the tests used to assess it.

The second part of this chapter compares Study 1, the patient sample, and each of the arms of Study 2, the general population sample. This is done in order to understand whether increasing the level of information provided to a general population sample can minimise preference differences from a patient sample.

Figure 5.1 provides an overview of the overall CIPN DCE study (Studies 1 and 2) as well as the separate arms of Study 2.

Figure 5.1 Outline of CIPN DCE Project: Study 1 and 2

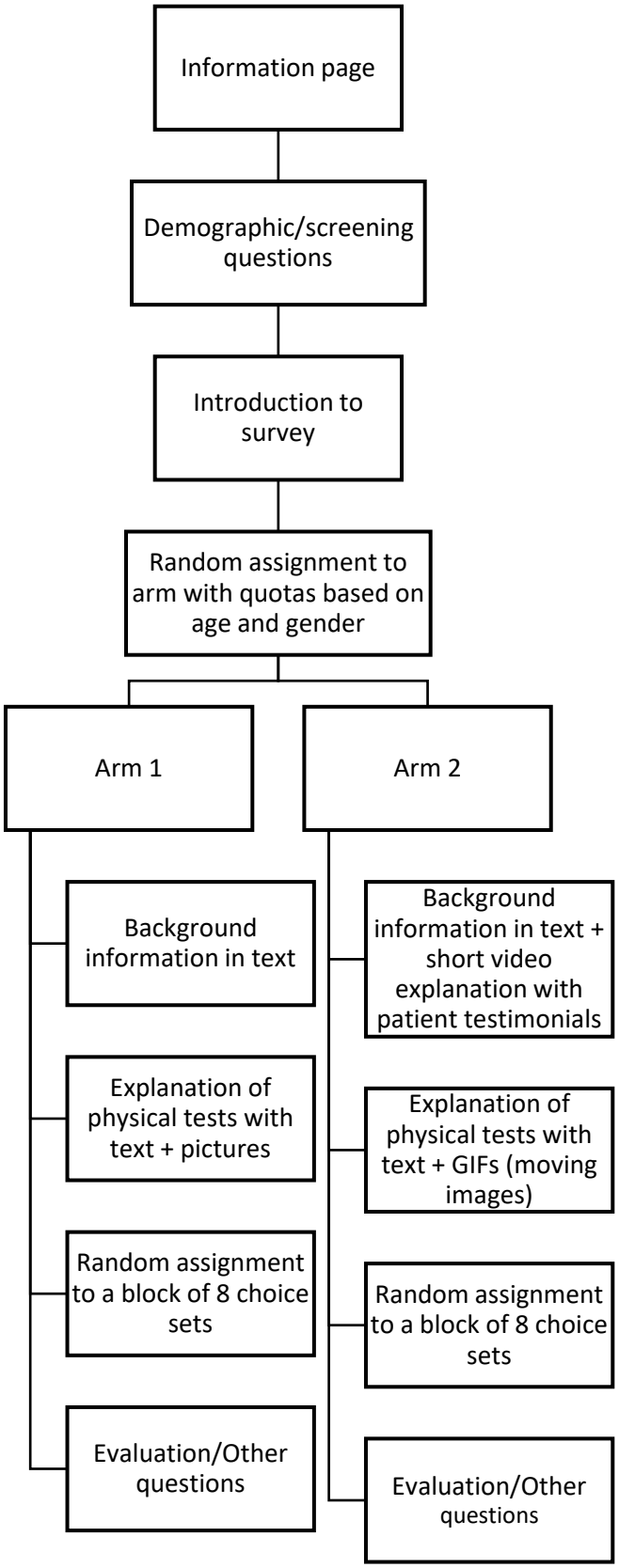


5.2 Study 2: Methods

5.2.1 Overview of Study 2

Figure 5.2 provides a flow chart detailing contents of the survey seen by both arms in Study 2. Potential respondents were all initially taken to an information page when they clicked on the survey link. This page provided details regarding the purpose of the study. At the bottom of the page was a question asking respondents if they consented to participating in the study. If respondents were not willing to consent, they could click 'no' and they were then redirected to a screen-out page. If respondents were willing to consent, they indicated consent by clicking 'yes' and continuing on to the next page. Respondents were then asked to answer some basic demographic questions relating to their age, gender and education level. Age and gender quotas were used to ensure respondents were representative of the Australian adult population. After these demographic questions was a page explaining that they would see some background information over the next few pages. The content of the background information differed between the two arms of the study, as described in the next section. After viewing the background information, respondents were randomly assigned to one of four blocks of choice sets to complete. Respondents were then asked some debriefing questions about the background information and choice sets as well as some questions related to their experience with CIPN. On the last page, respondents were given the option to provide further comments in a free-text box.

Figure 5.2 Survey flow for arms 1 and 2 in Study 2



5.2.2 Background information section of DCE

As noted, the two arms in Study 2 differed in how the background information about CIPN was presented. The content of Arm 1 was purposely kept as close as possible to what was seen by the patient sample. This allowed comparison of results by population type.

The only notable difference between the surveys shown to the patient sample and Arm 1 of the general population sample was in the spacing of the introductory information. The patient sample veiwed the introductory information over five pages while those assigned to Arm 1 of Study 2 saw the same information over nine pages. For instance, the first page of the background information given to the patient sample included three main topics: an introduction to CIPN, common symptoms and problems it can lead to in everyday life. For the general population sample, this information was divided into three pages as it was expected that for the majority of respondents from the general population, CIPN would be entirely new concept and the idea of undergoing physical tests for CIPN would also be unfamiliar. Hence, the respondents were given smaller amounts of information per page to allow them more time to absorb the information.

5.2.3 Experimental arms

The information provided to Arm 2 was more detailed than for Arm 1. In addition, the presentation format of information was also different between the arms. In Arm 1 (and in the survey for the patient sample), static images were used with text description of the types of physical tests that a CIPN assessment could involve. In contrast, for Arm 2, short videos on loop (i.e. GIFs) of the physical tests were used instead.

Arm 2 also included a video explanation of CIPN, which included some testimonials by patients about their experience with CIPN. The video used in Arm 2 was an extract from a video by the Swedish Health Services (Swedish Health Services, 2018) and was used with permission.

5.2.4 Summary of attributes, levels and choice sets

Attributes, levels and choice sets used were the same as for the patient sample in Study 1 and a full description of the development of these can be found in Chapter 4.

In brief, attributes and levels were initially identified by clinicians and two patient representatives who are part of the IN FOCUS Study (InFocus, 2019). The attributes and levels, and the survey more generally, were further refined by consultation with researchers familiar with DCEs and also through cognitive interviews with six current and former cancer patients.

Choice sets were the same as for the patient sample. In other words, there were 32 choice sets in total, partitioned into four blocks of eight choice sets each. Respondents were randomly assigned one block of eight choice sets to complete.

5.2.5 Development of survey

The patient sample survey was used as a starting point and modifications were made to the introduction section, as discussed above. The evaluation questions after the choice sets were also modified to suit the general population sample. Apart from questions about respondent understanding of choice sets, they were also asked to rate how informative the introduction section was. Respondents were asked about their experience with illness and cancer. A pilot version of the survey for the general population was given to a convenience sample of friends and colleagues. The survey was refined based on their comments and feedback. This was mostly in relation to spelling, grammatical error and presentation of survey. The final survey can be found at Appendix 5A.

5.2.6 Ethics approval

Ethics approval was obtained for Study 2 (UTS HREC REF NO. ETH18-2507). A copy of the letter of approval appears at Appendix 5B.

5.2.7 Recruitment

Respondents were recruited through an online Australian panel partnered with the online survey platform used for this survey, SurveyEngine (SurveyEngine GmbH, 2021). Respondents in each arm were representative of the Australian population in terms of age and gender. Quotas were based on Australian population statistics of June 2018, which were the latest available on the ABS website (Australian Bureau of Statistics, 2020) at the time of this survey. The goal of recruitment was to obtain a sample that had minimal or no knowledge or experience with CIPN. Hence, less emphasis was given to

the education level or digital literacy of respondents and more on ensuring a representative general population sample.

5.2.8 Planned sample size

Sample size was based on the need for a sufficient number of responses per choice set to allow for complex modelling, such as the mixed logit model, and to minimise asymptotic standard errors. As noted in Chapter 3, Burgess et al. (2011) find that average asymptotic standard errors increase noticeably once there are fewer than 20 responses per choice set. The sample size was chosen so that there would be 40 responses per choice set. This required 160 respondents per arm, and so 320 respondents in total needed to be recruited.

Another factor for consideration when determining the sample size needed was the planned analyses of responses to the evaluation questions after the choice sets. It was of particular interest to compare the responses to the evaluation questions by arm. Evaluation questions required rating on a 5-point Likert scale and responses between arms were analysed using the chi-squared test. The power of the chi-square tests based on the planned sample size of 320 respondents was calculated in order to understand the likelihood of detecting differences between the two arms as captured by responses, if differences were actually present. In particular Cohen's d effect sizes were used as a rule of thumb, with 0.1 is considered a small effect size and 0.3 a medium effect size (Cohen, 1988; NCSS Statistical Software, 2021). This was calculated in Mathematica, using code written by Street (2020b). For the planned sample size of 320 respondents, the probability of detecting a small effect size when it is present is only 26%. With the planned number of respondents, a medium effect size of 0.3 has an extremely high probability (over 99%) of being detected. In addition, for a sample size of 320 respondents, there is a 62% probability of detecting a small to medium effect size of 0.16. Hence, it should be kept in mind that with the planned sample size of 320 respondents, there is a reasonable probability of detecting small to medium effect sizes, if present.

5.2.9 Testing poolability of data from different sources

One approach to comparing preferences from different sources, such as different samples or arms, is to pool the data for a formal analysis of differences. However, this is

only possible if it is actually appropriate for the data to be pooled – that is, if there are no scale differences. The presence of scale differences between different sources needs to be investigated (Vass, Wright et al., 2018). The scaled MNL model or S-MNL model from the gmnl package in R (Sarrias & Daziano, 2017) was used to do this. More detail about the S-MNL model can be found in Chapter 2.

The S-MNL model can be used to test for the presence of error variances across respondents, and whether this heterogeneity can be accounted for by the DCE that respondents completed (Sarrias & Daziano, 2017). In other words, this can be used to test for any potential scale differences by arm or by sample type. This is done by including sources such as arm assignment or sample type as an individual characteristic when estimating the S-MNL model.

The S-MNL model was used to test for scale differences by arm in the general population sample, represented by the δ parameter in the model. To estimate the parameters in the S-MNL, R-Studio's default non-linear optimisation algorithm, the Broyden–Fletcher–Goldfarb–Shanno (BFGS), with 2000 Halton draws was used.

The S-MNL model was also used to test for scale differences between the patient and general population samples. However, the BFGS algorithm produced NAs (missing values) in the model estimates. As a result, an alternative algorithm was tried. The Berndt-Hall-Hall-Hausman (BHHH) algorithm was chosen over the Newton-Raphson (NR). The NR algorithm can be used to maximise any function; however it much slower than the BHHH algorithm because it maximises the sum of log likelihoods over the sample of observations (see Chapter 8 of Train (2009) for further detail). When estimating the S-MNL, as a proof of concept, 100 draws were used first, before moving on to 1000 draws. Draws were increased until results stabilised at 10,000 draws.

5.3 Study 2: Results

5.3.1 Sample characteristics

In total, 448 respondents agreed to participate and started the survey. There were 74 respondents that completed the demographic questions but did not go on to complete the rest of the survey. Of these 74 respondents, 38/74 (51%) were female and 36/74

(49%) were male). Respondents that did not finish the survey were much more likely to be older, with 36/74 (49%) aged 60 and above.

340 respondents completed the whole survey. A further five respondents were removed because their responses in the comments box (a string of random of letters in the comment box section; see Appendix 5C) indicated they may not have provided valid responses to the choices sets. As a result, 335 respondents were included for data analysis, with 167 respondents in Arm 1 and 168 respondents in Arm 2. A summary of the numbers at each point of the survey has been provided in Table 5.1.

Table 5.1 Survey status

Survey Status	No. (%)
Started survey	448 (100%)
Completed demographic questions	414 (92%)
Completed choice sets	359 (80%)
Completed some evaluation questions	341 (76%)
Completed whole survey	340 (76%)
Included for data analysis*	335 (75%)

*five respondents excluded due to unintelligible text in comment box section

Quotas were assigned for recruitment to ensure that each arm was representative of the Australian population in terms of age and gender (Australian Bureau of Statistics, 2020). Quotas were based on the latest statistics available at the time, which was June 2018. Table 5.2 provides a summary of the age and gender distribution across both arms.

Table 5.2 Frequency of gender and age of respondents

Gender	Age Range	Arm 1 No.	Arm 2 No.	Total No.	Total %	ABS %*
Female	18–39	33	33	66	20%	20%
	40–59	32	26	58	17%	17%
	60 and over	22	26	48	14%	14%
Male	18–39	32	31	63	19%	20%
	40–59	26	25	51	15%	16%
	60 and over	21	25	46	14%	13%
Other	Other	1	2	3	1%	n/a
Total		167	168	335	100%	100%

*Australian Bureau of Statistics (ABS) June 2018: age and gender distribution of the Australian population

Approximately 47% of respondents did not attend university, with many of these respondents selecting TAFE or secondary school as their highest form of education. Table 5.3 provides a summary by arm. Chi-square tests were carried out to test for

whether age, gender or education was dependent on arm assignment, no significant association was found ($p > 0.05$).

Table 5.3 Education level

Education level	Arm 1 (%)	Arm 2 (%)
No school certificate or other qualifications	1 (1%)	2 (1%)
Secondary school	25 (15%)	44 (26%)
Trade or apprenticeship	6 (4%)	8 (5%)
TAFE or vocational college	42 (25%)	29 (17%)
Bachelor's degree	59 (35%)	46 (27%)
Postgraduate degree	34 (20%)	39 (23%)
Total	167 (100%)	168 (100%)

Each respondent was randomly assigned to an arm and then randomly assigned to a block of eight choice sets. This has been summarised in Table 5.4. A chi-square test was performed and found no significant association between block assignment and arm ($p = 0.3$).

Table 5.4 Block assignment

Block	Arm 1 (%)	Arm 2 (%)
1	41 (25%)	41 (24%)
2	46 (28%)	39 (23%)
3	34 (20%)	49 (29%)
4	46 (28%)	39 (23%)
Total	167 (100%)	168 (100%)

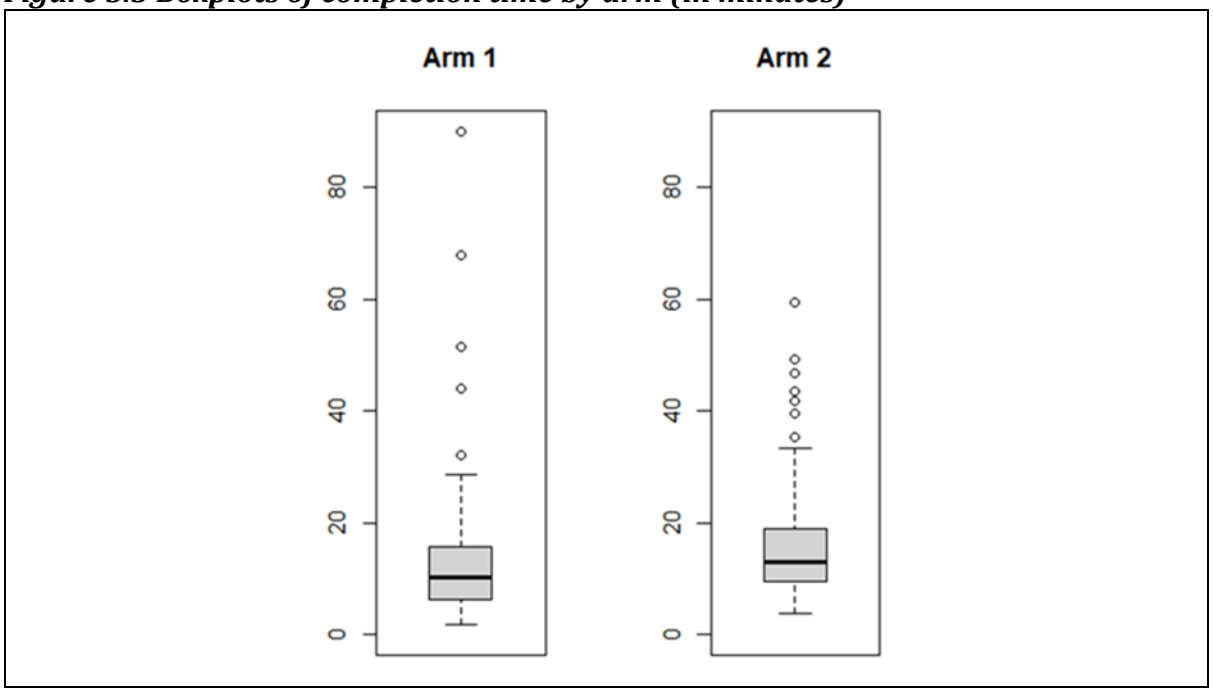
The time taken to complete the survey is summarised in Table 5.5. The median time to complete the survey was higher for Arm 2 at almost 13 minutes versus 10 minutes for Arm 1. A higher median time of completion is expected as those in Arm 2 were compelled to watch a video 1 minute and 35 seconds long and the 'next' button was disabled until the video had completed playing. In addition, lag and loading time for the video may have contributed to the longer median time for completion. Completion time at 10th percentile for Arm 1 was 4 minutes 31 seconds, for Arm 2 this was 6 minutes 40 seconds.

Table 5.5 Completion time

Time (minutes)	Mean	Median	Min	Max	SD
Arm 1	12.714	10.253	1.772	89.818	10.508
Arm 2	15.439	12.881	3.653	59.417	9.119

Boxplots were also used to examine the spread of completion times as shown in Figure 5.3. A Kruskal-Wallis test was used to examine whether completion time was dependent on assigned arm. This test was chosen as it does not require an assumption of a normal distribution. The Kruskal-Wallis test was estimated in R Studio (RStudio Team, 2020) and was significant ($p < 0.001$), indicating that completion time was significantly longer in Arm 2.

Figure 5.3 Boxplots of completion time by arm (in minutes)



5.3.2 Analysis of general population sample

Overview of analyses

To begin with, MNL models, which assume homogeneous preferences, were estimated for each arm. The MNL models were also estimated with the fastest 10% of respondents removed from each arm to test for sensitivity of results to length of time taken to complete the survey. No major differences were observed, and all respondents were therefore retained for the remaining analyses. Next, the S-MNL model was estimated to test whether Arm 1 and Arm 2 could be combined together for analysis. No significant

scale difference was found, allowing Arm 1 and Arm 2 to be pooled. A MNL model on the combined data was estimated to formally test for preference differences by arm. None was found; hence, the rest of the results section is focused on finding the model of best fit using the general population data as a whole. This includes the LCA and the MXL model, both of which relax the assumption for homogeneous preferences.

MNL model results by arm

Initially, separate MNL models for each arm were estimated. In general, it was found that the same terms were significant (see Table 5.6). The exception is the parameter estimate of the final level for the attribute 'Clinic Time', which was significant at the 1% level for Arm 1 ($p < 0.01$) and significant at the 10% level for Arm 2 ($p = 0.099$).

Table 5.6 MNL results by arm

	Arm 1	Arm 2
Log likelihood	−872	−872
Obs	1336	1344
Iterations	3	4
AIC	1769	1771
BIC	1837	1838

MNL by Arm	Arm 1		Arm 2	
	Estimate	P-value	Estimate	P-value
S&Q 2 (symptoms & usual activities)	−0.056	0.487	0.055	0.489
Det 2 (minor and major changes)	0.509	0.001***	0.659	0.00***
Q 2 (3 questions to answer)	−0.07	0.555	0.105	0.373
Q 3 (12 questions to answer)	−0.048	0.652	0.131	0.225
Q 4 (20 questions to answer)	0.009	0.94	0.064	0.585
PhyT 2 (clinician administered test)	0.461	0.001***	0.521	0.000***
PhyT 3 (patient activity based test)	0.5	0.001***	0.416	0.000***
PhyT 4 (technical test)	0.411	0.001***	0.547	0.000***
CT 2 (usual clinic time + 10 mins)	0.022	0.837	0.018	0.858
CT 3 (usual clinic time + 30 mins)	−0.064	0.504	−0.09	0.341
CT 4 (separate appointment, takes up to 60 mins)	−0.349	0.001***	−0.171	0.099
Res 2 (doctor may change your general care)	−0.24	0.003***	−0.156	0.048**
Res 3 (doctor may change your chemo/cancer treatment)	−0.29	0.001***	−0.234	0.003***

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Sensitivity analyses using the MNL model

The MNL model was used to test whether preferences were sensitive to potential response quality issues.

It was noted that the minimum completion times for arms 1 and 2 were 1:46 and 3:39 respectively, recalling that those in Arm 2 had to allow for a 1:35 video to play. These completion times appear to be unrealistic in order for a respondent to be able to comprehend and complete the whole survey. As a consequence, MNL models were estimated with the fastest 10% (n = 17) of respondents in both arms removed (see Appendix 5E). For Arm 1, in terms of the direction and significance of values, no difference was found when the 10% of respondents with the shortest completion times were removed. Similarly for Arm 2, the removal of the fastest 10% did not make a difference in terms of direction and significance of values, with the exception of the parameter estimate for a questionnaire with 12 questions, which was not significant previously (p = 0.225) but became significant at the 5% level (p = 0.028).

In the evaluation questions, respondents in Arm 2 were asked whether they were able to view the full video. Four respondents indicated they were unable to view the video due to technical difficulties. The MNL model was estimated for Arm 2 with these four respondents removed (see Appendix 5F). No differences in terms of the direction or significance of values were noted.

S-MNL model results

Considering how similar results were between the arms, the possibility of combining arms for analysis was explored using the S-MNL model. S-MNL model results are summarised in Table 5.7.

The δ_{arm} parameter estimate had a p-value of 0.123, indicating that the arm respondents were assigned to did not account for scale differences. That is, there were no significant scale differences in the parameters between the two arms. As a result, arms could be combined for analysis.

Table 5.7 S-MNL model results

Log likelihood	−1730
Obs	2680
Iterations	4
Draws	10, 000
AIC	3498
BIC	3587

S-MNL	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	0.071	0.052	0.170
Det 2 (minor and major changes)	0.697	0.116	0.000***
Q 2 (3 questions to answer)	−0.016	0.081	0.839
Q 3 (12 questions to answer)	0.068	0.079	0.388
Q 4 (20 questions to answer)	−0.003	0.081	0.972
PhyT 2 (clinician administered test)	0.444	0.106	0.000***
PhyT 3 (patient activity based test)	0.432	0.081	0.000***
PhyT 4 (technical test)	0.533	0.107	0.000***
CT 2 (usual clinic time + 10 mins)	0.039	0.074	0.598
CT 3 (usual clinic time + 30 mins)	−0.076	0.075	0.311
CT 4 (separate appointment, takes up to 60 mins)	−0.227	0.079	0.004**
Res 2 (doctor may change your general care)	−0.249	0.064	0.000***
Res 3 (doctor may change your chemo/cancer treatment)	−0.261	0.060	0.000***
τ	1.201	0.216	0.000***
δ_{arm}	0.266	0.173	0.123

*** p < 0.001; **p < 0.01; *p < 0.05

MNL model: testing preference differences by arm

A model was estimated with the combined data. A dummy indicator for the arm was defined and then interacted with each attribute to test for any differences in preferences by arm (see Table 5.8).

Table 5.8 MNL results with dummy indicator for arm

Log likelihood	−1740
Obs	2680
Iterations	4
AIC	3540
BIC	3693

MNL with dummy for arm	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	−0.056	0.081	0.486
Det 2 (minor and major changes)	0.509	0.083	0.000***
Q 2 (3 questions to answer)	−0.070	0.119	0.555
Q 3 (12 questions to answer)	−0.048	0.107	0.651
Q 4 (20 questions to answer)	0.009	0.119	0.940
PhyT 2 (clinician administered test)	0.461	0.109	0.000***
PhyT 3 (patient activity based test)	0.500	0.097	0.000***
PhyT 4 (technical test)	0.411	0.106	0.000***
CT 2 (usual clinic time + 10 mins)	0.022	0.105	0.836
CT 3 (usual clinic time + 30 mins)	−0.064	0.096	0.504
CT 4 (separate appointment, takes up to 60 mins)	−0.349	0.106	0.001***
Res 2 (doctor may change your general care)	−0.240	0.079	0.002***
Res 3 (doctor may change your chemo/cancer treatment)	−0.290	0.08	0.000***
Sy&UA 2:dummy	0.111	0.114	0.327
Det 2:dummy	0.150	0.119	0.207
Q 2:dummy	0.175	0.168	0.296
Q 3:dummy	0.180	0.152	0.238
Q 4:dummy	0.055	0.167	0.741
PhyT 2:dummy	0.060	0.154	0.699
PhyT 3:dummy	−0.084	0.137	0.539
PhyT 4:dummy	0.136	0.153	0.374
CT 2:dummy	−0.003	0.147	0.983
CT 3:dummy	−0.026	0.135	0.847
CT 4: dummy	0.178	0.148	0.228
Res 2:dummy	0.084	0.112	0.451
Res 3:dummy	0.057	0.112	0.615

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

None of the dummy interacted parameters were significant; that is, all had p-values > 0.05, indicating that there were no significant differences in preferences for attribute levels by arm.

Final MNL model: pooled data

The final MNL model included the combined data from arms 1 and 2 and will be the focus for the analysis of the general population sample. These results are summarised in Table 5.9.

Respondents preferred an assessment that picks up both minor and major nerve damage, whether it was important or not. Respondents did not have a significant preference between 'no questionnaire' and 'questionnaire between 3 to 20 questions long'. In other words, they did not have a significant preference for having a questionnaire as part of the assessment. In contrast, respondents strongly preferred an assessment that included a type of physical test.

Table 5.9 MNL results

Log likelihood	−1749.4
Obs	2680
Iterations	4
AIC	3524.7
BIC	3601.3

MNL	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	0.007	0.057	0.905
Det 2 (minor and major changes)	0.580	0.059	0.000***
Q 2 (3 questions to answer)	0.017	0.083	0.837
Q 3 (12 questions to answer)	0.039	0.076	0.604
Q 4 (20 questions to answer)	0.036	0.083	0.665
PhyT 2 (clinician administered test)	0.487	0.077	0.000***
PhyT 3 (patient activity based test)	0.459	0.068	0.000***
PhyT 4 (technical test)	0.474	0.076	0.000***
CT 2 (usual clinic time + 10 mins)	0.019	0.073	0.793
CT 3 (usual clinic time + 30 mins)	−0.076	0.067	0.257
CT 4 (separate appointment, takes up to 60 mins)	−0.260	0.074	0.000***
Res 2 (doctor may change your general care)	−0.204	0.056	0.000***
Res 3 (doctor may change your chemo/cancer treatment)	−0.265	0.056	0.000***

*** p < 0.001; **p < 0.01; *p < 0.05

Respondents did not mind spending up to an additional 30 minutes on top of their usual clinic time in order to be assessed for CIPN, although they were generally against the situation of a separate appointment of up to 60 minutes. In general, respondents were against the situation where the doctor alone makes decisions about a patient's general

care or chemotherapy treatment. Shared decision-making between the patient and the doctor was preferred.

Latent class analysis (LCA): pooled data

LCA was used to assess whether there were varying preferences by groups of respondents. Two-class and 3-class models were estimated. Based on the BIC, the 2-class model had a better fit than the 3-class model (BIC 3659.472 versus 3619.299 respectively). The 2-class model is reported in Table 5.10.

Looking further at the LCA results, over 60% of respondents were in class 2. Respondents in both classes 1 and 2 significantly preferred an assessment that asks about both minor and major nerve damage and did not significantly prefer having a questionnaire of any length.

Respondents in class 1 were also indifferent about the inclusion of a physical test and about whether or not there was shared decision-making when it came to the outcomes of assessment results. There was a particular concern with the length of time the CIPN assessment would take, with respondents in Class 1 unwilling to spend more than 10 minutes on top of usual clinic time on the assessment. Respondents in Class 2, on the other hand, felt the inclusion of a type of physical test was important. Respondents in this class also did not mind how long the assessment took, even if it meant a separate appointment up to 60 minutes long. This group of respondents also preferred shared decision-making over the situation where the doctor alone may change their chemotherapy treatment.

Table 5.10 LCA results

Log likelihood	−1703.1
Obs	2680
Iterations	468
AIC	3460
BIC	3619
Class 1	0.363
Class 2	0.637

LCA with two classes	Class 1 Estimate (S.E.)	Class 2 Estimate (S.E.)
S&Q 2 (symptoms & usual activities)	0.08(0.12)	−0.051(0.083)
Det 2 (minor and major changes)	−0.523(0.181)***	1.324(0.185)***
Q 2 (3 questions to answer)	0.179(0.223)	−0.111(0.159)
Q 3 (12 questions to answer)	−0.21(0.243)	0.163(0.16)
Q 4 (20 questions to answer)	−0.327(0.248)	0.278(0.159)
PhyT 2 (clinician administered test)	0.366(0.434)	0.705(0.275)**
PhyT 3 (patient activity based test)	0.161(0.262)	0.693(0.143)***
PhyT 4 (technical test)	0.318(0.433)	0.712(0.272)***
CT 2 (usual clinic time + 10 mins)	0.175(0.164)	−0.107(0.106)
CT 3 (usual clinic time + 30 mins)	−0.295(0.137)**	0.051(0.108)
CT 4 (separate appointment, takes up to 60 mins)	−0.864(0.207)***	0.066(0.125)
Res 2 (doctor may change your general care)	−0.034(0.241)	−0.314(0.17)
Res 3 (doctor may change your chemo/cancer treatment)	−0.317(0.184)	−0.258(0.109)**
Class 2		0.563(0.106)***

*** p < 0.001; **p < 0.01; *p < 0.05

Although the LCA model results are interesting, in terms of goodness of fit to data the MNL model still slightly outperformed the LCA (BIC 3601.3 and 3619.299 respectively). As a result, alternative models were explored to see whether there was a model with a better fit than the MNL model.

MXL model results: pooled data

An uncorrelated MXL model, which allows all parameters to be random, was estimated and was used to test for the presence of preference heterogeneity. Results are summarised in Table 5.11.

Table 5.11 MXL model with all parameters entered as random

Log likelihood	−1700
Obs	2680
Iterations	118
AIC	3455
BIC	3608
Draws	2000

MXL with all parameters	Estimate	S.E.	P-value
S&Q 2 Mean (symptoms & usual activities)	0.034	0.070	0.627
Det 2 Mean (minor and major changes)	0.875	0.117	0.000***
Q 2 Mean (3 questions to answer)	−0.020	0.101	0.845
Q 3 Mean (12 questions to answer)	−0.003	0.104	0.974
Q 4 Mean (20 questions to answer)	0.014	0.109	0.900
PhyT 2 Mean (clinician administered test)	0.652	0.103	0.000***
PhyT 3 Mean (patient activity based test)	0.584	0.097	0.000***
PhyT 4 Mean (technical test)	0.663	0.108	0.000***
CT 2 Mean (usual clinic time + 10 mins)	0.024	0.086	0.776
CT 3 Mean (usual clinic time + 30 mins)	−0.123	0.084	0.145
CT 4 Mean (separate appointment, takes up to 60 mins)	−0.355	0.106	0.001**
Res 2 Mean (doctor may change your general care)	−0.257	0.080	0.001**
Res 3 Mean (doctor may change your chemo/cancer treatment)	−0.313	0.070	0.000***
S&Q 2 SD	0.024	0.274	0.931
Det 2 SD	1.344	0.155	0.000***
Q 2 SD	0.031	0.282	0.913
Q 3 SD	0.157	0.366	0.667
Q 4 SD	0.633	0.149	0.000***
PhyT 2 SD	0.017	0.463	0.971
PhyT 3 SD	0.531	0.180	0.003**
PhyT 4 SD	0.413	0.186	0.026**
CT 2 SD	0.002	0.489	0.997
CT 3 SD	0.186	0.276	0.501
CT 4 SD	0.989	0.146	0.000***
Res 2 SD	0.438	0.173	0.011*
Res 3 SD	0.190	0.275	0.490

*** p < 0.001; **p < 0.01; *p < 0.05

The MXL model was re-estimated with only the attributes where at least one of the attribute level parameter estimates had significant standard deviations entered as random in the model. That is, all parameters except for the parameter associated with S&Q 2 were entered as random. By examining the AIC and BIC of the estimated models, it was determined that the MXL model with only select parameters entered as random was the model of best fit for the combined general population data, and this is therefore the focus of analysis in this section (AIC and BIC summarised in Table 5.12).

Table 5.12 Summary of AIC and BIC

Summary	MNL	LCA 2 classes	MXL (all random)	MXL (select random attributes)
Log likelihood	−1749	−1703	−1700	−1701
AIC	3525	3460	3455	3453
BIC	3601	3619	3608	3600

Results of the final MXL model are summarised in Table 5.13.

Table 5.13 Final MIXL model for pooled data

MXL model: select random attributes	Estimate	S.E.	P-value
S&Q 2 Mean(symptoms & usual activities)	0.035	0.07	0.613
Det 2 Mean (minor and major changes)	0.874	0.117	0.001***
Q 2 Mean (3 questions to answer)	−0.018	0.101	0.856
Q 3 Mean (12 questions to answer)	−0.003	0.103	0.975
Q 4 Mean (20 questions to answer)	0.019	0.109	0.863
PhyT 2 Mean(clinician administered test)	0.65	0.103	0.001***
PhyT 3 Mean (patient activity based test)	0.58	0.096	0.001***
PhyT 4 Mean (technical test)	0.661	0.108	0.001***
CT 2 Mean (usual clinic time + 10 mins)	0.023	0.086	0.787
CT 3 Mean(usual clinic time + 30 mins)	−0.125	0.084	0.136
CT 4 Mean (separate appointment, takes up to 60 mins)	−0.36	0.106	0.001***
Res 2 Mean (doctor may change your general care)	−0.257	0.079	0.001***
Res 3 Mean (doctor may change your chemo/cancer treatment)	−0.312	0.07	0.001***
Det 2 SD	1.344	0.154	0.001***
Q 2 SD	0.004	0.287	0.988
Q 3 SD	0.056	0.974	0.954
Q 4 SD	0.636	0.149	0.001***
PhyT 2 SD	0.065	0.498	0.896
PhyT 3 SD	0.539	0.181	0.003**
PhyT 4 SD	0.417	0.183	0.023*
CT 2 SD	0.009	0.509	0.985
CT 3 SD	0.104	0.423	0.806
CT 4 SD	0.979	0.146	0.001***
Res 2 SD	0.428	0.177	0.016*
Res 3 SD	0.175	0.293	0.55

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

The mean parameter estimates were very similar to the corresponding parameter estimates in the MNL model in terms of the pattern of preferences exhibited by respondents. Furthermore, parameter estimates with significant standard deviations were mostly indicative of respondents holding different strengths of preference rather than opposing preferences.

To examine the standard deviations further, density plots overlaid with the conditional mean values of respondents were used and are summarised in Figures 5.4 to 5.8. Figure 5.4 is a kernel density plot of the parameter estimate for an assessment that picks up both minor and major changes in CIPN (Det 2). The majority of the density curve lies above zero, as indicated by the red dotted vertical line, indicating that the majority of respondents do prefer an assessment that picks up both minor and major changes in their condition. Two peaks in the density curve were noted, clustered around the values 1 and 2 respectively. This indicates that there was one group of respondents that held a stronger preference for this parameter than the other.

Figure 5.5 is very similar to Figure 5.4, in the sense that the standard deviation is an indication of the strength of preference. The density curve mostly lies below zero, with two peaks in the curve. In other words, respondents generally preferred the CIPN assessment to take place during usual clinic time rather than requiring a separate appointment up to 60 minutes long. Respondents were generally clustered into two groups, with one group holding a stronger preference than the other. A significant standard deviation was also noted for the parameter of the level *the doctor may change your general care* (see Figure 5.6). However, this was also an indication of strength of preference with the vast majority of the density curve below zero, indicating that respondents generally preferred shared decision-making.

The kernel density plot in Figure 5.7 is very different. In particular, respondents appeared to be in disagreement about whether the inclusion of a questionnaire with 20 questions was preferred to having no questionnaire. Slightly over half of the density curve lies above zero, with slightly under half below zero. In other words, a slight majority of respondents support the inclusion of a questionnaire with 20 questions, while a large minority do not support it.

The parameters for a patient activity based test and technical test are plotted together in Figure 5.8. Both density curves are above zero, but the shapes of the density curves are very different. The density curve for patient activity-based test is very wide and flat, indicating that although a patient activity-based is preferred over no physical test, the strength of preference varies widely among respondents. In contrast, respondents have greater agreement in strength of preference for a technical test, as the density curve is much narrower with a distinct peak.

Figure 5.4 Plot of Det 2 density and conditional mean values

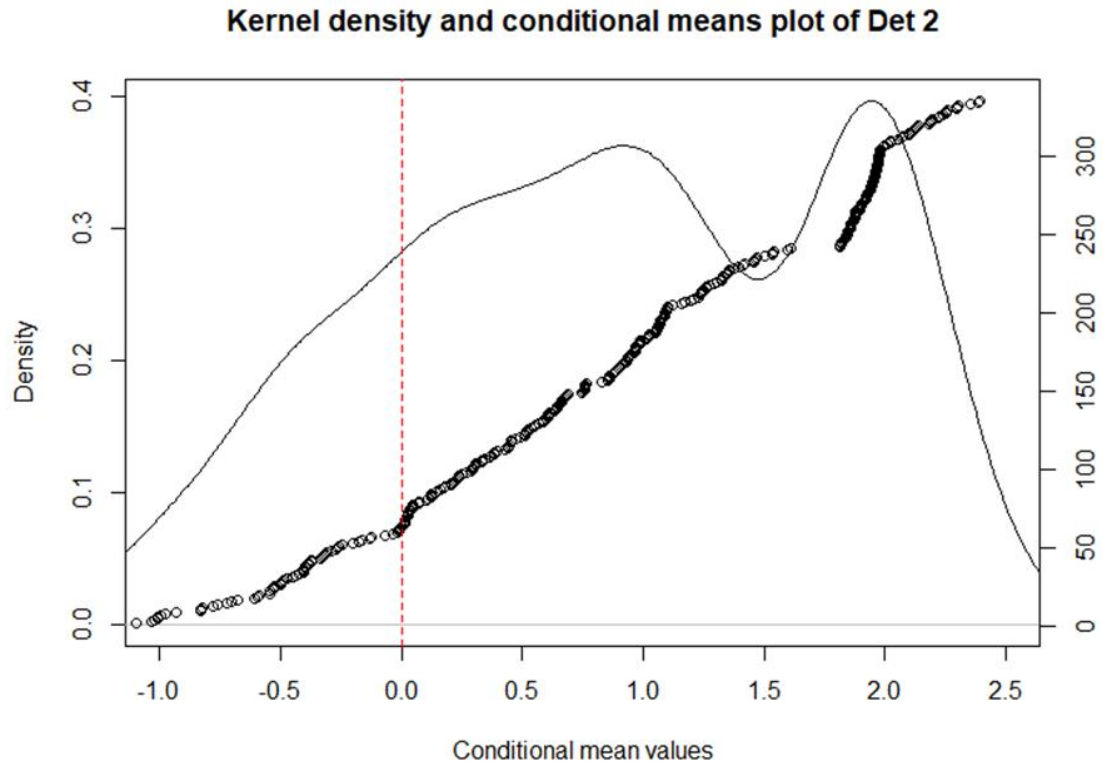


Figure 5.5 Plot of CT 4 density and conditional mean values

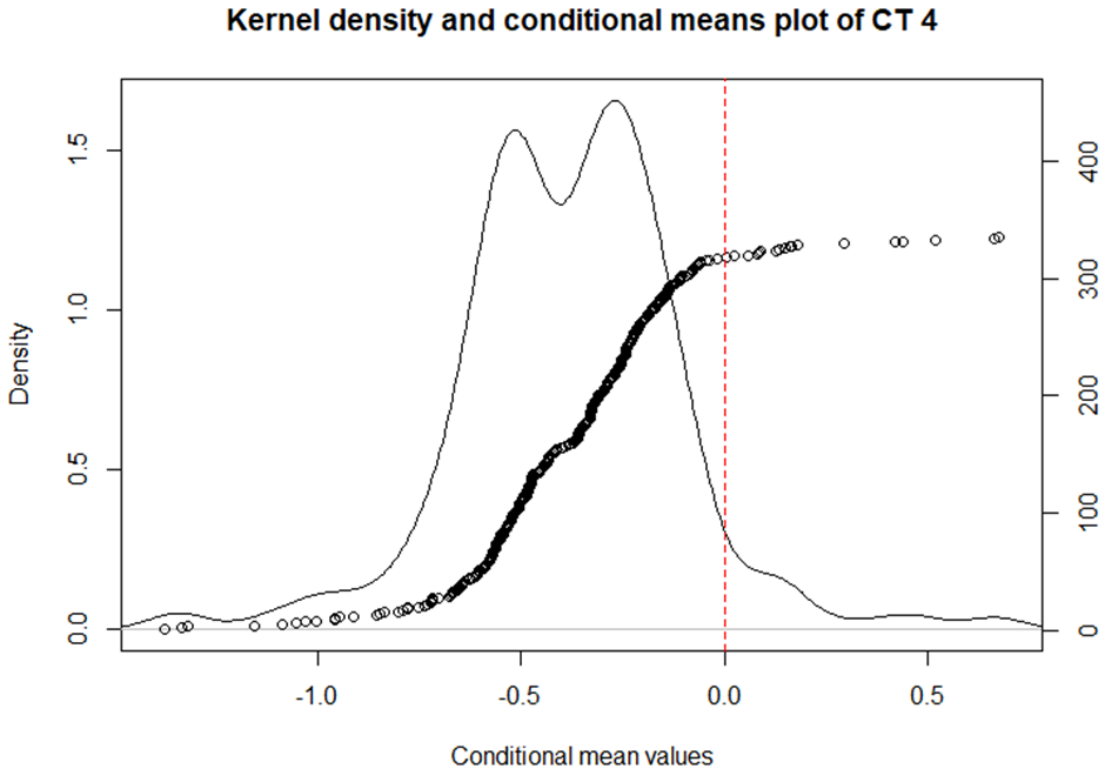


Figure 5.6 Plot of Res2 density and conditional mean values

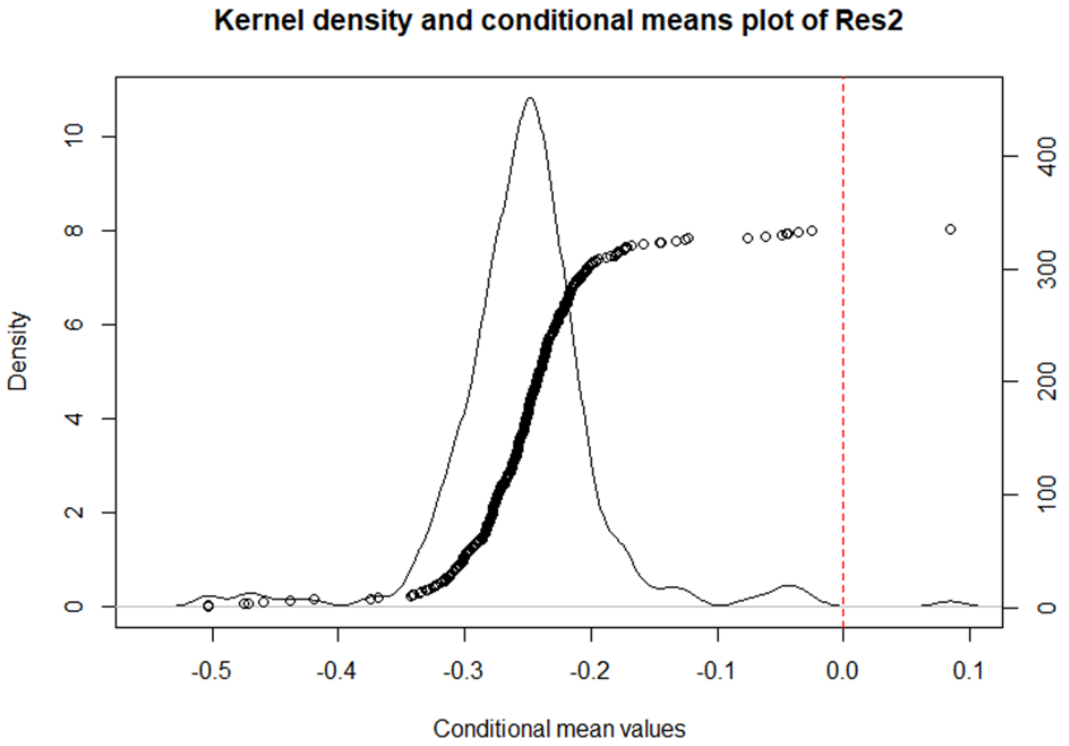


Figure 5.7 Plot of Q 4 density and conditional mean values

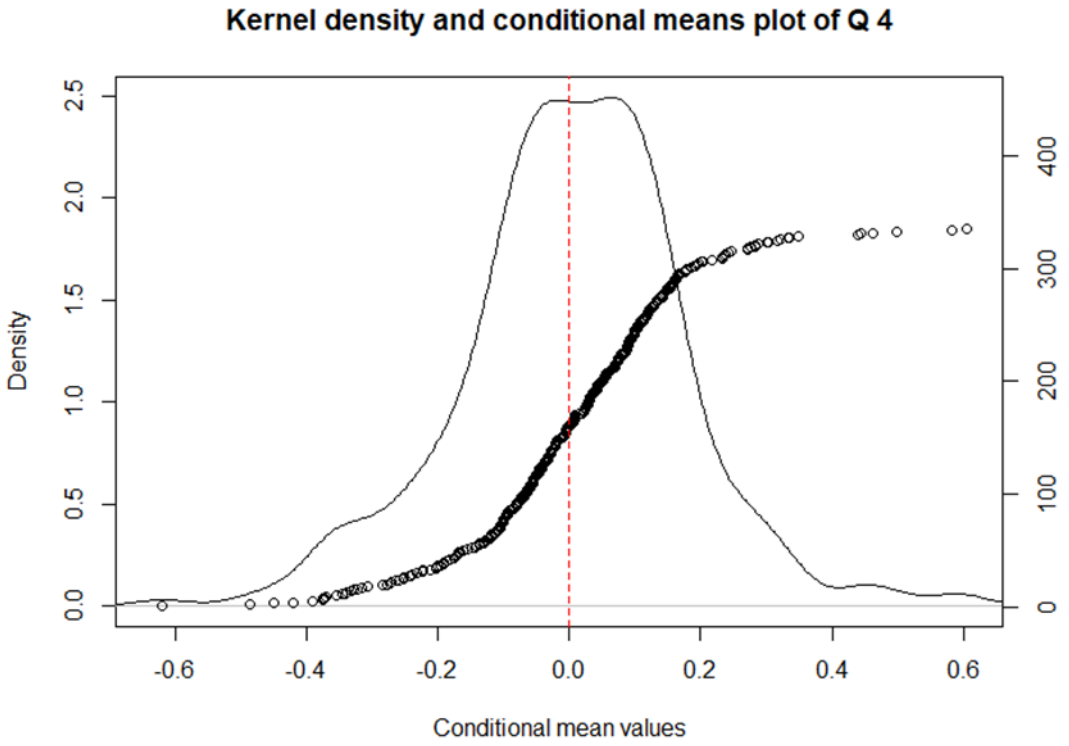
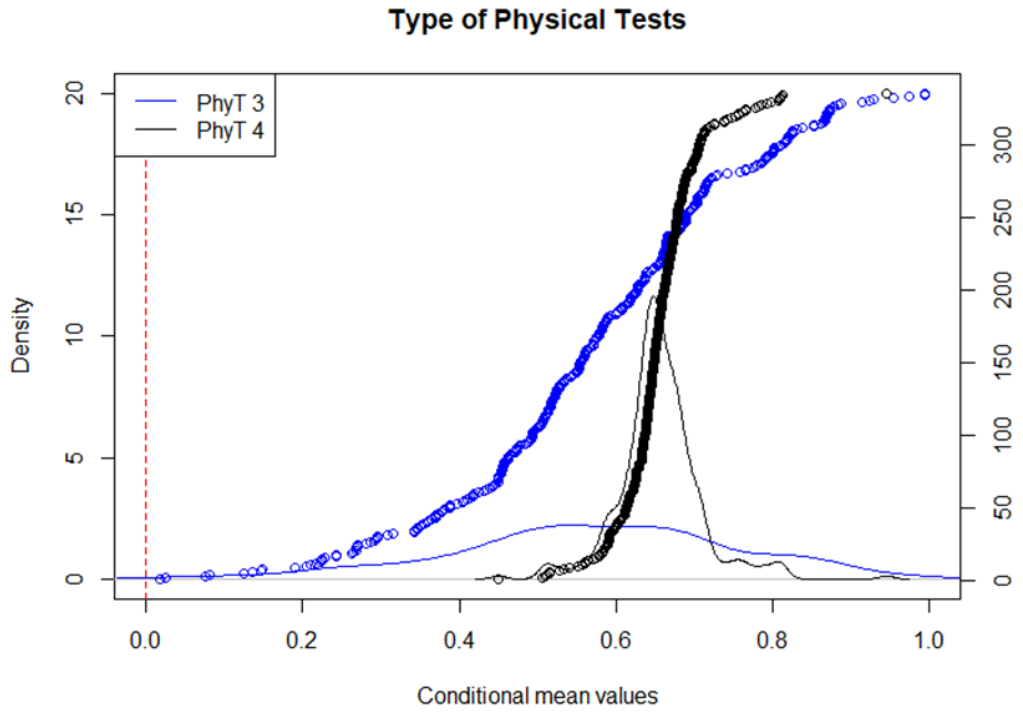


Figure 5.8 Plot of PhyT 3 and 4 density and conditional means



5.3.3 Responses to evaluation questions

After completing the eight assigned choice sets, respondents were asked a series of questions about the introductory information and the choice sets. Respondents were asked whether the introductory information was helpful to them; answers are summarised in Table 5.14. The vast majority of respondents found the introductory information helpful, with 85% of Arm 1 and 91% of Arm 2 selecting 'yes'. A small minority of respondents in both arms already knew about CIPN prior to this survey.

Table 5.14 Helpfulness of introductory information

Helpfulness of introduction information*	Arm 1 (%)	Arm 2(%)
Yes	142 (85%)	153 (91%)
No	6 (4%)	6 (4%)
No, I already knew this information prior to this survey	6 (4%)	6 (4%)
No, it was confusing	8 (5%)	2 (1%)
I don't remember	5 (3%)	1 (1%)

*Actual question: Prior to the choice questions, some information was provided to define peripheral neuropathy and to explain some key definitions and terms. Did this information help you understand the choice questions?

Respondents in Arm 2 were asked additional questions in relation to the video they had been asked to watch. The majority (98%, n = 164) were able to view the full length of

the video, and only four respondents experienced problems with viewing the video. For three of these respondents, this was due to lag in the video and for one respondent there was no sound.

Respondents who were able to view the full video were asked to rate video quality on a 5-point Likert scale, with answers summarised in Table 5.15. Eighty-nine per cent of these respondents reported that video quality was 'good' or 'excellent', with no respondent rating below 'okay'.

Table 5.15 Arm 2 video quality rating

Video Quality Rating	No. (%)
Okay	18 (11%)
Good	65 (40%)
Excellent	81 (49%)

Respondents were also asked about how much the introductory information assisted them in understanding the different types of physical tests. A summary of responses is provided in Table 5.16. To test for whether respondents' level of agreement depended on the arm to which they were assigned, a chi-square test and asymptotic linear by linear association test (ALbl) were used. The latter test was used as this test is sensitive to the ordinal nature of the 5-point Likert scale. The p-values from the tests are summarised in Table 5.17 below.

Table 5.16 Level of agreement with helpfulness of the physical test type explanations*

Frequencies (%)		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Sharp and dull test	Arm 1	2 (1%)	5 (3%)	19 (11%)	91 (55%)	50 (30%)
	Arm 2	1 (1%)	4 (2%)	13 (8%)	75 (45%)	75 (45%)
Tuning fork test	Arm 1	2 (1%)	4 (2%)	27 (16%)	85 (51%)	49 (29%)
	Arm 2	1 (1%)	3 (2%)	18 (11%)	80 (48%)	66 (39%)
Peg board	Arm 1	2 (1%)	4 (2%)	22 (13%)	84 (50%)	55 (33%)
	Arm 2	2 (1%)	2 (1%)	20 (12%)	68 (41%)	76 (45%)
Sway test	Arm 1	2 (1%)	4 (2%)	17 (10%)	92 (55%)	52 (31%)
	Arm 2	2 (1%)	3 (2%)	17 (10%)	77 (46%)	69 (41%)
Nerve conduction study	Arm 1	3 (2%)	4 (2%)	28 (17%)	77 (46%)	55 (33%)
	Arm 2	2 (1%)	3 (2%)	15 (9%)	77 (46%)	71 (42%)

* Actual Question: For each of the following types of physical tests, would you agree that the information provided prior to the choice questions helped you understand what they were?

Table 5.17 Testing for whether level of agreement with helpfulness of physical test explanations is dependent on arm

P-values	Chi-square test	Asymptotic linear by linear association (ALbl) test
Sharp and dull test	0.088	0.013*
Tuning fork test	0.294	0.033*
Peg board	0.214	0.071
Sway test	0.425	0.192
Nerve conduction study	0.178	0.027*

*p < 0.05; **p < 0.01; *** p < 0.001

For three of the five physical tests, there was a significant difference between arms ($p < 0.05$). Respondents in Arm 2 were more likely to strongly agree that information provided prior to the choice questions helped them understand the sharp and dull test, tuning fork test and nerve conduction study.

Respondents were also asked to rate their ability to imagine living with certain symptoms of CIPN and living with impacts to their daily lives resulting from CIPN symptoms. Table 5.18 and Table 5.19 provide a summary.

Table 5.18 Level of agreement with ability to imagine symptoms

Frequencies (%)		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Numbness	Arm 1	4 (2%)	10 (6%)	19 (11%)	99 (59%)	35 (21%)
	Arm 2	2 (1%)	6 (4%)	24 (14%)	85 (51%)	51 (30%)
Pins and needles	Arm 1	3 (2%)	7 (4%)	21 (13%)	101 (60%)	35 (21%)
	Arm 2	0 (0%)	8 (5%)	21 (13%)	85 (51%)	54 (32%)
Burning sensations	Arm 1	3 (2%)	13 (8%)	28 (17%)	93 (56%)	30 (18%)
	Arm 2	3 (2%)	9 (5%)	36 (21%)	75 (45%)	45 (27%)
Balance problems	Arm 1	3 (2%)	6 (4%)	22 (13%)	104 (62%)	32 (19%)
	Arm 2	3 (2%)	8 (5%)	23 (14%)	80 (48%)	54 (32%)
Muscle weakness	Arm 1	3 (2%)	11 (7%)	21 (13%)	97 (58%)	35 (21%)
	Arm 2	4 (2%)	7 (4%)	24 (14%)	83 (49%)	50 (30%)
Constipation	Arm 1	4 (2%)	7 (4%)	28 (17%)	87 (52%)	41 (25%)
	Arm 2	2 (1%)	11 (7%)	36 (21%)	73 (43%)	46 (27%)
Decreased reflexes	Arm 1	5 (3%)	13 (8%)	22 (13%)	96 (57%)	31 (19%)
	Arm 2	2 (1%)	9 (5%)	38 (23%)	84 (50%)	35 (21%)

*Actual question: In the context of peripheral neuropathy, I could imagine what it would feel like to live with

Table 5.19 Level of agreement with ability to imagine impacts on daily activities*

Frequencies (%)		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Trouble picking up/holding things	Arm 1	6 (4%)	10 (6%)	25 (15%)	90 (54%)	36 (22%)
	Arm 2	2 (1%)	9 (5%)	16 (10%)	87 (52%)	54 (32%)
Trouble buttoning clothes	Arm 1	4 (2%)	8 (5%)	29 (17%)	89 (53%)	37 (22%)
	Arm 2	2 (1%)	8 (5%)	20 (12%)	85 (51%)	53 (32%)
Trouble while walking	Arm 1	4 (2%)	6 (4%)	24 (14%)	96 (58%)	37 (22%)
	Arm 2	1 (1%)	10 (6%)	18 (11%)	86 (51%)	53 (32%)

*Actual question: In the context of peripheral neuropathy, I could imagine what it would feel like to live with

Chi-square tests and ALbl tests were used to test whether level of agreement with ability to imagine living with CIPN symptoms and impact on daily life depended on arm assignment. The p-values from these tests are summarised in Table 5.20. In general, no significant differences between arms were detected except for *trouble picking up/holding things*. Those in Arm 2 were significantly more likely ($p < 0.05$) to strongly agree that they could imagine what it would feel like to live with such a difficulty.

Table 5.20 Testing for whether ability to imagine is dependent on arm

P-values	Chi-square test	Asymptotic linear by linear association test (ALbl)
Numbness	0.179	0.111
Pins and needles	0.075	0.078
Burning sensations	0.155	0.359
Balance problems	0.060	0.273
Muscle weakness	0.291	0.294
Constipation	0.397	0.767
Decreased reflexes	0.120	0.751
Trouble picking up/holding things	0.104	0.014*
Trouble buttoning clothes	0.262	0.053
Trouble while walking e.g. tripping or stumbling	0.133	0.139

*$p < 0.05$ **$p < 0.01$ *** $p < 0.001$

Respondents were also asked about the strategy they used to choose between Assessment tools A and B in each choice question (see Table 5.21). A chi-square test was carried out to test for whether strategy of choice was associated with arm assignment, this was not significant ($p = 0.302$). In relative terms, respondents in Arm 1 and Arm 2

were similar in terms of *consideration of all features* being the most frequently chosen option followed by *consideration of all features that were different*. There were four respondents in Arm 1 and 1 respondent in Arm 2 who selected *other*. However, upon further inspection of comments, these respondents were categorised as using the strategy of considering only features that were most important to them. There was one respondent in Arm 1 who selected *other* with only the comment 'none'.

Table 5.21 Strategy for choosing preferred assessment

Statement	Arm 1 (%)	Arm 2 (%)
I considered all features for each assessment option	74 (44%)	84 (50%)
I considered all features which were different between the two assessment options	63 (38%)	62 (37%)
I considered only the features which were most important to me	29 (17%)	22 (13%)
Other	1 (1%)	0 (0%)
Total	167 (100%)	168 (100%)

Respondents were asked about how easily they could identify the differences between each assessment tool as well as how easy it was for them to choose between Assessment A and B in each choice question. Frequency of responses are summarised in Table 5.22 and Table 5.23.

Table 5.22 Level of agreement: ease of identifying differences between assessment tools*

Frequency (%)	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Arm 1	3 (2%)	9 (5%)	26 (16%)	91 (55%)	38 (23%)
Arm 2	1 (1%)	1 (1%)	20 (12%)	93 (55%)	53 (32%)

*Actual question: Most of the time, I could easily identify the differences between assessment options.

Table 5.23 Level of agreement: ease of choosing a preferred assessment tool*

Frequency (%)	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Arm 1	4 (2%)	8 (5%)	35 (21%)	92 (55%)	28 (17%)
Arm 2	2 (1%)	10 (6%)	25 (15%)	86 (51%)	45 (27%)

* Actual question: Most of the time, I could easily choose between the assessment options

Table 5.24 Testing for whether ease of identifying/choosing is dependent on arm

P-values	Chi-square test	Asymptotic linear by linear association test (ALbl)
Ease of identifying differences	0.030*	0.003***
Ease of choosing an assessment tool	0.152	0.068

*p < 0.05 **p < 0.01 *** p < 0.001

Chi-square and ALbl tests were performed to test if there were any differences by arm (see Table 5.24). Respondents in Arm 2 were more likely to agree that most of the time they could easily distinguish between assessment tools. This difference was identified with the chi-square test ($p < 0.05$) and, after accounting for the ordinal nature of the Likert scale, the ALbl test found this difference to be even stronger ($p < 0.01$). Although respondents in Arm 2 could more easily identify differences between assessment tools, they did not find it any easier to choose between tools than did respondents in Arm 1.

5.3.4 Experience with illness and cancer

Respondents were also asked about whether they suffered from a range of physical or mental conditions (see Table 5.25). The list of conditions is based on the Health Conditions section from the National Health Survey 2017–18 (Australian Bureau of Statistics, 2018). A chi-square test was used to test whether there were any differences between arms when it came to experience with the specified conditions and no significant difference was found ($p = 0.636$). Over half (54%) of Arm 1 and slightly over half (52%) of Arm 2 did not suffer from any of the conditions listed. Eight respondents from Arm 1 and 5 respondents from Arm 2 indicated they had cancer.

Those who indicated they had had cancer were asked about their cancer type; results are summarised in Table 5.26. Multiple diagnoses were common. Respondents were also asked about whether someone close to them had ever had cancer (see Table 5.27). Over half of respondents in Arm 1 and Arm 2 knew someone who had been diagnosed with cancer. In terms of the cancer type, this varied widely, with skin cancer, colorectal cancer, breast cancer and lung cancer selected most frequently. A summary is provided in Table 5.28.

Table 5.25 Frequency of conditions

Condition	Arm 1	Arm 2
Asthma	18 (11%)	21 (13%)
Arthritis	13 (8%)	21 (13%)
Cancer	8 (5%)	5 (3%)
Cardiovascular disease	9 (5%)	5 (3%)
Diabetes and high sugar levels	12 (7%)	7 (4%)
Kidney disease	2 (1%)	2 (1%)
Osteoporosis	4 (2%)	6 (4%)
Mental, behavioural and cognitive conditions	10 (6%)	13 (8%)
None of these	91 (55%)	88 (52%)
Total	167 (100%)	168 (100%)

Table 5.26 Frequency of cancer diagnosis type

Cancer diagnosis (can select more than one)	Arm 1	Arm 2
Breast cancer	1	4
Cancer of other female reproductive organs (including uterus, ovary)	1	0
Cancer of unknown primary site	1	0
Cervical cancer	2	0
Colon/rectum/bowel cancer (colorectal)	3	1
Leukaemia	1	0
Lung cancer (including trachea, pleura and bronchus)	1	0
Metastatic Bone Cancer	1	0
Non-Hodgkin lymphoma	1	0
Other type of lymphoma	1	0
Prostate Cancer	1	0
Skin cancer (Basal cell carcinoma)	3	1
Skin cancer (melanoma)	3	1
Skin cancer (other)	1	0
Total	21	7

Table 5.27 Someone you know with cancer?

Has someone close to you ever received a diagnosis for cancer?	Arm 1	Arm 2
Yes	95 (57%)	92 (55%)
No	72 (43%)	76 (45%)
Total	167 (100%)	168 (100%)

Table 5.28 What was their cancer diagnosis?

What was their cancer diagnosis (could select more than one)	Arm 1	Arm 2
Colon/rectum/bowel cancer (colorectal)	27	17
Breast cancer	20	23
Prostate cancer	13	9
Lung cancer (including trachea, pleura and bronchus)	16	10
Cervical cancer	7	3
Cancer of other female reproductive organs (including uterus, ovary)	7	2
Bladder/kidney cancer	2	6
Stomach cancer	7	9
Leukaemia	5	4
Non-Hodgkin lymphoma	5	2
Other type of lymphoma	3	0
Cancer of unknown primary site	3	0
Other cancer	10	9
Skin cancer (melanoma)	25	28
Skin cancer (basal cell carcinoma)	18	16
Skin cancer (squamous cell carcinoma)	5	11
Skin cancer (other)	1	0
Skin cancer (I don't know)	7	4
Total	181	153

5.4 Comparison of Study 1 and Study 2: patient and general population preferences

5.4.1 Overview

Now that Study 1 and 2 have been described and analysed separately, this section compares results from these studies. The goal of this section is to compare the preferences of the patient and general population samples.

5.4.2 Investigating the impact of extra information on scale differences between the patient and general population sample

One of the aims of the CIPN DCE project was to investigate whether the addition of extra information to the general population would reduce differences with a patient sample. This involves a comparison of results between the general population arms and the patient sample.

As the information received in each arm was slightly different, the comparison of the patient with the general population samples will be done separately by arm. Arm 1, of the general population sample, which received the same information as the patient sample, will be compared to the patient sample. Arm 2 of the general population sample, which received additional information, will also be compared to the patient sample. The impact of the extra information received in Arm 2 will be explored. For instance, this section also investigates whether Arm 2 preferences were more similar to the preferences of the patient sample than Arm 1 preferences.

As above, a S-MNL model was estimated to investigate the possibility of combining the patient and general population samples for analysis; this model is summarised in Table 5.29. This was used to test whether there were scale differences between the patient sample and the two arms of the general population sample.

In this model, the patient sample was used as the base with a separate parameter for Arm 1,δ_{arm1}, and Arm 2, δ_{arm2}. That is, Arm 1 and Arm 2 are each compared to the patient sample for scale differences.

Table 5.29 S-MNL: patient and general population samples

Log likelihood	−2300
Obs	3616
Iterations	6
Draws	10000
AIC	4623
BIC	4722
Estimation procedure	BHHH

S-MNL	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	0.205	0.081	0.011**
Det 2 (minor and major changes)	1.195	0.167	0.000***
Q 2 (3 questions to answer)	0.129	0.113	0.254
Q 3 (12 questions to answer)	0.246	0.114	0.031*
Q 4 (20 questions to answer)	0.165	0.124	0.183
PhyT 2 (clinician administered test)	0.883	0.164	0.000***
PhyT 3 (patient activity based test)	0.840	0.130	0.000***
PhyT 4 (technical test)	0.884	0.151	0.000***
CT 2 (usual clinic time + 10 mins)	0.090	0.107	0.397
CT 3 (usual clinic time + 30 mins)	−0.049	0.104	0.634
CT 4 (separate appointment, takes up to 60 mins)	−0.304	0.104	0.004**
Res 2 (doctor may change your general care)	−0.540	0.101	0.000***
Res 3 (doctor may change your chemo/cancer treatment)	−0.607	0.105	0.000***
τ	0.994	0.161	0.000***
δ_{arm1}	−0.685	0.162	0.000***
δ_{arm2}	−0.337	0.160	0.036*

*** p < 0.001; **p < 0.01; *p < 0.05

The het parameters were significant for both Arm 1 (p < 0.001) and Arm 2 (p < 0.05), indicating that there were scale differences between each of the general population arms compared to the patient sample. As a result, the patient sample could not be combined with either of the arms of the general population sample for analysis. The S-MNL model was re-estimated with the general population arms combined to investigate whether a larger sample size would reduce the scale difference (see Appendix 5F). This was possible as no significant scale differences (p > 0.05) was detected between the two general population arms. However, the δ parameter associated with the general population sample was still significant (p < 0.001), indicating the presence of scale differences between the two samples.

It is interesting that the parameter estimate of $\delta_{arm1}$was about twice the size of that of δ_{arm2}, relative to the patient sample. In other words, the scale difference between Arm 1

and the patient sample was larger than the scale difference between Arm 2 and the patient sample. This suggests that the extra information provided in Arm 2 contributed to reducing the scale difference between the general population sample and the patient sample. Nevertheless, the general population arms did exhibit significant scale differences with the patient sample.

5.4.3 Comparison of patient and general population results: using ratios

In this section, the goal was to compare results from the patient and general population samples. The impact of extra information was not the focus of this analysis, hence, it was not deemed necessary to separate Arms 1 and 2 of the general population sample for analysis. The general population arms were pooled, made possible by the lack of scale differences, and compared to the patient sample.

Due to the scale difference between the patient and general population samples, model results cannot be directly compared. However, ratios of parameter estimates between the respective samples can be used to investigate whether differences are due to scale alone or whether there were actual preference differences (Fiebig et al., 2009; Vass, Wright et al., 2018). This analysis was done according to the explanation given in Fiebig et al. (2009). Ratios were calculated to give an idea of whether differences were due to preferences and/or scale.

This is done by comparing the magnitude of ratios, with ratios closer in magnitude indicating the presence of a preference difference, after correcting for scale differences. The ratio is calculated by dividing parameter estimates from the patient sample with those from the general population sample, as summarised in Table 5.30.

Parameter estimates used were from the respective MXL models (models of best fit to the respective data). Only parameter estimates that were significant ($p < 0.05$) in both samples were compared. Ratios varied in magnitude across the parameters indicating the presence of both scale and preference differences between the two samples.

Table 5.30 Ratios of parameters

Mean estimates and ratios	Patients	Genpop*	Ratio**
Det 2 Mean (minor and major changes)	1.738	0.867	2.005
PhyT 2 Mean (clinician administered test)	1.054	0.644	1.637
PhyT 3 Mean (patient activity based test)	1.425	0.573	2.487
PhyT 4 Mean (technical test)	1.063	0.653	1.628
Res 2 Mean (doctor may change your general care)	−0.860	−0.254	3.386
Res 3 Mean (doctor may change your cancer treatment)	−0.985	−0.309	3.188

*Genpop = general population sample;
**patients divided by general population sample parameter estimate

An alternative method to examine whether differences between samples were due more to scale versus actual preference differences was to compare results by calculating ratios of marginal rates of substitution (MRS) (Fiebig et al., 2009). Firstly, the MRS across parameter estimates in each sample was calculated using the parameter for level of detail (Det 2) as the denominator. The denominator was chosen arbitrarily, as the purpose was to allow for comparison of samples rather than interpretation of the MRS value itself. The denominator is then set equal to 1 to allow for comparison of the ratios of MRS.

The ratios of MRS were calculated by dividing the patient sample MRS by the general population sample MRS, and are summarised in Table 5.31. The ratios of MRS that were not close to 1 suggest that the difference between the two samples was not due solely to scale. The ratios of MRS for the parameters *patient activity-based test*, *the doctor may change your general care* and *the doctor may change your cancer treatment* were larger than 1. In other words, there was more heterogeneity in preferences in the general population for these parameters compared to the patient sample.

In contrast, the ratio of MRS for *clinician administered test and technical test* were less than 1, suggesting the patient sample had more heterogeneity in preferences for these particular parameters than did the general population sample.

Table 5.31 Ratios of MRS

Ratios of substitution	Patients	Genpop	Ratio**
Det 2 Mean (minor and major changes)	1	1	1
PhyT 2 Mean(clinician administered test)	0.6064	0.7428	0.8164
PhyT 3 Mean (patient activity based test)	0.820	0.661	1.241
PhyT 4 Mean (technical test)	0.612	0.753	0.812
Res 2 Mean (doctor may change your general care)	−0.495	−0.293	1.689
Res 3 Mean (doctor machange your cancer treatment)	−0.567	−0.356	1.590

*Genpop = general population sample **patients divided by general population sample value

5.4.4 Comparison by utility ranking of possible assessment tools

Apart from using ratios, patient and general population findings were compared in terms of the combination of attribute levels that gave respondents the highest level of utility. The utility of each possible combination of attribute levels – that is, each possible assessment tool – were calculated and ranked for the patient and general population samples. There was a total of 768 possible assessment tools. The utility for each of these possible assessment tools was calculated separately for the patient and general population samples.

Calculations were made in Mathematica using code written by Street (2020a). To calculate the utility of each possible assessment tool, parameter estimates were taken from the models of best fit for the respective samples. If the parameter estimate was not significantly different from 0 ($p > 0.05$), a 0 was used for its utility value. A list of all the potential assessment tools was then created from the possible attribute level combinations. Each attribute level was dummy coded as had been done for the estimation of the parameters. Each attribute level combination was then multiplied by the transpose of the parameter estimates to calculate the utility of each potential assessment tool. Each assessment tool was then ranked by its utility value.

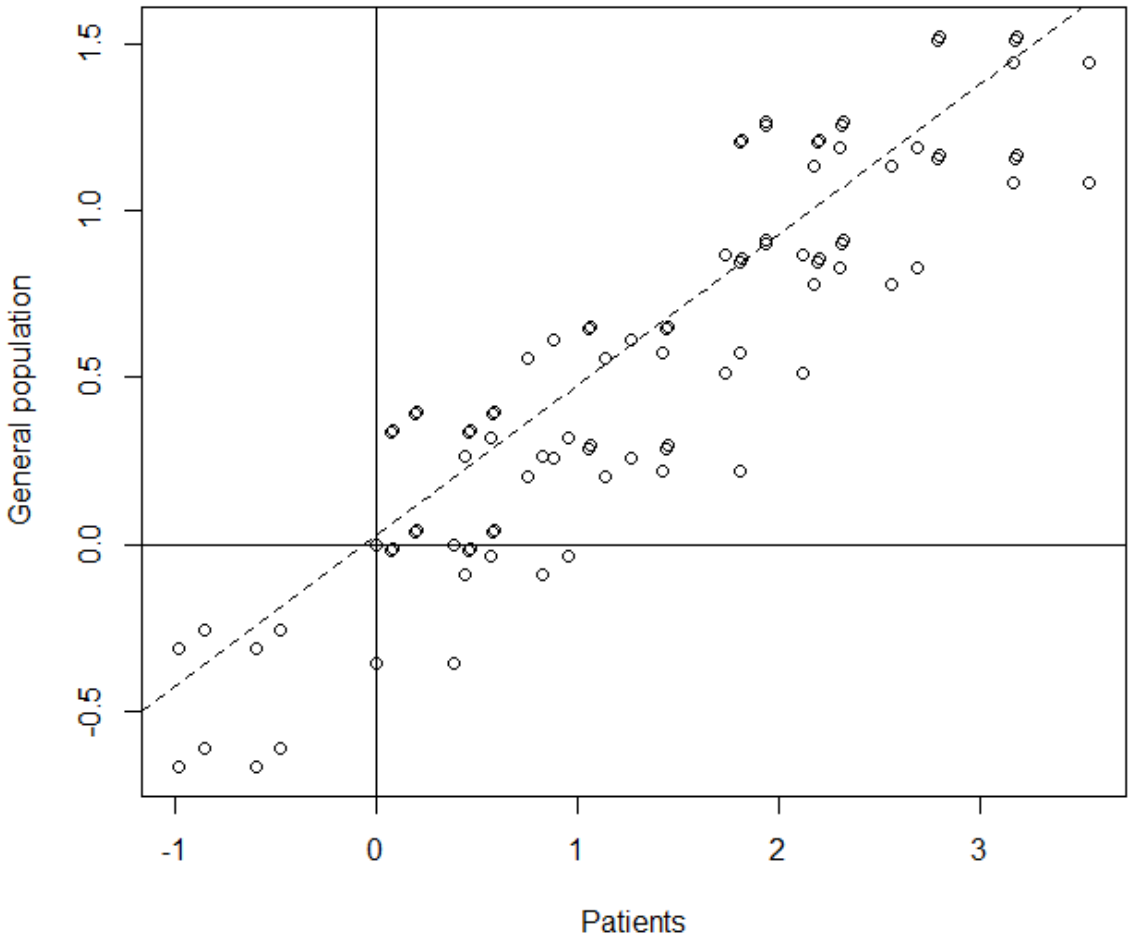
Figure 5.9 provides a graphical summary where the utility from any given assessment, from the patient and general population sample, represented by the dots, are plotted against each other. Vertical and horizontal solid black lines have been drawn to indicate where 0 lies on the scale. The dotted diagonal line is the line of best fit to the data using the lm function in R Studio. The darker dots represent overlap in the utility values for different assessment tool combinations. Along the horizontal axis, the patient sample had a wider range of utility values from -0.985 to 3.548 whereas utility values for the general population sample ranged from -0.664 to 1.52. This is partly reflective of the finding of significant scale differences between the patient and general population samples.

It was of note that there are some assessment tools that the patient sample preferred but were not preferred by the general population sample. For instance, looking at the utility value of 1 along the horizontal axis and where it intersects the solid line of 0.0 along the vertical axis. There is an assessment tool that gives the patient sample a

positive utility value of 1 i.e. this is an assessment tool that the patient sample would prefer. For the general population sample, this particular assessment tool gives a negative utility, or rather, would not be preferred.

There were 22 assessment tools that gave the highest level of utility in the patient sample, relative to the other possible assessment tools. Assessment tools that had the highest utility always had the features of asking about symptoms' impact on usual activities; picking up both minor and major changes in CIPN, regardless of whether it was important or not; a patient activity-based test; and the patient and doctor deciding together what results mean for care or treatment. As long as these features were included in the assessment tool, the patient sample was indifferent about the length of questionnaire and impact of assessment on clinic time.

Figure 5.9 Scatterplot of utility values



In contrast, there were 29 assessment tools that gave the highest level of utility for the general population sample. These assessment tools had the features of picking up both minor and major changes in CIPN, a technical test and shared decision-making about

what results mean for care or treatment. If these features were included in the assessment tool, respondents were indifferent about whether the assessment asks about symptoms or impact on usual activities, length of questionnaire and impact on clinic time.

The patient sample were comparatively more selective about the types of assessment tools they were satisfied with. There were fewer possible assessment tools from which the patient sample obtained the highest utility compared to the general population (22 versus 29 respectively). It was also noted that the assessment tools that gave the patient sample the highest utility, always asked about symptoms impact on usual activities. Whereas for the general population sample, assessments that gave the highest utility could ask about symptoms on usual activities or symptoms alone.

5.5 Discussion

In this chapter, general population preferences for the features of a CIPN assessment tool were examined. The impact of providing more information was also explored by comparing preferences between the two general population arms. In one arm, respondents received the same amount of information as the patient sample, while the other arm received additional information. These arms were also compared against the patient sample findings to identify whether the provision of extra information leads to more similar preferences between these two populations.

5.5.1 Study 2: Summary of findings

In general, the general population sample was indifferent about whether the assessment asks about symptoms or symptoms' impact on usual activities. In other words, they were not concerned about the assessment picking up quality-of-life impacts of CIPN. However, they definitely preferred to have the assessment pick up small changes in their condition whether they were significant or not. In terms of the actual assessment, including an objective measure of some type was strongly supported by respondents. Whether or not respondents had to complete a questionnaire was not important to them. Respondents also valued including a CIPN assessment as part of their clinical consultations and were willing to devote up to 30 minutes of extra clinic time for this. The general population sample was also strongly supportive of the notion

of patient involvement in decision-making when it came to the impact of the CIPN assessment results on their care or treatment.

The general population sample exhibited heterogeneous preferences for many of the attributes. The majority of this heterogeneity was in regards to the strength of preference rather than a different direction of preferences, with the exception of the parameter for the inclusion of an assessment with a questionnaire including 12 questions. Respondents were divided almost equally, with half preferring to have such an assessment as opposed to having no questionnaire and the other half preferring to have no questionnaire, seeing a questionnaire with 12 questions as a negative feature.

Results from Study 2 suggest that while the inclusion of more information improved respondent comprehension of the DCE task, the level of information did not necessarily influence preferences. In particular, the inclusion of moving images and a short video did not lead to any significant differences in estimated parameters between arms within the general population sample. However, those in Arm 2 did report finding it easier, compared to those in Arm 1, to identify differences between assessment options when completing choice sets. This suggests the extra information provided to respondents in Arm 2 did improve understanding of choice sets.

The findings from Study 2, are relevant to DCE researchers. These findings provide evidence that it is important to explain unfamiliar terminology to respondents. This can increase the researcher's confidence in results as it can help respondents make more informed choices. Evidence from Study 2 also supports the notion that provision of extra information can be beneficial to respondents without necessarily influencing the preferences they exhibit. This is desirable for many health DCE researchers, as it is often necessary to explain terminology to respondents, but undesirable for this explanation to influence preferences.

5.5.2 Comparison of Study 1 and 2: summary of main findings

Scale differences prevented the direct comparison of the general population results with the patient sample results, but it was notable that the provision of extra information in Arm 2 had the effect of reducing the scale difference between the general population and the patient samples. This suggests that providing extra information can

narrow the difference between naïve respondents and those with experience or knowledge of a particular health condition.

Although results between the general population and patient samples were not directly comparable due to the scale difference, indirect comparisons could be made. In particular, there was evidence that the difference between the two samples was due not only to scale but also to actual differences in preferences. Although there was overlap in the type of assessment tools that the patient and general population samples considered the 'best', there were also key differences. For instance, although both samples preferred the inclusion of a physical test, the patient sample preferred a patient activity-based test, which requires active participation on the part of the patient. In contrast, the general population sample preferred a technical test.

Another key difference was the importance the patient sample placed on the assessment asking about the impact of symptoms on usual activities, which was not considered important by the general population sample. In other words, the patient sample was more conscious of the quality-of-life implications of CIPN than the general population sample.

Regardless of population type, the inclusion of a physical test as part of the assessment and shared decision-making when it comes to implications of assessment results on care and treatment was valued. This is an important finding for clinicians, particularly when it comes to the design of assessment tools and their role in treatment. The findings also suggest that clinicians should consider that new patients with no prior knowledge of or experience with cancer treatment and its associated side effects may place less importance on the quality of life consequences of CIPN than they might once they have more understanding and experience of the treatments. Hence, when deciding with patients about care and treatment options, this information should be brought to their attention so that they can make an informed decision.

5.5.3 Strengths and weaknesses of the study

Although there has been previous research comparing patient and general population preferences, as far as the author is aware this is the first study specifically designed to test whether level of information provided has an impact on differences between these two populations.

It should be kept in mind that the gender and age distributions of the patient sample were very different to the general population sample. In particular, the majority of the patient sample was female, with a median age of 64. Respondents in the patient sample were also all volunteers and recruited on a convenience basis. In contrast, the general population sample was recruited from a research panel and respondents were representative of the Australian population in terms of age and gender. Apart from demographic differences, the number of respondents in the general population sample was more than twice that of the patient sample. That said, despite the differences in demographic and sample sizes in the two samples, the preferences elicited were actually quite similar with only a few exceptions, as summarised above. It was interesting to note, that despite how similar preferences exhibited were, scale differences still existed between the patient and general population sample that prevented direct comparison of model results.

5.5.4 Implications of findings and directions for future research

The current findings add to the literature on comparisons of preferences between patient and general population samples. The current project finds that although preferences were, ultimately, quite similar, some important differences were noted between the general population and patient preferences. Differences between patient and general population preferences have been noted in other studies (Najafzadeh et al., 2019; Ogorevc et al., 2019). In particular, evidence from the literature suggests that patients are more conscious of quality-of-life implications than the general population. Najafzadeh et al. (2019) investigated preferences for anticoagulant treatment outcomes. They find that patients considered the negative quality-of-life consequences of a non-fatal cardiovascular event as worse than death. This was in contrast to the general population, which considered death worse than a non-fatal cardiovascular event.

In contrast, other studies have found that patients are conscious of both quality- and quantity-of-life implications of health states. In particular, Goodwin et al. (2021) used cognitive interviews to investigate how patients with multiple sclerosis (MS) and the general population value different health states. No patients at any stage mentioned that quality of life was more important than quantity of life – rather, it was some of the general population respondents who suggested this. While informative, it should be

kept in mind that the study by Goodwin et al. (2021) involved only 12 patients and 14 members of the public.

The current study adds to the literature investigating preferences between different population types. For instance, there are many studies in the literature that have compared preferences from health care providers with other population types, such as patient, caregiver, general population etc. (Harrison et al., 2017).

A recent systematic review by Harrison et al. (2017) finds that in the majority of included studies there were differences in preferences between health care providers and other population groups. That said, it was most common for studies to find mixed results – that is, some agreement and some disagreement in preferences. Interestingly, Harrison et al. (2017) note that the level of agreement or disagreement of preferences did vary depending on the type of attributes. In particular, health care providers were much concerned about attributes related to structure (e.g. infrastructure, such as hospital beds per room, organisational or human resources) and outcomes (e.g. mortality, effects of treatment or care on patients). In contrast, patients placed more importance on process-related attributes such as safety, delivery and timing of treatment or care.

The considerable number of studies examining preference differences between health providers and other population types, particularly patients, have been motivated in part by a move towards a patient-centred approach to health care – that is, allowing for patient involvement in the decision-making process and responding to a patient's desire for information (Harrison et al., 2017; Little et al., 2001; Stewart, 2001). In particular, there is evidence that ensuring that patients are informed improves patient satisfaction and promotes more efficient use of health resources (Harrison et al., 2017; Stacey et al., 2017).

Overall, when comparing Study 1 and Study 2, preferences were quite similar. The only major difference seen was in the concern of the patient sample regarding quality-of-life implications of CIPN, which the general population sample did not share. This is despite more information being given to Arm 2 of the general population sample, including a video with patient testimonials of their experience of CIPN symptoms. The reduced scale difference between Arm 2 and the patient sample, when compared with Arm 1 and

the patient sample, suggests that further scale reductions are possible and would be useful. Potentially, scale differences could be reduced further or even disappear altogether were respondents in Arm 2 given information emphasising the quality-of-life implications of CIPN symptoms. Future studies could investigate this further. It would also be valuable to understand whether there is any potential to reduce scale differences between a general population and a patient population when it comes to health conditions other than cancer. In other words, future studies could investigate the generalisability of the findings from this current project to health conditions overall.

Another point for consideration is that provision of extra information may reduce scale differences between populations but may not reduce preference differences. It is possible that after accounting for scale differences, different population types may still have differences in their preferences. This is suggested by the comparisons of the patients and general population estimates through ratios of parameters and MRS. In particular, these ratios suggest the presence of both scale and preference differences. In other words, increasing the level of information provided may reduce scale differences without necessarily reducing preference differences between the populations. If this is indeed the case then, apart from increasing respondent understanding of choice sets, the purpose of extra information provision may be increasing the ease of statistical analysis. That is, if providing the right type or amount of extra information can reduce the scale difference entirely between a patient and general population sample, parameter estimates can be directly compared between samples. This will be particularly useful for DCE studies where WTP, or the use of ratios of some type, is not the most easy to interpret for comparison purposes. This is something that can also be explored in future studies.

Chapter 6. Anchoring choice sets to immediate death or full health during the valuation of the EQ-5D-5L

6.1 Overview

In the previous two chapters, the impact of two types of presentation differences in a DCE were investigated. This included varying the amount of information provided and using alternative text formats. The impact of these presentation differences was explored in the context of using a DCE to design a health assessment tool. Understanding whether respondents were sensitive to the amount of information provided in a DCE, as well as the presentation format of information, is important, as health economists often need to include, and decide how to present, explanations of unfamiliar concepts and terms used in the DCE.

Another type of presentation difference that is of relevance to health economists is understanding the impact of anchoring. This is of particular relevance to the valuation of quality-of-life (QoL) instruments, where anchoring of choice sets is often used. In this chapter, a QoL valuation DCE is described with three options per choice set. The impact of anchoring in a choice set by using a third option of immediate death or full health for a shorter duration is investigated. This investigation was conducted in the context of the EQ-5D-5L, a QoL instrument. The next section details the role of the author. This is followed by relevant background information including a brief introduction to the concept of a quality-adjusted life year (QALY) and QoL, including the EQ-5D-5L and current issues in the valuation of the EQ-5D-5L. The formal aims of this chapter are then introduced.

6.2 Author involvement in the project

The data used in this chapter are from a larger project funded by the EuroQol Research Foundation (EuroQol Research Foundation, 2021). This project involved collaboration among a number of EuroQol members, with different members responsible for different parts of the project. The role of the author of this thesis was to undertake the analysis of two approaches to presenting the health state and duration information in the DCE. The author was not involved in the initial design of the study or of the data collection but did

undertake much of the data analysis of the DCE_{TTO} choice set component of the project, part of which is used for this thesis chapter. The data from this project were chosen as it was considered an appropriate study for investigating the potential impact of anchoring of choice sets on respondent preferences.

6.3 Background

6.3.1 Incorporating health effects into economic evaluation

There are multiple factors to consider when health care decision-makers and governments make decisions about whether to adopt a new health technology, program or policy. This is, in part, due to the limited nature of health care resources and the increasing pressures on national health care budgets, which mean that not all health technologies, programs or policies can be implemented (Drummond et al., 2015). The adoption of one health technology or program means that these resources will not be available for an alternative technology or program (Drummond et al., 2015). It is important to understand how these resource allocation decisions affect the health of a population (Drummond et al., 2015). As a result, there is an increasing need for a systematic approach that considers the clinical and economic evidence for the relative value of alternative health programs to assist in the allocation of health resources. This includes assessing the cost of the program or technology and weighing it against the consequences (e.g. health effects or potential gains). Furthermore, the cost and gains from a proposed health program or technology must be assessed against a comparator or alternative program or health technology, which might be the current health technology in place. This allows the evaluation of a new program or technology in the context of whether it is a better or worse use of health resources (Drummond et al., 2015). This type of assessment has generally been referred to as an economic evaluation (Drummond et al., 2015; Torrance, 1986). A widely established generic measure of health effect, commonly used in economic evaluation, is the quality-adjusted life year or QALY (Drummond et al., 2015; Klarman & Rosenthal, 1968; Torrance, 1987).

6.3.2 Defining quality-adjusted life years (QALYs)

The QALY is a widely used measure of health outcome that simultaneously accounts for changes in both quality of life and quantity of life (Drummond et al., 2015; Klarman & Rosenthal, 1968). Quantity of life refers to the survival or number of years a person is

alive; quality of life, in the context of this thesis, refers to health-related quality of life (HRQoL) during the period for which a person is alive. This includes both physical (e.g. physical disabilities, pain) and emotional wellbeing (e.g. emotional distress, depression) (Torrance, 1987). One QALY represents a year of life in full health. QALYs associated with different health conditions are calculated by multiplying the duration of the condition (or health state) by a QALY weight (Drummond et al., 2015).

QALY weights are the relative utility or weight of any two or more health states (Norman, Cronin, et al., 2013; Torrance, 1987). In economic evaluations, QALY weights can be combined with information about the probability of health outcomes and survival associated with these outcomes to estimate QALYs (Drummond et al., 2015). QALY weights can take a range of values but must include two values that help to 'anchor' the scale. The conventional QALY scale has perfect health anchored at 1 and death anchored at 0, with health states with a QALY weight of less than 0 considered worse than death (Drummond et al., 2015). QALY weights for specific health states are derived through hypothetical (stated preference) studies, which aim to measure the preferences of the population of interest regarding the health states (Drummond et al., 2015). A stated preference approach is necessary because the value of health states cannot be observed in revealed preference data.

There are a range of different stated preference techniques researchers may use to derive the QALY weights of particular health states from populations of interest – for example, time trade-off, standard gamble and, more recently, DCEs (Drummond et al., 2015; Torrance, 1987). To calculate the QALY weights using DCEs, a random sample of the population of interest is asked to express their preferences for options presented in a series of DCE choice sets. The options are health states made up of levels of different health attributes. Usually, there is also a duration attribute featuring the length of time respondents would experience the particular health state (Bansback et al., 2012; Viney et al., 2014). A duration attribute is included as it provides information about the trade-off respondents make between quantity and quality of life, and provides a numéraire for estimating QALY weights (Viney et al., 2014). Analysis of the data from the DCE can then be used to calculate the QALY weights.

6.3.3 The use of multi-attribute utility instruments (MAUIs)

Multi-attribute utility instruments (MAUIs) are a specific type of QoL instrument that describe health on a number of dimensions or attributes and also include a scoring algorithm to derive the QALY weights associated with the instrument's health states (Viney et al., 2014; Viney et al., 2011). QALY weights derived from these MAUIs can be used to express and compare health outcomes for cost effectiveness and cost utility analyses (Drummond et al., 2015; Torrance, 1987). In other words, they can be used to decide whether a particular health technology or intervention would be good value for a population of interest.

Increasingly, when it comes to reimbursement schemes (e.g. the Pharmaceutical Benefits Scheme in Australia), governments prefer QALYs to be estimated using MAUIs embedded in clinical trials (Eldessouki & Smith, 2012; Kennedy-Martin et al., 2020). Kennedy-Martin et al. (2020) report that there are 34 countries where official pharmacoeconomic guidelines are present. Of these, 29 recommend the use of a MAUI, with 15 requiring the use of a specific MAUI.

A number of different MAUIs have been developed and used in the literature; these include the Six Dimensional Health State Short Form (SF-6D), which is the scoring algorithm used for the Health Survey Short Form 36 (SF-36) (Brazier et al., 2002; Viney et al., 2011; Ware Jr, 2000); the Health Utilities Index (HUI) (Torrance et al., 1995; Viney et al., 2011); and the Assessment for Quality of Life (AQoL) (Hawthorne et al., 2001; Viney et al., 2011). However, by far the most commonly used MAUI is the EQ-5D (Kennedy-Martin et al., 2020; Norman, Cronin, et al., 2013; Viney et al., 2011). Kennedy-Martin et al. (2020) found that the EQ-5D was the most preferred and most used MAUI across 34 different guidelines from around the globe. It has also been identified as the most used MAUI in a review of cost utility analyses (Wisløff et al., 2014). Wisløff et al. (2014) report that 77% (87/113) of studies that explicitly reference a single MAUI use the EQ-5D-3L.

6.3.4 The EQ-5D-5L

The EQ-5D instrument was developed by the EuroQol Group, beginning in 1987. The goal was to develop a generic instrument to describe QoL (Devlin & Brooks, 2017; EuroQol Research Foundation, 2021). There are now three main versions of the EQ-5D:

the EQ-5D-3L (3-level version), EQ-5D-5L (5-level version) and the EQ-Y (EuroQol Research Foundation, 2019). The EQ-Y is suitable for children aged 8–15 years. The EQ-5D-3L was created in 1990 and is the most widely used version, being available in more than 180 languages (EuroQol Research Foundation, 2019). The EQ-5D-5L was introduced in 2009 and extends the EQ-5D-3L by increasing the number of levels for each domain from three to five to provide increased sensitivity to QoL changes (EuroQol Research Foundation, 2019; Norman, Cronin, et al., 2013).

Both the EQ-5D-3L and the EQ-5D-5L measure health using five dimensions or attributes: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The number of levels available for each dimension is either three (EQ-5D-3L) or five (EQ-5D-5L). The EQ-5D-5L domains and levels are provided in Table 6.1, which was based on the UK self-complete paper version of the EQ-5D-5L User Guide (EuroQol Research Foundation, 2019). Each attribute has five levels. For instance, for mobility the levels are no problems in walking about; slight problems walking about; moderate problems walking about; severe problems walking about; and unable to walk about. Each combination of levels from each of the five attributes represents a unique health state, meaning there are $5^5 = 3125$ possible health states in total. To illustrate, a respondent may indicate they have no problems with mobility (level 1), moderate problems with self-care (level 3), slight problems with usual activities (level 2), severe problems with pain/discomfort (level 4) and extreme problems with anxiety/depression (level 5). This respondent's health state can be represented by the 5-digit code 13245, as summarised in Table 6.1. If a respondent rates all five dimensions as no problems (i.e. best possible level in each dimension), then they have the health state 11111, usually called full health.

Table 6.1 EQ-5D-5L levels with health state 13245 underlined and bolded for illustration

Attributes	Levels	Level description
Mobility	***1***	***No problems in walking about***
	2	Slight problems in walking about
	3	Moderate problems in walking about
	4	Severe problems in walking about
	5	Unable to walk about
Self-care	1	No problems washing or dressing self
	2	Slight problems washing or dressing self
	3	***Moderate problems washing or dressing self***
	4	Severe problems washing or dressing self
	5	Unable to wash or dress self
Usual activities (e.g. work, study, housework, family or leisure activities)	1	No problems doing usual activities
	2	***Slight problems doing usual activities***
	3	Moderate problems doing usual activities
	4	Severe problems doing usual activities
	5	Unable to do usual activities
Pain/Discomfort	1	No pain or discomfort
	2	Slight pain or discomfort
	3	Moderate pain or discomfort
	4	***Severe pain or discomfort***
	5	Extreme pain or discomfort
Anxiety/Depression	1	Not anxious or depressed
	2	Slightly anxious or depressed
	3	Moderately anxious or depressed
	4	Severely anxious or depressed
	5	***Extremely anxious or depressed***

6.3.5 Deriving QALY weights for the EQ-5D-5L

Although health states can be summarised using the 5-digit code, the code itself cannot be used directly for economic evaluation. Rather, for each health state that is represented by a 5-digit code (from 11111 to 55555), a specific QALY weight can be assigned, and it is these weights that are used as inputs in health-related economic evaluations of new interventions using cost utility analysis (EuroQol Research Foundation, 2019; Kennedy-Martin et al., 2020; Yang et al., 2019). The QALY weight sets are obtained by asking representative samples of the population (e.g. for a country or

region) to value different EQ-5D-5L health states using a preference-based task that includes a numéraire (most often survival duration). The scoring algorithm for the EQ-5D-5L has typically been valued using a time trade-off task, and there is now a standardised valuation study protocol called the EQ-VT for obtaining value sets for the EQ-5D-5L. This standard protocol generally uses the composite time trade-off (cTTO) valuation technique to obtain value sets (Andrade et al., 2020; Janssen et al., 2017; Rencz et al., 2020; Welie et al., 2020). The cTTO is completed face-to-face with an interviewer. The cTTO follows the typical TTO format when respondents are considering health states better than death (i.e. respondents are faced with *x* years in full health versus 10 years in a disease state and *x* is varied until respondents are indifferent between the two options). For health states considered worse than death, respondents are faced with a modified TTO format of *x* years in full health versus a fixed life of 10 years in full health followed by 10 years in a specified health state (Stolk et al., 2019).

6.3.6 Using DCEs to derive QALY weights for the EQ-5D-5L

DCEs have become increasingly popular as a method for obtaining QALY weights (or value set) for health states described by MAUIs such as the EQ-5D-5L. A recent structured review identifies 63 papers published between 1997 and 2018 that used DCE methods for health state valuation (Mulhern et al., 2019). As the use of DCEs to elicit values for the EQ-5D has increased, questions have been raised about how the format, design and implementation of the DCE impact on the utility values produced (Jonker et al., 2018b; Lim et al., 2018; Mulhern, Norman, Shah, et al., 2018; Mulhern, Norman, Street, et al., 2018; Norman et al., 2016).

It is important to note that utility values derived from any DCE are arbitrarily defined. That is, DCEs produce values that can be used to determine the relative position of each level within a particular dimension, and also the relative value of levels across different dimensions. The values produced can also be combined to compare the relative utility of different health states. However, the values produced cannot be directly compared across models as they are on a latent scale; hence, they are not immediately useful for valuing health for the purposes of cost-utility analysis. This is because the QALY model imposes additional assumptions to ensure that it is based on a cardinal utility scale with anchors at full health (1) and death (0). Therefore, one focus of valuation research has

been on methods to anchor values on the utility scale from 1 to 0 required in order derive QALY weights (Norman, Cronin, et al., 2013).

One common method of anchoring values has been to include duration as an attribute in each option of the choice set, called the DCE_{TTO} (Bansback et al., 2012). For valuation of the EQ-5D using the DCE_{TTO}, each option consists of a health state described using the five dimensions of the EQ-5D and an additional attribute for duration, giving six attributes in total. Each respondent's utility is assumed to be defined by the product of the utility of the specified health state and the duration in each option. As such, when duration is zero the health state does not matter, as it assumed that respondent preferences would depend on both duration and the health state. This is known as the zero-condition assumption (Miyamoto et al., 1998; Bansback et al., 2012). The QALY model also imposes the assumption that constant proportional time trade-off holds. That is, the trade-off between duration and quality of life gain is a fixed proportion and is independent of the duration or quantity of life years presented to respondents. For instance, a respondent could be faced with two health states, where one health state was 15 years in duration and the other was a better and improved health state but for a duration of 12 years (20% reduction). A respondent may choose the improved health state, sacrificing the extra three years' duration. For the same improvement in health state, it is assumed that if respondents were faced with 20 years versus 16 years, they would make the same choice. That is, they would choose the improved health state with an equivalent proportional loss in duration – a 20% reduction, which in this case is four years (Bansback et al., 2012; Pliskin et al., 1980).

One concern that has been raised about this approach (in which two or more health state duration combinations are presented to respondents) is the determination of where being dead lies on the scale, as respondents have not been asked explicitly to consider whether a health state is better or worse than being dead, as they are in other tasks such as the EQ-VT protocol. In such a situation, the position of being dead on the utility scale is inferred through the modelling process (Bansback et al., 2012; Norman, Cronin, et al., 2013).

An alternative approach is to explicitly include consideration of death in the valuation of health states of the EQ-5D-5L. One approach has been to include immediate death as an option in each choice set (Norman, Cronin, et al., 2013; Norman et al., 2016). This is

based on an approach to DCE_{TTO} used to develop values for the EQ-5D-3L (Viney et al., 2014) , EQ-5D-5L (Norman, Cronin, et al., 2013) and SF-6D (Norman, Viney, et al., 2013). This approach presents two health state duration combinations (A and B) and a third option (C) described as immediate death. Options A and B are represented by an EQ-5D-5L health state experienced for a specified duration (which may be the same for the two health states or may differ). Option C is specified as death (no duration and no health state). In the standard format, respondents are asked to choose both the best and the worst option of the three in order to obtain a complete ranking of health states from those considered worse than death to full health (Norman, Cronin, et al., 2013). This will be called the *immediate death* approach in this chapter.

Another approach is to include full health for a specified duration (rather than immediate death) as the third option within each choice set (Jonker et al., 2017; Lim et al., 2018). This approach to DCE_{TTO} was developed in The Netherlands (Jonker et al., 2017). This approach also presents two health states (A and B) but now experienced for the same duration. State C is full health experienced for a shorter duration than that presented for A and B. In the approach that has been most commonly implemented, respondents first choose between health states A and B, and then between health state B and health state C. The specific task format was aimed to simplify the trade-offs respondents make by splitting the choice set into a two-stage task. Another reason was to improve the internal consistency of health states by reducing the possibility of respondents facing extreme differences in health states – for example, no problems versus extreme problems (Jonker et al., 2017). This will be called the *full health* approach in this chapter.

The *immediate death* and *full health* approaches that incorporate duration and different anchors into health state valuation provide an important case study of how information presented in a DCE may be important in influencing results.

6.4 Aims

The aim of this chapter is to investigate whether anchoring choice sets to either immediate death or full health had any impact on the valuation of the EQ-5D-5L. This is investigated in the context of respondents from two countries, Peru and Denmark.

6.4.1 Details of the project

In this project, the DCE_{TTO} component involved a 2×2 between-subjects design. Along with the two anchoring approaches to the DCE_{TTO}, two construction methods of the choice sets were also used – namely, an efficient generator-developed design and an algorithmic construction of an efficient design. This resulted in a four-arm study. For the purposes of this chapter, it was decided that results from only one construction approach would be reported, as the two construction methods produced very similar results. Reporting on one construction method was sufficient to tell the overall story of the data and to address the aim of this chapter. As a consequence, the focus of this chapter will be on the on the relevant Peruvian and Danish data from the DCE_{TTO} choice sets constructed using an efficient generator-developed design.

6.5 Methods

6.5.1 The data

This project was conducted with the purpose of eliciting values for the EQ-5D-5L health states from general population samples in both Peru and Denmark. Data collection took place in the respective countries. This was conducted in accordance with the EQ-5D-5L valuation protocol version 2.1 (Oppe et al., 2016; Stolk et al., 2019). This protocol involved respondents completing cTTO tasks followed by the DCE_{TTO} choice sets (Oppe et al., 2016; Stolk et al., 2019). The data analysed in this chapter are from the responses to the DCE_{TTO} choice sets.

Data collection was undertaken from April 2018 to February 2019 in Peru. For Denmark, interviews were completed between October 2018 and December 2019.

6.5.2 Data collection: Peru

The procedure to collect data in Peru is described briefly here. More detail about respondent demographics and the data collection process can be found in Augustovski et al. (2020). Participants completed the DCE_{TTO} as part of the national EQ-VT valuation study in Peru. In total, for the 4 arms, a population-based random sample of 1000 adults aged 18–75 years was used, drawn from three major cities in Peru, one located on the coast (Lima, the capital), one in the highlands (Arequipa) and one in the jungle (Iquitos). A multistage sampling design was used, consisting of four stages stratified by

socioeconomic level (census enumeration areas, blocks in each census enumeration area, households from each block using systematic sampling, selecting one household member 18–75 years old). The National Institute of Health of Peru hired 11 medical students who were specifically trained for this protocol. Each interviewer underwent a rigorous face-to-face 5-day training session and the decision whether to use each interviewer was based on their performance after performing 5–10 pilot interviews. The interviews were performed at participants' homes and administered through the EuroQol EQ-VT platform on laptops.

6.5.3 Data Collection: Denmark

Data collection in Denmark is described briefly in this section. More detail about respondent demographics and data collection process can be found in Jensen et al. (2021).

Similarly to the data collection in Peru, the DCE$_{TTO}$ was completed as part of a national study in Denmark to develop the first Danish value set for the EQ-5D-5L. There was a target sample size of 1200 respondents, across the four arms. The contact information of a random sample representative of the Danish adult population (aged 18 years and over) was provided by Statistics Denmark. Individuals were sent an email invitation to participate. Individuals who agreed to participate could choose to be interviewed at home or at a nearby public institution. The study was administered as computer-assisted personal interviews. Interviewers had a Master's degree in public health or medical market access. They completed 2.5 days of training prior to data collection.

6.5.4 Arms in study

The DCE$_{TTO}$ tasks were presented following the completion of the main valuation protocol, which was based on the cTTO discussed above.

Respondents were randomly assigned to one of four arms. However, for the purposes of this chapter, only two arms were used for analysis (see section 6.4.1 for further explanation), summarised in Table 6.2.

Table 6.2 Arms in study

Anchoring approach to DCE_{TTO}	Arm	Description
Immediate death	death_arm	Respondents saw immediate death as the third option.
Full health	full health_arm	Respondents saw full health for a shorter duration compared to health state B.

6.5.5 List of attributes and levels

Table 6.3 summaries the attributes and levels used in the DCE. Abbreviations used for attribute levels are also listed.

Table 6.3 List of attributes and levels

Attributes	Levels	Level description
Mobility	MO1	No problems in walking about
	MO2	Slight problems in walking about
	MO3	Moderate problems in walking about
	MO4	Severe problems in walking about
	MO5	Unable to walk about
Self-care	SC1	No problems washing or dressing self
	SC2	Slight problems washing or dressing self
	SC3	Moderate problems washing or dressing self
	SC4	Severe problems washing or dressing self
	SC5	Unable to wash or dress self
Usual activities (e.g. work, study, housework, family or leisure activities)	UA1	No problems doing usual activities
	UA2	Slight problems doing usual activities
	UA3	Moderate problems doing usual activities
	UA4	Severe problems doing usual activities
	UA5	Unable to do usual activities
Pain/Discomfort	PD1	No pain or discomfort
	PD2	Slight pain or discomfort
	PD3	Moderate pain or discomfort
	PD4	Severe pain or discomfort
	PD5	Extreme pain or discomfort
Anxiety/Depression	AD1	Not anxious or depressed
	AD2	Slightly anxious or depressed
	AD3	Moderately anxious or depressed
	AD4	Severely anxious or depressed
	AD5	Extremely anxious or depressed

6.5.6 The task

As mentioned previously, choice sets were anchored in one of two ways. Each choice set included three options and was answered as a two-stage task, with respondents asked to choose between the different health state options, A and B, followed by a choice between health states B and C. Depending on the arm to which respondents were assigned, health state C was shown either as *immediate death* or as *full health* for a shorter duration than health state B. Participants completed a practice choice set followed by 12 choice sets used for data analysis. Each choice set was presented in two stages; thus, 24 choices were made in total. Choice sets were constructed with the intention that each respondent would be randomly assigned 12 choice sets from the total number of choice sets. However, the survey system was not able to do this. Instead, the survey system grouped choice sets into blocks of 12, and each respondent was assigned one of these blocks to complete.

An example of the two stage choice set from the *death_arm* for the Peruvian respondents is represented in Figures 6.1 and 6.2. These are not actual replications of the choice set but a translated English language version produced for illustrative purposes. An example of an actual task from the DCE_{TTO} component of the study using full health as health state C is available in Augustovski et al. (2020), a paper that describes the data collection of the Peruvian data.

The implementation of the choice set for the arms with *immediate death* differed from the way these tasks have been implemented in previous work (Norman, Cronin, et al., 2013; Viney et al., 2014). In previous studies, respondents were asked to indicate the most preferred of the three options in the choice set and then indicate the less preferred of the remaining two. This ensures that there is a complete ranking of the three items. The method used in this version of the task did not guarantee that a complete ordering would result, and so could lose as much as one-third of the information available in the intended implementation of this approach. However, to impose a similar preference elicitation task procedure across arms, this adapted format was used for both the *immediate death* and *full health* approach to the DCE_{TTO}.

Thus in the death arm, respondents first choose between health states A and B, which could have differing durations. They then choose between health state B and immediate

death (health state C). In the *fullhealth arm* respondents chose between health states A and B, which have the same duration. Respondents were then asked to choose between health state B and full health for a shorter duration.

Table 6.4 Arm death_arm: mock-up of health state A vs B (different durations)

Which option would you prefer?	A	B
Duration	10 years in this health state then you die	15 years in this health state then you die
1. Problems with mobility	Slight problems in walking about	Slight problems walking about
2. Problems with self-care	Moderate problems washing or dressing yourself	Slight problems washing or dressing yourself
3. Problems with usual activities (e.g. work, study, housework, family or leisure activities)	Severe problems doing usual activities	No problems doing usual activities
4. Problems with pain/discomfort	No pain or discomfort	Severe pain or discomfort
5. Problems with anxiety/depression	Not anxious or depressed	Moderately anxious or depressed

Table 6.5 Arm death_arm: mock-up of health state B vs C (anchored to immediate death)

Which option would you prefer?	B	C
Duration	15 years in this health state then you die	Immediate death
1. Problems with mobility	Slight problems walking about	
2. Problems with self-care	Slight problems washing or dressing yourself	
3. Problems with usual activities (e.g. work, study, housework, family or leisure activities)	No problems doing usual activities	
4. Problems with pain/discomfort	Severe pain or discomfort	
5. Problems with anxiety/depression	Moderately anxious or depressed	

6.5.7 Choice set construction method

The efficient generator-developed designs used in this study were constructed by Professor D. J. Street using the approach in Street and Burgess (2007) , with the additional constraint that two of the health state attributes were to be at the same level in options A and B for the full health design. As noted earlier, the efficient generator-developed construction method does not require priors to inform the design.

Originally the total number of choice sets to be constructed was 125. For this construction method, the order of presentation of health states A and B was fixed in the order that they were constructed in the design. In the *death_arm* an initial set of 125 5-level 6-tuples of resolution 3 was constructed, to which the generator (1, 2, 3, 4, 1, 3) was added to give the pairs of health states. For *fullhealth_arm*, initial sets of 25 5-level 6-tuples of resolution 3 were used as the first option and five generators of weight 3 were chosen (Street & Burgess, 2007), with each being added to a specific initial set. The four choice sets in each set of 25 in which one health state dominated the other were removed, so the final design had 105 choice sets with no dominant pairs. For both arms duration was obtained from an initial coding of levels. In the *fullhealth_arm* the full health duration was chosen randomly from an allowable set of shorter durations. The option to be compared to full health in each set was fixed.

6.6 Data Analysis

6.6.1 Modified utility specification

The specification of the utility model used in this chapter is a modification of that described in Chapter 2. The utility model in this chapter follows the utility function originally proposed by Bansback et al. (2012) and since used in other valuation studies of the EQ-5D-5L (Norman, Cronin, et al., 2013; Viney et al., 2014). It is assumed that the utility function is linear with respect to time (duration), which imposes the constant proportional trade-off assumption of the QALY model. Similarly, as discussed in the background section of this chapter, it is assumed that the zero condition holds. That is, the utility function is defined as the interaction between the attribute levels that enter the model and the the duration attribute, such that when duration is 0, utility is also 0 (Bansback et al., 2012; Miyamoto et al., 1998). It is also assumed that constant proportional time trade-off holds (Bansback et al., 2012; Pliskin et al., 1980).

6.6.2 MNL model with duration

Initially, a multinomial logit (MNL) model was estimated for each of the four arms. In the models, there is either a full health or an immediate death option, and in each case these are treated as a health state with a fixed duration (Norman et al., 2016). The advantage of this approach is that the utility specification will remain the same and the conventional QALY structure holds.

The MNL model specification was based on Norman, Cronin et al. (2013), where utility for health state *n* for individual *i* in choice set *j* is specified as:

$$U_{nij} = \alpha Duration_{nij} + Duration_{nij} X_{nij}^{T} \beta + \varepsilon_{nij}, \quad n = 1, \dots, N;\ i = 1, \dots, I;\ j = 1, \dots, J, \tag{1}$$

where α = the utility associated with duration,

$Duration$ = period of time associated with the health state,

X^T = vector of length 20 and is the set of dummies corresponding to the EQ-5D 5L health state dimension levels (with level 1 omitted for each dimension as it is used as the base level),

β = vector of length 20. Entries to the vector are the associated utilities levels 2–5 of each dimension of the EQ-5D-5L (level 1 is the base level, which is omitted). Specifically, parameters 1–4 of the vector are associated with the dimension of mobility, parameters 5–8 are associated with the dimension of self-care, parameters 9–12 are associated with the dimension of usual activities, parameters 13–16 are associated with the dimension of pain/discomfort and parameters 17–20 are associated with the dimension of depression/anxiety.

The MNL models were estimated in STATA 15.1 (StataCorp, 2017).

6.6.3 Reporting of MNL model parameter estimates: calculating QALY weights

Raw parameter estimates from the MNL model are on the latent scale and represent the interaction between duration and the attribute level. The raw parameters from the MNL model estimation can be converted to be on the QALY scale. This is done by taking the raw MNL parameter estimates and dividing by the utility associated with duration (i.e.

α) in order to obtain the parameter estimates on the QALY scale (Gu et al., 2014). Thus all the parameters are divided by the estimate of α. Parameter estimates are reported on the QALY scale to allow comparison of values across arms.

Parameter estimates on the QALY scale can then be used to calculate the QALY weight or utility of each of the 3125 possible health states. This is because the zero condition allows an arbitrary linear transformation of the data to get to the QALY scale. The utility is equal to one plus the sum of the parameter estimates associated with each attribute level in the health state (Gu et al., 2014; Norman, Cronin, et al., 2013). The utility of health state *n* for individual *i* in choice set *j* is specified as

$$U_{nij} = 1 + \sum X_{ni}^{T} \frac{\beta}{\alpha}, \quad n = 1, \dots, N;\ i = 1, \dots, I;\ j = 1, \dots, J. \tag{2}$$

Utility decrements are therefore represented by $\frac{\beta}{\alpha}$ (Gu et al., 2014).

6.6.4 MXL model with duration

The potential presence of preference heterogeneity was examined through the mixed logit model. In the mixed logit model, parameter estimates associated with each individual are assumed to be drawn from an underlying distribution (Train, 2009). In this model, the normal distribution has been used for the underlying distribution and has a diagonal covariance matrix. For the deterministic component of utility, each decision-maker is allowed to have their own vector of parameter estimates, which is represented by β_i.

The random utility of option *n* for individual *i* in choice set *j* is given by

$$U_{nij} = \alpha Duration_{nij} + Duration_{nij}\, X_{nij}^{T} \beta_i + \varepsilon_{nij}, \quad n = 1, \dots, N;\ i = 1, \dots, I;\ j = 1, \dots, J, \tag{3}$$

where

$$\beta_i \sim MVN\ (\beta, \Sigma).$$

The unobservable component of utility for individual *i* in choice set *n* and associated with option *j* is represented by ε_{nij}, and is assumed to be iid extreme value (Sarrias & Daziano, 2017; Train, 2009).

MXL models were estimated in R Studio (R Development Core Team, 2020) using the gmnl package (Sarrias & Daziano, 2017). In order to estimate the MXL model, the number of Halton draws, the primes to use and the number of elements to drop from each sequence were specified. Two thousand Halton draws were used with 500 of the initial sequence of elements dropped for each of the primes. This was done to reduce the risk of strong collinearity among the Halton sequences (Andersen, 2014; Czajkowski & Budziński, 2019). MXL model results are reported as raw parameter estimates on the latent scale.

6.7 Data cleaning and screening of respondents

During the cleaning process of the raw data, 27 respondents in the Peruvian data set and 36 respondents in the Danish data set were dropped due to low-quality responses. Quality of responses was determined by interviewer compliance with the interview methodology and the face validity of responses (Augustovski et al., 2020; Jensen et al., 2021; Purba et al., 2017; Ramos-Goñi et al., 2017). For instance, interviewers were monitored and given weekly feedback on their interviews, in accordance with the EuroQoL Group's EQ-VT quality assurance processes. The data quality was also monitored by looking at key indicators such as time taken to complete each task and potential interviewer effects on the distribution of responses (Augustovski et al., 2020).

Respondents were only included for analysis if they completed all 24 assigned choice sets. As a consequence, one additional respondent from the Peruvian data and four respondents from the Danish data were removed due to incomplete choice data. Durations used in choice sets, and the combination of durations used in health states have been summarised and are available in Appendix 6A. Table 6.6 summarises the number of respondents by arm and country. Table 6.7 summarises the total number of choice sets in each arm. The choice sets used were the same across countries.

Table 6.6 Total number of respondents by arm and country

Country	Peru		Denmark	
Arm	death_arm	fullhealth_arm	death_arm	fullhealth_arm
No. of respondents	221	241	253	228

Table 6.7 Total number of choice sets by arm

Arm	death_arm	fullhealth_arm
No. of choice sets	125	105

6.7 Results

6.7.1 Poolability of arms across countries

The choice sets used for each arm were the same for both countries. The possibility of being able to combine the data for each arm across country for analysis was explored. The presence of potential scale differences between countries was investigated by estimating an s-mnl model on the combined data across countries for each arm, with a dummy indicator ($\delta_{dummy\_country}$) for country. A summary of the results can be found in Appendix 6B. The dummy indicator for country had a p-value < 0.001 for each of the arms, indicating that scale differences did exist between the two countries. This meant that the data from the two countries cannot be combined for analysis. Hence, the following analyses will be by arm and by country.

6.7.2 MNL model results

MNL parameter estimates by arm and by country are summarised in Table 6.6 on the QALY scale; raw parameter estimates on the latent scale are available in Appendix 6C.

Examining the parameter estimates of each attribute from the Peruvian data, there was a general pattern of decrement in utility as severity increased – that is, parameter estimates were generally monotonically decreasing. This was consistent across both arms with the exception of parameter estimates for SC2 and SC3 in the fullhealth_arm, although these parameter estimates were not significantly different from the base level SC1 ($p > 0.05$). That is, respondents in this arm did not consider having slight or moderate problems with washing or dressing themselves to be significantly worse than having no problems with washing or dressing themselves. That said, the number of

respondents in each arm are small. Hence, the non-significant parameters may be confounded by a potential lack of power.

Table 6.8 MNL results

QALY scale	Peru		Denmark	
	death_arm	fullhealth_arm	death_arm	fullhealth_arm
MO2	−0.018	−0.002	−0.028	0.041
MO3	−0.023	−0.122***	−0.085*	−0.138***
MO4	−0.171***	−0.183***	−0.169***	−0.135***
MO5	−0.462***	−0.577***	−0.232***	−0.236***
SC2	−0.002	−0.039	−0.014	−0.056*
SC3	−0.06	−0.011	−0.061*	−0.067*
SC4	−0.255***	−0.188***	−0.113***	−0.185***
SC5	−0.338***	−0.389***	−0.187***	−0.253***
UA2	−0.008	−0.073*	0.003	−0.016
UA3	−0.122**	−0.088*	−0.121***	−0.081***
UA4	−0.298***	−0.329***	−0.229***	−0.207***
UA5	−0.473***	−0.61***	−0.272***	−0.3***
PD2	−0.037	−0.067*	0.032	−0.079***
PD3	−0.05	−0.077*	−0.032	−0.126***
PD4	−0.218***	−0.323***	−0.331***	−0.388***
PD5	−0.523***	−0.598***	−0.588***	−0.568***
AD2	−0.004	0.015	0.021	−0.01
AD3	0.032	−0.098**	−0.03	−0.116***
AD4	−0.131***	−0.183***	−0.381***	−0.413***
AD5	−0.314***	−0.392***	−0.564***	−0.623***
PIT State	−1.109	−1.567	−0.844	−0.98

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute
PIT State: utility associated with the worst health state 55555 calculated by 1 + (sum of parameter estimates associated with level 5 of each attribute)

Parameter estimates in the Danish data also followed the pattern of increasing utility decrement as attribute level severity increased. A number of inconsistencies were noted (e.g. for parameters UA2, PD2 and AD2 in the *death_arm*); however, these parameters were not significantly different from 0 (p > 0.05).

Predicted health states utilities and position of 'dead' on the scale

Based on the parameter estimates, utility values for all possible health states were calculated. An example of the calculation of the health state utilities using the QALY scale parameter estimates has been provided in Table 6.7. Parameter estimates that were not significantly different from 0 (i.e. p > 0.05), were constrained to 0 in the calculation of health state utilities, an example of such a calculation has been provided

using the health state 23345, summarised in Table 6.8. It is common to constrain non-significant coefficients in order to ensure the model has consistent ordering (Mulhern et al., 2020).

Table 6.9 Example of calculation of utility weight using the PIT state

Levels associated with the PIT state (i.e. worst health state 55555)	Peru: death_arm
MO5	−0.462***
SC5	−0.338***
UA5	−0.473***
PD5	−0.523***
AD5	−0.314***
Utility weight = 1 + (-0.462-0.338-0.473-0.523-0.314)	−1.109

*p < 0.05; ** p < 0.01; *** p < 0.001

Table 6.10 Example of calculation of utility weight with constrained values

Levels associated with state 23345	Peru: death_arm
MO2 (p> 0.05, utility constrained to 0)	0
SC3 (p> 0.05, utility constrained to 0)	0
UA3	−0.122**
PD4	−0.218***
AD5	−0.314***
Utility weight = 1 + (0 + 0 - 0.122 - 0.218 - 0.314)	0.346

*p < 0.05; ** p < 0.01; *** p < 0.001

Utility weights of a few key health states are presented in Table 6.9. The proportion of health states considered to be worse than death (i.e. < 0) is also summarised in the table. The arms using the *immediate death* approach produced fewer values below zero or states considered worse than death. It was also noted that the arms using *immediate death* generally produced a narrower range of utility values compared to that produced when *full health* was used as the third option, regardless of country.

Table 6.11 Summary of health state utilities for Peruvian respondents

Health State	Peru		Denmark	
	death_arm	fullhealth_arm	death_arm	fullhealth_arm
11111	1	1	1	1
22222	1	0.859	1	0.860
33333	0.866	0.604	0.746	0.481
35433	0.355	−0.030	0.502	0.165
44444	−0.081	−0.212	−0.215	−0.333
51525	−0.312	−0.692	−0.073	−0.235
55555	−1.174	−1.613	−0.868	−0.962
%<0 *	20%	36%	20%	29%

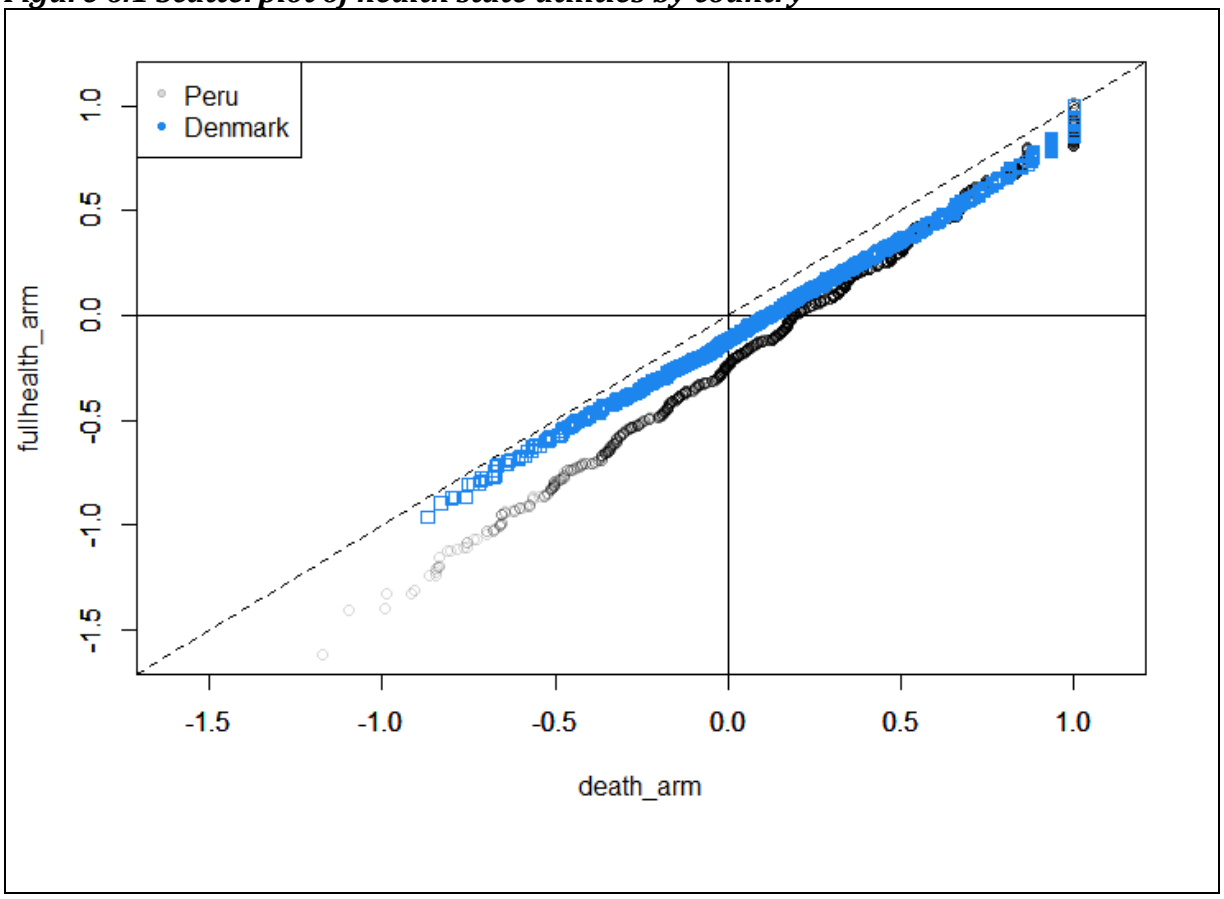
*Percentage of health states considered worse than death i.e. health state utility < 0

The utility weight values between arms for the same health states suggest that anchoring to full health versus immediate death did appear to have an impact on elicited values from respondents. For instance, it was found that across arms, respondents regarded health state 33333 as better than death (i.e. utility weight > 0) where there were moderate problems with mobility, self-care and usual activities, and moderate pain or discomfort and moderate anxiety or depression. However, health state 35433 demonstrates some of the differences between arms. Peruvian respondents in the *fullhealth_arm* considered this reasonably mild health state of having moderate problems with walking about, being unable to wash or dress self, having severe problems with usual activities, moderate pain or discomfort and moderate anxiety or depression as worse than death i.e. a point estimate that was less than 0. In contrast, Peruvian respondents in the *immediate death_arm* considered this health state as better than death, with a point estimate above 0. It was also noted that for the Danish respondents in the *fullhealth_arm* their utility for this health state was also close to zero. The utility differences in arms continue with health states considered worse than death. An example can be seen in health state 51525 (Table 6.9), which involves an inability to walk about, no problems with washing or dressing, unable to do usual activities, slight pain or discomfort, and extreme anxiety or depression. Across arms and countries, respondents consider this health state to be worse than death. It was noted that respondents in the *fullhealth_arms* considered this health state as having a greater impact on their utility compared to the *death_arms*.

Pairwise scatterplots were used to compare the arms in each country. These have been summarised in Figure 6.1. Pearson's correlation coefficient was also used to examine the correlation between arms. A near perfect correlation of 0.999 was noted between arms regardless of country.

The greatest disparity seen was in terms of the 'worst' health states, with the shape of the scatterplot less linear in the lower left quadrant for both countries. It was also noted that the greatest differences between countries was also in the states worse than death, with a noticeable divergence of the plot for health state utilities below zero.

Figure 6.1 Scatterplot of health state utilities by country*



*1. 0 = full health, 0.0 = death, <0.0 = states considered worse than death

6.7.3 MNL model results: Cross country comparison

In order to compare the Peruvian and Danish respondents further, kernel density curves of the health state utilities calculated from the MNL model parameter estimates were plotted with non-significant parameter estimates constrained to zero once again (Figure 6.2 and Figure 6.3). The kernel density curve can be thought of as a smoothed histogram with the height representing the frequency of certain utility values. The

death_arm in the Peruvian data, had the greatest number of coefficients that were not significantly different from zero ($p > 0.05$). Hence, it had the greatest number of coefficients constrained to zero, leading to the 'waviness' of the kernel density plot for this particular arm.

In both plots for the Peruvian and Danish data, it was observed that most health states lay above zero. That is, most of the health states were considered better than death. In general, the range of utility values for health states among the Danish respondents was much narrower than for the Peruvian respondents.

Figure 6.2 Peru: kernel density plots of utilities

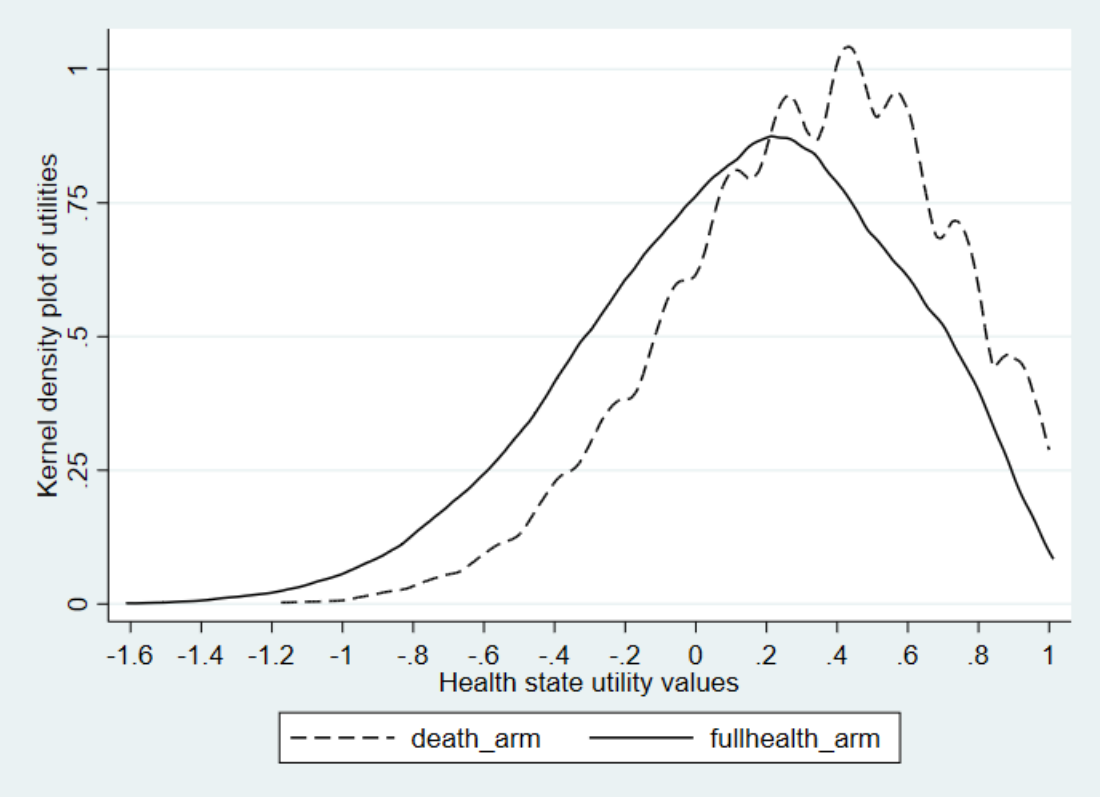
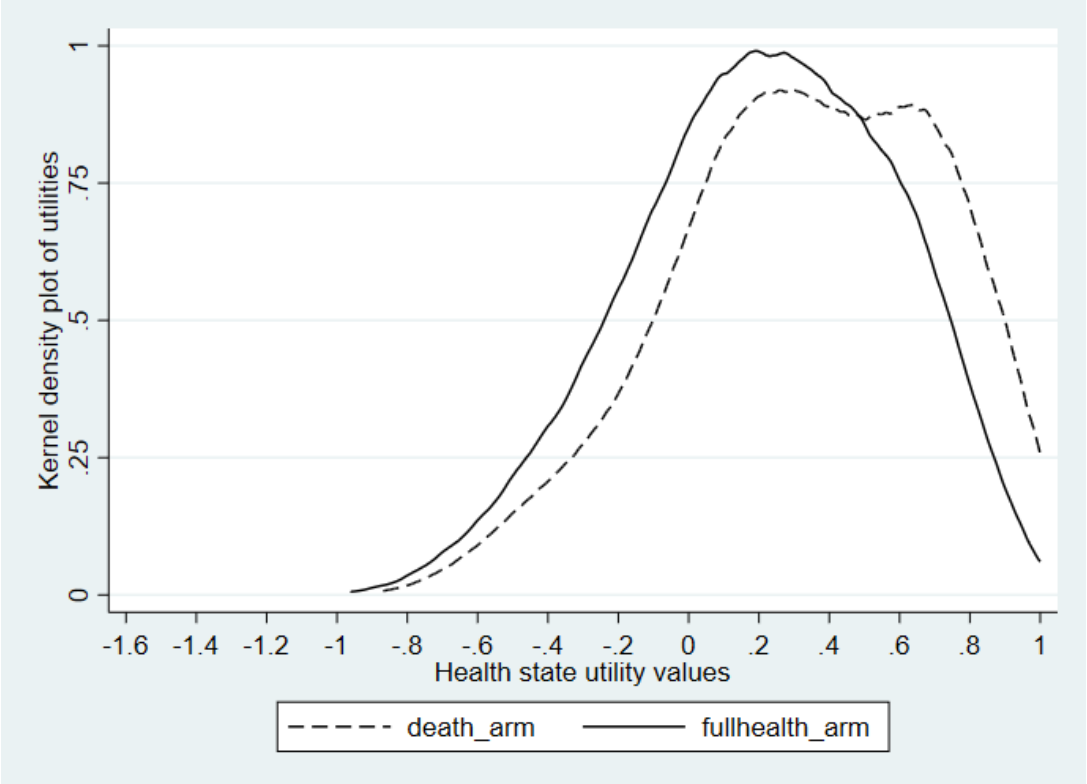


Figure 6.3 Denmark: kernel density plots of utilities



Some differences were also noted in terms of the shapes of the curves; however, the overall pattern was very similar. That is, the kernel density plot of the *fullhealth_arm* had a noticeable shift to the left compared to the *death_arm* in both the Peruvian and Danish data.

6.7.4 MNL model results: significance of adjacent utility decrements

So far, the MNL models have been estimated with parameters dummy coded such that the base level is level 1 for each attribute. For this type of dummy coding the default test is whether the parameter estimate is significantly different from zero ($p < 0.05$). In this section, the MNL models were recoded so that the default test was whether utility decrements for adjacent levels were significantly different from zero. This was done by using incremental dummy coding for each of the attributes. Instead of the usual dummy coding, where a 1 is used to indicate different level, an additional 1 adjoins to the next level. An example of the usual dummy coding and incremental dummy coding for a 5-level attribute appears in Table 6.12 and Table 6.13.

Table 6.12 Dummy coding for five level attribute

Dummy coding	d1	d2	d3	d4
Level 1	0	0	0	0
Level 2	1	0	0	0
Level 3	0	1	0	0
Level 4	0	0	1	0
Level 5	0	0	0	1

Table 6.13 Incremental dummy coding for five level attribute

Incremental dummy coding	d1	d2	d3	d4
Level 1	0	0	0	0
Level 2	1	0	0	0
Level 3	1	1	0	0
Level 4	1	1	1	0
Level 5	1	1	1	1

Table 6.14 summarises the MNL results using incremental coding. Looking first at the Peruvian death_arm parameter estimate MO2 as an example, MO2 is the difference of the latent scale parameter estimate between MO2 and MO3, which is equal to −0.005 to (−0.006) = 0.001. This decrement was not significant, indicating that the Peruvian respondents were indifferent between slight and moderate problems with mobility,

although these respondents did care about a change in mobility from moderate to severe and severe to extreme.

It was interesting to note that across arms and across countries, respondents were indifferent between slight and moderate problems with self-care (e.g. washing or dressing themselves). Peruvian respondents appeared to have a higher tolerance for pain and discomfort than Danish respondents, as they perceived an increase from slight to moderate pain/discomfort as not significant.

Table 6.14 MNL model results: significance of adjacent utility decrements

Latent scale	Peru		Denmark	
	death_arm	fullhealth_arm	death_arm	fullhealth_arm
Duration	0.249***	0.258***	0.389***	0.498***
MO2	0.001	0.031***	0.022*	0.089***
MO3	0.037***	0.016	0.033***	−0.001
MO4	0.073***	0.102***	0.025*	0.051***
MO5	−0.115***	−0.149***	−0.09***	−0.118***
SC2	0.014	−0.007	0.018	0.006
SC3	0.049***	0.046***	0.021	0.059***
SC4	0.021*	0.052***	0.029**	0.034**
SC5	−0.084***	−0.1***	−0.073***	−0.126***
UA2	0.028**	0.004	0.048***	0.032**
UA3	0.044***	0.062***	0.042***	0.063***
UA4	0.044***	0.073***	0.017	0.046***
UA5	−0.118***	−0.157***	−0.106***	−0.15***
PD2	0.003	0.003	0.025*	0.023*
PD3	0.042***	0.063***	0.116***	0.131***
PD4	0.076***	0.071***	0.1***	0.09***
PD5	−0.13***	−0.154***	−0.229***	−0.283***
AD2	−0.009	0.029**	0.02	0.053***
AD3	0.041***	0.022*	0.137***	0.148***
AD4	0.046***	0.054***	0.071***	0.105***
AD5	−0.078***	−0.101***	−0.22***	−0.31***

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute

6.7.5 MXL model results

The presence of preference heterogeneity was investigated by estimating MXL models. Table 6.14 and Table 6.15 provide a summary of the AIC and BIC of the MNL and MXL models. Relaxing the assumption of homogeneous preferences, which is made in the

MNL model, consistently provided a better fit to data regardless of the arm or the country respondents were from.

Table 6.15 Peru: AIC and BIC

Peru	AIC		BIC	
	MNL	MXL	MNL	MXL
death_arm	6235	5269	6388	5545
fullhealth_arm	7205	5828	7359	6108

Table 6.16 Denmark: AIC and BIC

Denmark	AIC		BIC	
	MNL	MXL	MNL	MXL
death_arm	6304	4729	6459	5011
fullhealth_arm	6091	5024	6244	5302

Raw parameter estimates of the means and standard deviations from the MXL models are summarised in Tables 6.17 and 6.18. Regardless of design approach or country, there was preference heterogeneity among respondents with respect to the duration attribute. This has potential implications for how QALY weights are calculated if health decision-makers seek to incorporate these differences in taste preferences for duration.

Respondents also had heterogeneous preferences for most of the attribute levels. In terms of utility decrement by severity of attribute level, mean parameter estimates were very similar to those estimated using the MNL model. As with the MNL model results, it was noted that the mean parameter estimates for level 2 of many of the dimensions for the Peruvian and Danish respondents were not significant ($p > 0.05$), although standard deviations were significant for some. For instance, heterogeneous preferences were noted for *MO2 × Duration* in the *fullhealth_arm* in both the Peruvian and Danish samples as well as in the *death_arm* in the Danish sample. This indicates that while some respondents were indifferent between slight and no problems in walking about, there were also some respondents who did care about having slight problems in walking about.

Table 6.17 Peru: MXL model latent scale results

MXL	death_arm		fullhealth_arm	
Peru	Mean	SD	Mean	SD
Duration	0.698***	0.373***	1.035***	1.146***
MO2 x Duration	−0.032	0.043	−0.023	0.075*
MO3 x Duration	−0.055*	0.097***	−0.096***	0.104**
MO4 x Duration	−0.168***	0.097***	−0.167***	0.098***
MO5 x Duration	−0.316***	0.188***	−0.505***	0.242***
SC2 x Duration	−0.003	0.115***	−0.032	0.026
SC3 x Duration	−0.045*	0.102***	−0.028	0.073**
SC4 x Duration	−0.125***	0.068*	−0.172***	0.029
SC5 x Duration	−0.224***	0.127***	−0.32***	0.202***
UA2 x Duration	−0.015	0.087**	−0.044*	0.101***
UA3 x Duration	−0.091***	0.054	−0.048*	0.078*
UA4 x Duration	−0.173***	0.128***	−0.227***	0.115***
UA5 x Duration	−0.324***	0.216***	−0.453***	0.273***
PD2 x Duration	−0.004	0.104***	−0.058**	0.027
PD3 x Duration	−0.03	0.102***	−0.089***	0.083**
PD4 x Duration	−0.1***	0.015	−0.252***	0.167***
PD5 x Duration	−0.315***	0.159***	−0.512***	0.306***
AD2 x Duration	−0.043	0.117***	−0.046*	0.111***
AD3 x Duration	−0.029	0.053	−0.103***	0.089***
AD4 x Duration	−0.095***	0.104***	−0.195***	0.122***
AD5 x Duration	−0.226***	0.183***	−0.364***	0.217***

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute

Table 6.18 Denmark: MXL model raw parameter estimates

MXL on latent scale	death_arm		fullhealth_arm	
Denmark	Mean	SD	Mean	SD
Duration	1.839***	0.609***	2.157***	1.058***
MO2 x Duration	−0.069*	0.09*	−0.005	0.216***
MO3 x Duration	−0.188***	0.207***	−0.189***	0.161***
MO4 x Duration	−0.235***	0.259***	−0.294***	0.089**
MO5 x Duration	−0.394***	0.299***	−0.488***	0.223***
SC2 x Duration	0.02	0.32***	−0.079*	0.134***
SC3 x Duration	−0.092**	0.044	−0.149***	0.015
SC4 x Duration	−0.14***	0.229***	−0.396***	0.1***
SC5 x Duration	−0.308***	0.329***	−0.485***	0.274***
UA2 x Duration	0.021	0.194***	−0.094***	0.076
UA3 x Duration	−0.114***	0.006	−0.194***	0.162***
UA4 x Duration	−0.34***	0.205***	−0.367***	0.176***
UA5 x Duration	−0.351***	0.233***	−0.525***	0.157***
PD2 x Duration	0.043	0.191***	−0.153***	0.187***
PD3 x Duration	−0.079*	0.039	−0.241***	0.065
PD4 x Duration	−0.507***	0.205***	−0.78***	0.393***
PD5 x Duration	−0.909***	0.532***	−1.183***	0.561***
AD2 x Duration	−0.102**	0.128***	−0.13***	0.214***
AD3 x Duration	−0.226***	0.178***	−0.311***	0.078*
AD4 x Duration	−0.736***	0.092*	−0.817***	0.329***
AD5 x Duration	−1.032***	0.573***	−1.356***	0.703***

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute

6.8 Discussion

The purpose of this chapter was to investigate whether the choice of anchor by researchers had an impact on the valuation of the EQ-5D-5L health states by respondents[2]. In the literature, researchers have anchored DCE_{TTO} choice sets by using a third option of either immediate death or full health (Jonker et al., 2017; Norman, Cronin, et al., 2013). The impact of these anchors was investigated in the context of populations from two different countries, Peru and Denmark.

6.8.1 Summary of findings

The *full health* arms produced a wider utility range than the *immediate death* arms. In addition, respondents in the *full health* arms valued a higher proportion of health states as worse than death compared to respondents in the *immediate death* arms. This may be a reflection of the unwillingness of respondents in the *full health* arms to trade off against full health. In contrast, the prospect of immediate death may have prompted respondents in the *immediate death* arms to be more willing to accept living in worse health states. It could also be the case that by presenting death and a particular health state as a comparison, respondents had to explicitly consider where the specified health state sat on the full health to death scale. This may mean that respondents are more likely to consider where death sits on the QALY scale. Results from this study provide evidence that respondent preferences are impacted by the anchor used in the choice sets. These findings were robust across countries.

Overall, accounting for preference heterogeneity in the data did improve the model fit. The results demonstrate that preference heterogeneity between respondents exists. Apart from detected preference heterogeneity in the parameter estimates for levels of dimensions, it was also noted that respondents had significant preference heterogeneity

[2] Most countries use TTO methods which includes death in the scale. However, TTO methods can be quite labour intensive as they are often completed as face to face interviews. Hence, alternative methods for health state valuation including DCEs have been explored. The use of DCEs in health state valuation started as a paired approach i.e. comparison of two health states, which then used external data such as TTO data to anchor the DCE values. But this had the issues of using different methods and different samples even. This led to the development of the latent DCE with duration as an extra attribute (Bansback et al., 2012). However, this is not anchored to death. This led to the development of the DCE with immediate death as a third choice set (Viney et al., 2014) and is used in government decision making in Australia. Then came the Dutch group with DCE with full health (Jonker et al., 2017; Jonker et al., 2012). The use of anchoring in DCEs is a relatively recent development. Further information about the use of DCEs in health state valuation can be found in (Mulhern et al., 2019).

in regards to duration. This raises the question of whether and how to incorporate this into QALY weights used in cost utility analyses. For example, heterogeneity as identified using the MXL model could potentially be used as an indicator for the amount of variation within a population for a sensitivity analysis.

6.8.2 Limitations and direction for future research

The use of DCE_{TTO} has grown in recent times and it is becoming accepted as a method for health state valuation. As far as this author is aware, this was the first study to compare the anchor used in choice sets for the valuation of the EQ-5D-5L, especially across two countries.

This study has a number of limitations that impact on the comparisons made and the interpretation of the results. The sample in each arm was quite small, with only 200–250 respondents. Hence, it is uncertain how generalisable the current findings are. Future studies could replicate the tasks in this study with a larger sample to investigate whether the findings hold.

It must be acknowledged that, unlike previous studies that include immediate death as the third option in the DCE_{TTO} choice sets (Viney et al., 2014, Norman et al., 2013), information regarding the full ranking of health states and immediate death was limited. That is, only the preference between health state B and immediate death was explicitly asked for, and therefore the preference between health state A and immediate death could not be elicited for all response patterns. Future studies that seek to compare framing of DCE_{TTO} choice sets could use the best and worst elicitation approach in order to obtain a full ranking of preferences for when immediate death is used as the third option versus when full health for a shorter duration is used as the third option in the choice set. It would be interesting to compare results from the current study, which uses the two-stage format, with those from a best-worst format.

There are other features of the study design that could limit the generalisability of the results. For example, DCE_{TTO} is usually administered online as a standalone task, rather than as an add-on, and in this study data were collected face-to-face following an EQ-VT interview. However, there is evidence from previous studies that the mode of administration does not impact responses once any potential demographic differences are controlled for, although this was in the context of choice sets with two options

(Mulhern et al., 2013) Respondent fatigue could also have been an issue, as the DCE_{TTO} was administered last.

6.8.3 Implications of findings

Evidence from this study suggests that when respondents are presented with different anchors in a DCE, it can impact the preferences elicited. Preferences were more a reflection of how researchers chose to present choice sets than of the true underlying preferences of respondents. More specifically, the current study suggest that in the valuation of health states, preferences elicited by respondents can be influenced by the choice of anchor used in DCE_{TTO} choice sets.

It should be noted that there is no 'gold' standard for how to anchor choice sets. Both anchors are just different methods to obtain utility weights, and both require respondents to engage in hypothetical thinking. However, health decision-makers and researchers conducting health evaluation studies such as cost utility analyses should be aware that QALY weights are sensitive to the anchor used in DCE_{TTO} choice sets, as evidenced from the current study.

Although from a statistical perspective value sets from different sources should be able to be compared, it still may not be appropriate to make direct comparisons. Evidence from this study suggests that value sets produced from DCE_{TTO} choice sets anchored to full health will produce more values considered worse than death than value sets where the anchor was immediate death. As a consequence, the choice of anchor in the preference elicitation task should be made explicit in health valuation studies. Health decision-makers should also be cautious about using QALY weights obtained from other countries. In particular, there was evidence of scale differences between Peru and Denmark, indicating that results should be considered separately by country.

Chapter 7. Discussion

Introduction

Discrete choice experiments (DCEs) can provide analysts and policy-makers with valuable information about population preferences when revealed data are unavailable. DCEs can contribute to shaping health policy by ensuring that it is reflective of what is of most value to patients, and to society more generally. However, DCEs can only be useful to decision-makers if DCE respondents are able to understand the choices they are being asked to make and are interpreting choice sets in the way the researcher has designed them. DCE researchers need to ensure that the decisions they make about the structure, format and design of the DCE are conducive to respondents making meaningful choices that reflect their 'true' underlying preferences.

A key advantage of DCEs is the control that researchers have over the development, presentation and design of the DCE, which means they can target the experiment to the specific policy questions. However, this can be a double-edged sword as, ultimately, the researcher has designed the data collection. This means that DCE researchers must carefully choose all features of the DCE, from the construction of the choice sets, to how choice sets are presented to respondents, to the content before and after the choice sets. A goal of this thesis was to explore the extent to which decisions researchers make about the features of the DCE drive respondent preferences as opposed to estimating respondents' actual underlying preferences.

This thesis aimed to investigate different ways information can be presented in a DCE and to understand whether choices about how the information in a DCE is presented influence the stated preferences of respondents. To investigate these aims, a scoping review was undertaken (Chapter 3) and then three empirical studies (Chapters 4–6) covering two different areas of policy were conducted.

The goal of the empirical studies was to investigate the impact of different forms of DCE information presentation commonly used in the health literature. The first two empirical studies investigated the impact of varying the amount and format of information provided before the choice tasks are presented. The third empirical study

investigated the presentation of an anchor in each choice task. These studies aimed to investigate whether and how these presentation differences impact DCE findings. The impact of the amount of information provided and the use of alternative presentation formats to text was investigated in the context of comparing patient and general population preferences. The studies reported in Chapters 4 and 5 aimed to understand what features different population groups prefer in an assessment tool for chemotherapy-induced peripheral neuropathy (CIPN). The study reported in Chapter 4, described the preferences of a cancer patient sample, whereas that of Chapter 5 described the preferences of a general population sample. Preferences were compared between the two populations to identify differences; this is important, because a common criticism of DCEs using general populations is that non-patients do not have experience of the specific health problem and so their preferences are not relevant or important. Chapter 6 investigated the impact of different anchors on estimated preferences of health states described using the EQ-5D quality of life questionnaire. In this study, preferences were compared across two different ways of describing choice sets. Choice sets included either immediate death or full health for a shorter duration as the third option.

Chapter 3: Scoping review

Summary of findings

The scoping review identified and summarised studies that have explored the impact of different ways that DCE information can be presented to respondents. Studies were included if they used a between-subjects design where respondents were allocated to two or more arms and the DCE presented information differently in some way between arms. In total, there were 69 studies included in the review. The identified studies varied widely in terms of discipline area and country where the study was conducted, and in terms of the amount of detail relating the design and construction of the DCE.

In the review, it was found that the ways information can be presented in a DCE could be classified into three forms:

1) researchers could vary the presentation of the DCE in terms of the amount of information provided;

2) researchers could vary the number of levels, attributes or options in choice sets (i.e. the structure of the choice sets); or
3) researchers could vary the content or phrasing of the text presented in the DCE or use alternative presentation formats to text, such as pictures, graphs and other visual methods.

Varying the text information was most common, although varying the DCE in terms of amount of information provided was also common. The studies included in the review were most likely to use one type of DCE presentation difference, although there were a few studies that used two or three differences.

Findings from the scoping review suggest that DCE results can be sensitive to how information is presented to respondents in the DCE, with over 95% of included studies finding a statistically significant difference between the results in the different arms. This is in contrast to the findings of a review of the environmental economics literature, in which Rakotonarivo et al. (2016) report that in the majority of included studies respondents were not sensitive to differences in information presented in a DCE. This suggests the possibility of a publication bias towards studies that find statistically significant differences in results in disciplines outside of environmental economics.

Implications of findings

In health DCEs, extra information is often presented to respondents to improve respondent understanding of complicated health or medical terminology with which respondents may not otherwise be familiar. Alternatively, for DCEs used for valuation purposes, the inclusion of some information and the way it is presented may be necessary to elicit the valuation data. In both instances, there is no intention for the presentation of information to influence respondent choices, but rather to aid respondent understanding of the choice sets or as an essential component of data elicitation.

Evidence from the scoping review indicates that the results from a DCE can be influenced by the way information is presented to respondents. However, evidence from a similar review (Rakotonarivo et al., 2016) suggests that more often than not, results from a DCE are not influenced by the way it was presented Overall, it appears there are instances where presentation differences impact results and cases where they

do not. This is an important finding, as it means that how information is presented in a DCE may not necessarily influence elicited preferences, and therefore impact true underlying preferences.

This raises the important question of whether it is possible to design health DCEs such that the way in which information is presented in a DCE does not impact respondent choices. The three empirical studies that follow the scoping review aim to answer this question. The first two empirical studies investigated whether providing extra information and the use of alternative formats to text could assist in improving respondent understanding of choice sets and whether this had an impact on their preferences. The third empirical study investigated whether two alternative ways of presenting a choice set for the purposes of eliciting QALY weights made a difference to the values elicited.

Chapters 4 and 5: Preferences for the assessment of CIPN

Summary of findings

In the first two empirical studies (covered in Chapters 4 and 5), preferences for a health intervention were investigated. Preferences from two different population groups, patients and the general population, were explored. Differences in how information was presented in the DCE to the general population was used to explore how information can be used to make general population preferences more relevant to the decision-making context. The specific case used was preferences for the features in CIPN assessment tools, which was an important policy area and one highly relevant to clinicians.

In the first empirical study, patients with cancer who were currently receiving, or had previously received, chemotherapy treatment were recruited. The DCE included some introductory material describing CIPN as well as text and pictures to explain different types of physical tests that could be included as part of the CIPN assessment tool. The study found that patients held a greater concern for the impact of CIPN on their daily life compared to experiencing symptoms alone. Patients were also supportive of the inclusion of a physical test of some sort as part of the assessment. Patients also wanted some measure of consultation about, and control over, their future treatment. For example, there was a strong preference that the results of any assessment tool should

inform a joint decision made between the doctor and patient about future treatment options. There was noticeable preference heterogeneity in terms of the strength of preference that patients had for particular features of a CIPN assessment tool.

The second empirical study, reported in Chapter 5, followed on from the first study by investigating preferences for the features of a CIPN assessment tool, but this time amongst the general population. As with many DCEs, a general population sample was used as a proxy for a patient sample given that, if fully informed, it seems reasonable to expect that their preferences would approximate those of the patient sample. The general population sample was split into two arms, with the amount of information provided to respondents differing between the arms. Respondents allocated to Arm 1 received the same amount of information as the patient sample, to allow comparison between the patient and general population samples. Respondents allocated to Arm 2 received additional information in the form of a video and moving images; the aim of the additional information was to bridge the gap in health experiences between cancer patients and non-cancer patients.

Compared to Arm 1, respondents in Arm 2 reported having a better understanding of the choice task. However, no significant differences were found in terms of preferences between the two arms, suggesting that varying the amount of information provided did not lead to significant differences in preferences in the general population sample. Scale differences between Arm 1 of the general population sample and the patient sample were noted, with the scale of Arm 1 of the general population being twice as large as that of Arm 2 of the general population sample and of the patient sample. The increased amount of information provided in Arm 2 seems to have reduced the scale difference between the patient and general population samples. That is, there were differences in terms of heterogeneity of preferences between the patient and general population samples. Provision of extra information reduced these differences in preference heterogeneity between the two population types.

A qualitative comparison of model results suggested that preference patterns, were generally similar between the general population sample and patient sample. Both the patient and general population samples preferred an assessment that picks up small changes in their condition, whether it was important or not. Both samples were indifferent regarding the inclusion of a questionnaire and did not mind adding extra

time to their usual appointment in order to assess for CIPN. The general population sample were willing to add an extra 30 minutes to the consultation time whereas the patient sample was willing to go even further by having a separate appointment devoted exclusively to assessing for CIPN. Both samples preferred that any changes to treatment or care resulting from assessment results be a combined decision between the patient and doctor. Both were also in favour of the inclusion of a physical test in the assessment tool, although the specific physical test that was most preferred did differ by sample. The key difference in terms of preferences was that the patient sample was more concerned about the quality of life implications of living with CIPN symptoms than were general population respondents in either arm. This may suggest that the way the quality of life information was presented to the general population did not fully capture what cancer patients may experience.

Implications of findings

Findings from Chapters 4 and 5 provide evidence that, at least within the same population sample, the amount of information provided does have an effect on respondent understanding of choice sets. Increasing the amount of information provided improved respondents' stated understanding of choice sets, although this did not appear to influence the preferences exhibited by respondents. This is an important finding, as it suggests that DCE researchers can be more confident that increasing the amount of information provided aids respondents and helps to establish their underlying preferences.

Another interesting finding was that the amount of information provided reduced heterogeneity. In this experiment, the provision of extra information reduced the scale differences between the two different populations. As far as this author is aware, this is the first study to examine the effect of the amount of information provided on scale differences. Although scale differences were not eliminated entirely in the study, it is possible that this could be achieved by providing further information, or perhaps by targeting knowledge areas that have been found to be lacking in one population compared to another. For instance, it was found that the general population sample was not concerned about the quality of life implications of living with CIPN symptoms while the patient sample were. This was despite the provision of text information about the potential impacts of CIPN on daily life. This finding was consistent with previous studies

which have found that general population samples do not value quality of life implications as much as patients do (Najafzadeh et al., 2019; Ogorevc et al., 2019).

These findings suggest that increasing the amount of information provided has a role to play when results from two different population samples are not easily compared. In previous studies that have compared patient and general population samples, results were able to be meaningful compared by dividing through by a common denominator to obtain the marginal rates of substitution (MRS) (Fiebig et al., 2009). This eliminates the scale, allowing results to be compared between different population samples (Vass, Wright et al., 2018). Previous studies that have examined differences between different population samples have used a common denominator that allows a natural or intuitive interpretation of results. For instance, utility weights are a common measure that can be used to directly compare results between different samples in studies that valued the EQ-5D-5L by dividing parameter estimates by the duration parameter estimate (Ludwig et al., 2021; Ogorevc et al., 2019). Najafzadeh et al. (2019) used risk of fatal bleeding as the common denominator to calculate MRS. Results could be interpreted as the average magnitude of risk that participants are willing to accept in one of the attributes in exchange for achieving a 1% decrease in the risk of fatal bleeding. In these examples, MRS provides a meaningful and useful interpretation of results. However, there may be cases where a DCE does not have a common denominator that can be used to provide an intuitive interpretation of MRS. In such cases, scale differences could potentially be removed by providing increased and targeted information to general population samples. This would allow patient and general population sample results to be combined together for comparison using formal statistical analyses.

This study is consistent with previous literature, in that there appears to be a lack of concern about the implications of reduced quality of life among general population samples compared to patient samples. This study also adds to the debate on the use of patient and general population samples by providing evidence that general population samples may be a viable alternative to patient populations if provided the right amount and type of information. For instance, general population samples could serve as opportune proxy for patient preferences where it may be too expensive or difficult to recruit a patient sample. There may also be situations where researchers may be interested in exploring preferences in relation to rare conditions where the number of

affected patients willing to participate in research may be too low for a meaningful sample. Or it might be the case that particular patient populations are unable to complete DCEs – for instance, patients with advanced dementia. There are also many other advantages to using general population samples, which often require less stringent ethics application processes and can be cheaper to recruit.

Chapter 6: Anchoring to immediate death versus full health

Summary of findings

The third empirical study, described in Chapter 6, explored a different and widely used policy-relevant application of DCEs to test the role of presentation of information. In this study, the information presentation question focused on a specific type of framing, the framing of choices in the context of valuation of health states. In particular, the study explored different approaches to providing an anchor in a DCE task, against which respondents were assessing options.

In valuation studies, DCEs have been used to elicit utility weights. In the DCE, respondents are faced with a series of choice sets where options are health states for a specified duration; hence, respondents are asked to compare one health state for duration *x* to another health state for duration *y* (where *x* and *y* may be equal). Some studies also include an additional option of immediate death, while others include an option of full health but for a shorter duration (in which case $x = y$ above). This study explored the impact on preferences of these two anchoring methods using the EQ-5D-5L (the most widely used general QoL instrument). Respondents completed all choice sets presented to them in two stages. In stage 1, they were asked to choose between two health states, health state A and health state B, each presented for a specific duration. In stage 2, respondents were asked to choose between health state B and immediate death (in one arm) or between health state B and full health for a shorter duration (in the other arm).

The study found that the choice of anchor made a difference to the range of the resulting value sets. It was found that respondents who were assigned to the arm to choose between health state B and full health considered a greater proportion of health states to be worse than death compared to those asked to choose between health state B and

immediate death. This finding was robust and consistent across general population samples in both Peru and Denmark.

Findings from this study provide evidence that respondents are sensitive to the type of anchor presented to them in DCEs, at least in the context of the framing of text in choice sets. Responses are more reflective of the choice of anchor rather than the underlying preferences of respondents. The impact of the anchor is also consistent with prospect theory (Kahneman & Tversky, 1979). In particular, respondent choices were consistent with the reflection effect whereby respondents are risk averse when faced with a positive outcome i.e. full health, and are risk taking when faced with a negative outcome i.e. death. Respondents are much more receptive to worse outcomes when faced with death. In contrast, respondents are less willing to live in worse health states when faced with the prospect of living in full health, albeit for a shorter period of time. It is also possible that loss version may partially explain some of these findings. Respondents may be behaving to the principle that 'loss looms larger than gains' (Gal & Rucker, 2018; Kahneman & Tversky, 1979) by choosing worse outcomes in the face of death.

Implications of findings

It should be noted that there is no gold standard for the best way to anchor choice sets. The purpose of the anchor is to allow explicit consideration for where 'dead' lies on the utility scale. This has been done in previous studies by asking respondents to compare health states to immediate death or full health (Jonker et al., 2017; Viney et al., 2014). As far as this author is aware, this is the first study to compare the impact of anchoring choice sets to immediate death versus full health in the context of valuing the EQ-5D-5L. Findings from this study have implications for utility weights derived from DCEs that are used in economic evaluation. It is important that researchers compare utility weights that use the same anchor, otherwise results can be misleading. For some valuation studies, it may be considered unethical to ask a particular population (e.g. small children, terminally ill patients) to explicitly consider immediate death when comparing health states. In such circumstances, anchoring to full health may be considered a more ethical method for valuing health states.

The impact of the choice of anchor was found in this study to be much larger on respondent preferences than the impact of variations in amount of information

provided. Within the anchoring effect literature, it has been noted that the anchoring effect is much more prominent when respondents are not familiar with or have no previous knowledge of the task (Furnham & Boo, 2011; van Exel et al., 2006; Wilson et al., 1996). Several studies have looked at mitigating the anchoring effect, with mixed results, although these were not in the health evaluation space (Fudenberg et al., 2012; Furnham & Boo, 2011; LeBoeuf & Shafir, 2009). It could be argued that the current study has attempted to minimise non-familiarity with the task, as the way in which the study was conducted did allow respondents to familiarise themselves with related concepts and provided the opportunity to ask clarifying questions. In particular, the current study included face-to-face interviews and was completed after a number of cTTO tasks in which respondents were asked to make trade-offs regarding different health states, some better and some worse than death and some involving contemplation of full health and duration. As such, respondents would be somewhat familiar with elements of the DCE_{TTO} prior to completing it. They were also able to clarify any questions they may have had, as an interviewer was present. Another point to consider is that the current study did require respondents to think about very serious concepts – i.e. matters of life and death – that are not bounded by the borders of countries. This may have contributed to the robust anchoring effect seen in the study across countries.

Limitations of findings and directions for future research

The scoping review and the three empirical studies have provided important and interesting findings that add to the DCE and health literature. However, when considering the implications of these findings, several limitations of the findings from this thesis should be noted. The aim of the scoping review was to understand what is currently available in the existing literature. As such, key details of the DCE have been reported without an assessment of the quality of studies, nor was there any assessment in terms of the suitability of models used for data modelling in studies. Future reviews could investigate whether findings differ if a quality selection criterion is applied so that only 'high-quality' DCEs are included. There is also the opportunity for future studies to narrow the scope to look only at studies that use a single type of presentation difference and investigate in detail the impact on results.

It also should be acknowledged that the samples used and compared in Chapters 4 and 5 were different in many respects. The patient sample consisted of volunteers. A large majority of respondents in the patient sample were female, older and were university-educated. The sample was also small, consisting of only 117 respondents. In contrast, the general population sample was recruited through an online panel and was representative of the Australian population in age and gender. The sample was also much larger, at 335 respondents. Future studies could recruit patient and general population samples that were comparable in terms of gender, age and education. Future studies could also explore further the effects of varying the type and amount of information provided on scale differences between patient and general population samples. For instance, future studies could examine whether scale differences can be further reduced, or even eliminated altogether, if quality of life implications of living with certain side effects or conditions are highlighted, perhaps by providing text/video testimonials of patients experiencing the difficulties of living with them. It would also be interesting to explore whether differences in scale could be narrowed or reduced entirely between other types of populations (e.g. between a provider/clinician sample and a patient or general population sample).

Chapter 6 included two different anchors for choice sets. It should be noted that the choice set format for choice sets with immediate death as an anchor was used differently from previous studies. Choice sets were shown as two-stage tasks in this study. In previous studies with immediate death as an anchor, all three options were shown simultaneously in the choice set, with respondents asked to indicate the best and worst health state. In this study, to ensure comparability of results, both sets of choice sets, regardless of anchor, were shown as a two-stage task. Future studies could compare results from using immediate death or full health as the anchor in choice sets using the best-worst format instead.

Conclusions

DCEs have become an important tool in health policy analysis and have been used to answer many health policy questions. They have been used to understand the impact of health programs (Brown et al., 2016; Franco et al., 2016; Salloum et al., 2015), as well as the design of health programs, systems and policies (Baji et al., 2016; Munger et al., 2017; Whitty et al., 2015). DCEs have also been a useful tool for understanding patient

preferences by eliciting information about how they value features of health assessments and treatments (Flood et al., 2017; Marshall et al., 2016; Wright et al., 2017).

Given the increasingly important role of DCEs as a policy tool, it is essential that there is an understanding of how the presentation of the DCE can influence results. This thesis has attempted to investigate this question through two distinct case studies covering different applications of DCEs.

What are the main ways a DCE can be presented to respondents?

The findings from this thesis have important implications for how researchers think about designing a DCE and how they interpret the findings. There are many DCE studies that have sought to investigate the impact of presenting DCEs in different ways to respondents, as evidenced by the number of studies included in the scoping review alone. However, a pattern evident in the literature is the lack of comparability of findings from such studies and a lack of clarity around how to use such findings to inform future DCEs, particularly within health.

This thesis presented a framework for classifying the different types of presentation differences used in DCE studies. Specifically, three main types were identified: varying the amount of information provided, varying structure of choice sets, and varying how text information is presented or using alternative formats to text. Any impacts of these presentation differences were also examined.

Does changing how the DCE is presented to respondents lead to differences in results?

This thesis also adds to the literature by examining the impact of DCE presentation differences on findings, with implications on the design of future DCEs. The studies were conducted in the context of how DCEs are used in health services research. For many health researchers using DCEs, it important that elicited preferences reflect the true underlying preferences of respondents rather than being a function of how features of the DCE were presented to respondents.

The findings of this thesis demonstrate that DCE results are sensitive to the choice of anchor used in choice sets. In particular, anchoring can affect how health states are

valued relative to death. DCE results are also sensitive to variation in the amount of information presented in the DCE, although not in the same way nor to the same extent. Variation in information has no apparent effect on respondent preferences; rather, increasing the amount of information presented in the DCE can aid a naïve sample of respondents in understanding the topic and terminology without affecting preferences. It also has effects in terms of reducing the heterogeneity between an experienced and naïve sample.

Final remarks

The impacts of presentation differences in a DCE are complex. However, this complexity can be broken down by researchers identifying the type of presentation difference of interest or of most relevance to their research.

Another important outcome of this thesis was that when researchers are thinking about how to present the DCE to respondents, they should be aware that different ways of presenting a DCE have different potential influences on findings. This thesis provides evidence that the choice of presentation type can have potential impacts in terms of respondent understanding, on scale as well as on elicited preferences.

References

Amano, I., Kurisu, K., & Hanaki, K. (2016). Evaluation of consumers' preferences on drinking waters considering the information provision effect. *Water Science and Technology: Water Supply, 16*(4), 1057–1067. https://doi.org/10.2166/ws.2016.019

Andersen, L. M. (2014). Obtaining reliable likelihood ratio tests from simulated likelihood functions. *PLoS ONE*, *9*(10), 1–12. https://doi.org/10.1371/journal.pone.0106136

Andrade, L. F., Ludwig, K., Goni, J. M. R., Oppe, M., & de Pouvourville, G. (2020). A French value set for the EQ-5D-5L. *Pharmacoeconomics*, *38*, 413–25.

Antoniou, C., & Kostovasilis, K. (2017). How may external information affect traffic risk perception? *Journal of Transportation Safety and Security*, *9*(3), 347–368. https://doi.org/10.1080/19439962.2016.1212444

Arrua, A., Curutchet, M. R., Rey, N., Barreto, P., Golovchenko, N., Sellanes, A., Velazco, G., Winokur, M., Gimenez, A., & Ares, G. (2017). Impact of front-of-pack nutrition information and label design on children's choice of two snack foods: comparison of warnings and the traffic-light system. *Appetite*, *116*, 139–146. https://doi.org/10.1016/j.appet.2017.04.012

Augustovski, F., Belizán, M., Gibbons, L., Reyes, N., Stolk, E., Craig, B. M., & Tejada, R. A. (2020). Peruvian valuation of the EQ-5D-5L: a direct comparison of time trade-off and discrete choice experiments. *Value in Health*, *23*(7), 880–888. https://doi.org/10.1016/j.jval.2020.05.004

Australian Bureau of Statistics. (2018). *4363.0 - National health survey: users' guide, 2017–18*. Commonwealth of Australia. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4363.0~2017-18~Main%20Features~Health%20conditions~4

Australian Bureau of Statistics. (2020). *3101.0 – Australian demographic statistics, Jun 2019*. https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Jun%202019?OpenDocument

Baba, Y., Kallas, Z., Costa-Font, M., Maria Gil, J., & Realini, C. E. (2016). Impact of hedonic evaluation on consumers' preferences for beef attributes including its enrichment with n-3 and CLA fatty acids. *Meat Science*, *111*, 9–17. https://doi.org/10.1016/j.meatsci.2015.08.005

Baji, P., García-Goñi, M., Gulácsi, L., Mentzakis, E., & Paolucci, F. (2016). Comparative analysis of decision maker preferences for equity/efficiency attributes in reimbursement decisions in three European countries. *The European Journal of Health Economics, 17*(7), 791–799. https://doi.org/10.1007/s10198-015-0721-x

Bandos, H., Melnikow, J., Rivera, D. R., Swain, S. M., Sturtz, K., Fehrenbacher, L., Wade, J. L., III, Brufsky, A. M., Julian, T. B., Margolese, R. G., McCarron, E. C., & Ganz, P. A. (2017). Long-term peripheral neuropathy in breast cancer patients treated with adjuvant chemotherapy: NRG oncology/NSABP B-30. *JNCI: Journal of the National Cancer Institute, 110*(2). https://doi.org/10.1093/jnci/djx162

Bansal, P., & Daziano, R. A. (2018). Influence of choice experiment designs on eliciting preferences for autonomous vehicles. *Transportation Research Procedia, 32*, 474–481. https://doi.org/10.1016/j.trpro.2018.10.044

Bansback, N., Brazier, J., Tsuchiya, A., & Anis, A. (2012). Using a discrete choice experiment to estimate health state utility values. *Journal of Health Economics, 31*(1), 306–318. https://doi.org/10.1016/j.jhealeco.2011.11.004

Bao, T., Basal, C., Seluzicki, C., Li, S. Q., Seidman, A. D., & Mao, J. J. (2016). Long-term chemotherapy-induced peripheral neuropathy among breast cancer survivors: prevalence, risk factors, and fall risk. *Breast Cancer Research & Treatment, 159*(2), 327–333. https://doi.org/10.1007/s10549-016-3939-0

Bateman, I. J., Day, B. H., Jones, A. P., & Jude, S. (2009). Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management*, *58*(1), 106–118.

Belle, N., & Cantarelli, P. (2021). Your money, your life, or your freedom? A discrete-choice experiment on trade-offs during a public health crisis. *Public Administration Review* (pre-print). https://doi.org/10.1111/puar.13429

Benning, T. M., Kimman, M. L., Dirksen, C. D., Boersma, L. J., & Dellaert, B. G. C. (2012). Combining individual-level discrete choice experiment estimates and costs to inform health care management decisions about customized care: the case of follow-up strategies after breast cancer treatment. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research, 15*(5), 680–689. https://doi.org/10.1016/j.jval.2012.04.007

Bijlenga, D., Bonsel, G. J., & Birnie, E. (2011). Eliciting willingness to pay in obstetrics: comparing a direct and an indirect valuation method for complex health outcomes. *Health Economics, 20*(11), 1392–1406. https://doi.org/10.1002/hec.1678

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., & Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological Economics*, *18*(3), 243–253.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*(2), 271–292. https://doi.org/10.1016/S0167-6296(01)00130-8

Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., Johnson, F. R., & 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. https://doi.org/10.1016/j.jval.2010.11.013

Brooks, K., & Lusk, J. L. (2010). Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *American Journal of Agricultural Economics*, *92*(4), 1229–1241.

Brown, Z. S., Kramer, R. A., Ocan, D., & Oryema, C. (2016). Household perceptions and subjective valuations of indoor residual spraying programmes to control malaria in northern Uganda. *Infectious Diseases of Poverty, 5*(1), 100. https://doi.org/10.1186/s40249-016-0190-1

Burgess, L., Street, D. J., & Wasi, N. (2011). Comparing designs for choice experiments: a case study. *Journal of Statistical Theory and Practice, 5*(1), 25–46. https://doi.org/10.1080/15598608.2011.10412048

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*(24), 4279–4292. https://doi.org/10.1002/sim.2673

Burton, M., & Rigby, D. (2012). The self selection of complexity in choice experiments. *American Journal of Agricultural Economics, 94*(3), 786–800. https://academic.oup.com/ajae/issue

Cantillo, J., Martín, J. C., & Román, C. (2020). Discrete choice experiments in the analysis of consumers' preferences for finfish products: a systematic literature review. *Food Quality and Preference*, 84, 103952. https://doi.org/10.1016/j.foodqual.2020.103952

Carlsson, F., García, J. H., & Löfgren, Å. (2010). Conformity and the demand for environmental goods. *Environmental and Resource Economics*, *47*(3), 407–421.

Clark, M. D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, E. W. (2014). Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*, *32*(9), 883–902.

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics, 21*(6), 730–741. https://doi.org/10.1002/hec.1739

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd edn)(Hillsdale, NJ: Lawrence Earlbaum).

Cole, A., Shah, K., Mulhern, B., Feng, Y., & Devlin, N. (2018). Valuing EQ-5D-5L health states 'in context' using a discrete choice experiment. *European Journal of Health Economics, 19*(4), 595–605. https://doi.org/10.1007/s10198-017-0905-7

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures: new approaches to missing data. *Psychological Methods*, *6*(4), 330–351.
Czajkowski, M., & Budziński, W. (2019). Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling, 31*, 73–85. https://doi.org/10.1016/j.jocm.2019.04.003
de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics, 21*(2), 145–172. https://doi.org/10.1002/hec.1697
Dellaert, B. G. C., Donkers, B., & van Soest, A. (2012). Complexity effects in choice experiment-based models. *Journal of Marketing Research, 49*(3), 424–434. https://doi.org/10.1509/jmr.09.0315
De-Magistris, T., Gracia, A., & Nayga Jr, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics, 95*(5), 1136–1154. https://doi.org/10.1093/ajae/aat052
Determann, D., Lambooij, M. S., Steyerberg, E. W., de Bekker-Grob, E. W., & de Wit, G. A. (2017). Impact of survey administration mode on the results of a health-related discrete choice experiment: online and paper comparison. *Value in Health, 20*(7), 953–960. https://doi.org/10.1016/j.jval.2017.02.007
Devarasetty, P. C., Burris, M., & Douglass Shaw, W. (2012). The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation Research, 46*(8), 1227–1240. https://doi.org/10.1016/j.tra.2012.05.002
Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol Group: past, present and future. *Applied Health Economics and Health Policy, 15*(2), 127–137. https://doi.org/10.1007/s40258-017-0310-5
Diener, A., O'Brien, B., & Gafni, A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics, 7*(4), 313–326. https://doi.org/10.1002/(SICI)1099-1050(199806)7:4
Dolders, M. G. T., Zeegers, M. P. A., Groot, W., & Ament, A. (2006). A meta-analysis demonstrates no significant differences between patient and population preferences. *Journal of Clinical Epidemiology, 59*(7), 653–664. https://doi.org/10.1016/j.jclinepi.2005.07.020
Domínguez-Torreiro, M. (2014). Alternative experimental design paradigms in choice experiments and their effects on consumer demand estimates for beef from endangered local cattle breeds: an empirical test. *Food Quality and Preference, 35*, 15–23. https://doi.org/10.1016/j.foodqual.2014.01.006
Drennan, J. (2003). Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57–63. https://doi.org/10.1046/j.1365-2648.2003.02579.x
Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (3rd ed.). Oxford University Press.
Eldessouki, R., & Smith, M. D. (2012). *Health care system information sharing: a step toward better health globally*. https://tools.ispor.org/peguidelines/
European Network for Health Technology Assessment. (2015). *Methods for health economic evaluations – a guideline based on current practices in Europe*. https://www.eunethta.eu/wp-content/uploads/2018/03/Methods_for_health_economic_evaluations.pdf
EuroQol Research Foundation. (2019). *EQ-5D-5L user guide*. https://euroqol.org/publications/user-guides
EuroQol Research Foundation. (2021). *About us*. https://euroqol.org/euroqol/
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. https://doi.org/10.1016/j.socscimed.2009.03.002

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science, 29*(3), 393–421. https://doi.org/http://mktsci.journal.informs.org/content/by/year

Finkelstein, E. A., Mansfield, C., Wood, D., Rowe, B., Chay, J., & Ozdemir, S. (2017). Trade-offs between civil liberties and national security: a discrete choice experiment. *Contemporary Economic Policy, 35*(2), 292–311. https://doi.org/10.1111/coep.12188

Flood, E. M., Bell, K. F., de la Cruz, M. C., & Ginchereau-Sowell, F. M. (2017). Patient preferences for diabetes treatment attributes and drug classes. *Current Medical Research and Opinion, 33*(2), 261–268. https://doi.org/10.1080/03007995.2016.1253553

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: what it can do for health care research and how to do it. *Journal of Health Economics, 26*(1), 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002

Franco, M. R., Howard, K., Sherrington, C., Rose, J., Ferreira, P. H., & Ferreira, M. L. (2016). Smallest worthwhile effect of exercise programs to prevent falls among older people: estimates from benefit–harm trade-off and discrete choice methods. *Age and Ageing, 45*(6), 806–812. https://doi.org/10.1093/ageing/afw110

Freeman, A. M. (2003). *The measurement of environmental and resource values: theory and methods*. Resources for the Future.

Frew, E. J., Whynes, D. K., & Wolstenholme, J. L. (2003). Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. *Medical Decision Making*, *23*(2), 150–159.

Froberg, D. G., & Kane, R. L. (1989). Methodology for measuring health-state preferences – III: population and context effects. *Journal of Clinical Epidemiology*, *42*(6), 585–592.

Fudenberg, D., Levine, D. K., & Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics*, *4*(2), 131–145.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gajic, A., Cameron, D., & Hurley, J. (2012). The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. *European Journal of Health Economics, 13*(6), 789–799. https://doi.org/10.1007/s10198-011-0332-0

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? Journal of Consumer Psychology, 28(3), 497-516.

Garza, A. G,, & Wyrwich, K. W. (2003). Health utility measures and the standard gamble. *Academic Emergency Medicine, 10*(4), 360–363. https://doi.org/10.1111/j.1553-2712.2003.tb01349.x

Gong, J., Zhang, Y., Yang, Z., Huang, Y., Feng, J., & Zhang, W. (2013). The framing effect in medical decision-making: a review of the literature. *Psychology, Health & Medicine*, *18*(6), 645–653.

Goodwin, E., Davey, A., Green, C., & Hawton, A. (2021). What drives differences in preferences for health states between patients and the public? A qualitative investigation of respondents' thought processes. *Social Science & Medicine, 282*, 114–150. https://doi.org/10.1016/j.socscimed.2021.114150

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *Journal of Marketing, 54*(4), 3. https://doi.org/http://dx.doi.org/10.2307/1251756

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological, 37*(8), 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2

Grisolía, J. M., Longo, A., Hutchinson, G., & Kee, F. (2018). Comparing mortality risk reduction, life expectancy gains, and probability of achieving full life span, as alternatives for presenting CVD mortality risk reduction: a discrete choice study of framing risk and health behaviour change. *Social Science & Medicine, 211*, 164–174. https://doi.org/10.1016/j.socscimed.2018.06.011

Gu, N. Y., Botteman, M. F., Gerber, R. A., Ji, X., Postema, R., Wan, Y., Sianos, G., Anthony, I., Cappelleri, J. C., Szczypa, P., & van Hout, B. (2013). Eliciting health state utilities for Dupuytren's contracture using a discrete choice experiment. *Acta Orthopaedica, 84*(6), 571–578. https://doi.org/10.3109/17453674.2013.865097

Gu, Y., Norman, R., & Viney, R. (2014). Estimating health state utility values from discrete choice experiments - a QALY space model approach. *Health Economics, 23*(9), 1098–1114. https://doi.org/10.1002/hec.3066

Hall, J., Viney, R., Haas, M., & Louviere, J. (2004). Using stated preference discrete choice modeling to evaluate health care programs. *Journal of Business Research, 57*(9), 1026–1032. https://doi.org/10.1016/S0148-2963(02)00352-1

Hampton, J. (2018). The between-subjects experiment. In Julia Nunn (Ed.), *Laboratory psychology* (pp. 15–37). Psychology Press.

Harris, P., Whitty, J. A., Kendall, E., Ratcliffe, J., Wilson, A., Littlejohns, P., & Scuffham, P. A. (2015). The Australian public's preferences for emergency care alternatives and the influence of the presenting context: a discrete choice experiment. *BMJ Open, 5*(4), e006820. https://doi.org/10.1136/bmjopen-2014-006820

Harrison, M., Milbers, K., Hudson, M., & Bansback, N. (2017). Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. *BMJ Open, 7*(5), e014719. https://doi.org/10.1136/bmjopen-2016-014719

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., & Payne, K. (2014). Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient, 7*(2), 151–170. https://doi.org/10.1007/s40271-014-0048-1

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals of Medicine, 33*(5), 358–370. https://doi.org/10.3109/07853890109002090

Heinzle, S. L. (2012). Disclosure of energy operating cost information: a silver bullet for overcoming the energy-efficiency gap? *Journal of Consumer Policy, 35*(1), 43–64. https://doi.org/10.1007/s10603-012-9189-6

Helter, T. M., & Boehler, C. E. H. (2016). Developing attributes for discrete choice experiments in health: a systematic literature review and case study of alcohol misuse interventions. *Journal of Substance Use*, *21*(6), 662–668.

Hess, S., Hensher, D. A., & Daly, A. (2012). Not bored yet – revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice, 46*(3), 626–644. https://doi.org/10.1016/j.tra.2011.11.008

Hollin, I. L., Craig, B. M., Coast, J., Beusterien, K., Vass, C., DiSantostefano, R., & Peay, H. (2020). Reporting Formative Qualitative Research to Support the Development of Quantitative Preference Study Protocols and Corresponding Survey Instruments: Guidelines for Authors and Reviewers. The Patient - Patient-Centered Outcomes Research, 13(1), 121-136. https://doi.org/10.1007/s40271-019-00401-x

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics, 69*(8), 1595–1603. https://doi.org/10.1016/j.ecolecon.2010.04.011

Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*(3), 307–317. https://doi.org/10.1177/002224379603300305

Hyland, M., Frei, C., Frei, A., & Mahmassani, H. S. (2018). Riders on the storm: exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behaviour and Society, 13*, 44–60. https://doi.org/10.1016/j.tbs.2018.05.001

InFocus. (2019). *InFocus Study*. Retrieved 22 April from https://www.infocusstudy.org.au/

Jahn, S., Schuch-Haellmigk, J., Dannewald, T., & Boztug, Y. (2018). How category average reference points affect choice of sugary foods. *Appetite, 126*, 201–209. https://doi.org/10.1016/j.appet.2018.04.001

Janssen, E. M., Longo, D. R., Bardsley, J. K., & Bridges, J. F. P. (2017). Education and patient preferences for treating type 2 diabetes: a stratified discrete-choice experiment. *Patient Preference and Adherence, 11*, 1729–1736. https://doi.org/10.2147/PPA.S139471

Jarvis, S. L. (2012). *Stated preference methods and models: analyzing recreational angling in New England groundfisheries*. Unpublished PhD thesis, University of Maryland. https://drum.lib.umd.edu/handle/1903/11556

Jensen, C. E., Sørensen, S. S., Gudex, C., Jensen, M. B., Pedersen, K. M., & Ehlers, L. H. (2021). The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. *Applied Health Economics and Health Policy 19*, 579–591. https://doi.org/10.1007/s40258-021-00639-3

Jin, J., Wang, W., He, R., & Gong, H. (2017). Valuing health risk in agriculture: a choice experiment approach to pesticide use in China. *Environmental Science & Pollution Research, 24*(21), 17526–17533. https://doi.org/10.1007/s11356-017-9418-2

Johnson, F. R., Yang, J.-C., & Reed, S. D. (2019). The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. Value in Health, 22(2), 157-160. https://doi.org/https://doi.org/10.1016/j.jval.2018.07.876

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., & Scarpa, R. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319–405.

Johnston, R. J., Holland, B. M., & Liuyang, Y. (2016). Individualized geocoding in stated preference questionnaires: implications for survey design and welfare estimation. *Land Economics, 92*(4), 737–759. https://doi.org/10.3368/le.92.4.737

Jonker, M. F., Attema, A. E., Donkers, B., Stolk, E. A., & Versteegh, M. M. (2017). Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Economics, 26*(12), 1534–1547. https://doi.org/10.1002/hec.3445

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018a). Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value in Health, 21*(8), 993-1001. https://doi.org/10.1016/j.jval.2018.01.016

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018b). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in Health, 21*(7), 767–771. https://doi.org/10.1016/j.jval.2017.10.002

Jonker, M. F., Stolk, E. A., & Donkers, B. (2012). Valuing EQ-5D-5L using DCE duration: Too complex or not.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185

Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., Busschbach, J., & Boye, K. S. (2020). Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *The European Journal of Health Economics*, *21*, 1245–1257.

Kim, D., & Park, B.-J. (2017). The moderating role of context in the effects of choice attributes on hotel choice: a discrete choice experiment. *Tourism Management, 63*, 439–451. https://doi.org/10.1016/j.tourman.2017.07.014

Klaiman, K., Ortega, D. L., & Garnache, C. (2016). Consumer preferences and demand for packaging material and recyclability. *Resources, Conservation and Recycling, 115*, 1–8. https://doi.org/10.1016/j.resconrec.2016.08.021

Klarman, H. E., & Rosenthal, G. D. (1968). Cost effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care*, *6*(1), 48–54.

Knox, S. A., Viney, R. C., Gu, Y. Y., Hole, A. R., Fiebig, D. G., Street, D. J., Haas, M. R., Weisberg, E., & Bateson, D. (2013). The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products. *Social Science & Medicine, 83*, 70–80. https://doi.org/10.1016/j.socscimed.2012.12.025

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing, 33*(1), 11–26. https://doi.org/10.1016/j.ijresmar.2014.12.004
Kroes, E. P., & Sheldon, R. J. (1988). Stated preference methods: an introduction. *Journal of Transport Economics and Policy*, 11–25.
Kuhfeld, W. (2006). *Orthogonal arrays*. http://support.sas.com/techsup/technote/ts723_Designs.txt
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157. http://www.jstor.org/stable/1828835
Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics, 26*(8), 661–677. https://doi.org/10.2165/00019053-200826080-00004
Lanz, B., & Provins, A. (2015). Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics, 47*(3), 300–324. https://doi.org/10.1007/s11149-015-9272-4
LeBoeuf, R. A., & Shafir, E. (2009). Anchoring on the 'here' and 'now' in time and distance judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 81–93. https://doi.org/10.1037/a0013665
Liebe, U., Meyerhoff, J., & Hartje, V. (2012). Test–retest reliability of choice experiments in environmental valuation. *Environmental and Resource Economics*, *53*(3), 389–407.
Lim, S., Jonker, M. F., Oppe, M., Donkers, B., & Stolk, E. (2018). Severity-stratified discrete choice experiment designs for health state evaluations. *Pharmacoeconomics, 36*(11), 1377–1389. https://doi.org/10.1007/s40273-018-0694-6
Little, P., Everitt, H., Williamson, I., Warner, G., Moore, M., Gould, C., Ferrier, K., & Payne, S. (2001). Preferences of patients for patient centred approach to consultation in primary care: observational study. *BMJ*, *322*(7284), 468.
Lockshin, L., Corsi, A. M., Cohen, J., Lee, R., & Williamson, P. (2017). West versus east: measuring the development of Chinese wine preferences. *Food Quality and Preference, 56*, 256–265. https://doi.org/10.1016/j.foodqual.2016.02.014
Loewenstein, G., & Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics, 92*(8), 1795–1810. https://doi.org/10.1016/j.jpubeco.2007.12.011
Logar, I., & Brouwer, R. (2018). Substitution effects and spatial preference heterogeneity in single- and multiple-site choice experiments. *Land Economics, 94*(2), 302–322. https://doi.org/10.3368/le.94.2.302
Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, *3*(3), 57–72.
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: theory, methods and applications*. Cambridge University Press.
Ludwig, K., Ramos-Goñi, J. M., Oppe, M., Kreimeier, S., & Greiner, W. (2021). To what extent do patient preferences differ from general population preferences? *Value in Health*, *24*(9), 1343–1349. https://doi.org/10.1016/j.jval.2021.02.012
Mann, R., Brazier, J., & Tsuchiya, A. (2009). A comparison of patient and general population weightings of EQ-5D dimensions. *Health Economics, 18*(3), 363–372. https://doi.org/10.1002/hec.1362
Mark, T. L., & Swait, J. (2004). Using stated preference and revealed preference modeling to evaluate prescribing decisions. *Health Economics*, *13*(6), 563–573.
Marshall, D. A., Deal, K., Bombard, Y., Leighl, N., MacDonald, K. V., & Trudeau, M. (2016). How do women trade-off benefits and risks in chemotherapy treatment decisions based on gene expression profiling for early-stage breast cancer? A discrete choice experiment. *BMJ Open*, *6*(6), e010981. https://doi.org/10.1136/bmjopen-2015-010981
Marshall, H. S., Chen, G., Clarke, M., & Ratcliffe, J. (2016). Adolescent, parent and societal preferences and willingness to pay for meningococcal B vaccine: a discrete choice experiment. *Vaccine, 34*(5), 671–677. https://doi.org/10.1016/j.vaccine.2015.11.075

McCrary, J. M., Goldstein, D., Boyle, F., Cox, K., Grimison, P., Kiernan, M. C., Krishnan, A. V., Lewis, C. R., Webber, K., Baron-Hay, S., Horvath, L., & Park, S. B. (2017). Optimal clinical assessment strategies for chemotherapy-induced peripheral neuropathy (CIPN): a systematic review and Delphi survey. *Supportive Care in Cancer, 25*(11), 3485–3493. https://doi.org/10.1007/s00520-017-3772-y

McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351–378.

Meenakshi, J. V., Banerji, A., Manyong, V., Tomlins, K., Mittal, N., & Hamukwala, P. (2012). Using a discrete choice experiment to elicit the demand for a nutritious food: willingness-to-pay for orange maize in rural Zambia. *Journal of Health Economics, 31*(1), 62–71. https://doi.org/10.1016/j.jhealeco.2012.01.002

Mendelsohn, R. (2019). An examination of recent revealed preference valuation methods and results. *Review of Environmental Economics & Policy, 13*(2), 267–282. https://doi.org/10.1093/reep/rez003

Meyerhoff, J., & Glenk, K. (2015). Learning how to choose – effects of instructional choice sets in discrete choice experiments. *Resource and Energy Economics, 41*, 122–142. https://doi.org/http://www.sciencedirect.com/science/journal/09287655

Miyamoto, J. M., Wakker, P. P., Bleichrodt, H., & Peters, H. J. M. (1998). The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. *Management Science, 44*(6), 839–849. https://doi.org/10.1287/mnsc.44.6.839

Morgenstern, O., & Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton University Press.

Mühlbacher, A. C., & Juhnke, C. (2013). Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Applied Health Economics and Health Policy, 11*(3), 163–180. https://doi.org/10.1007/s40258-013-0023-3

Mühlbacher, A. C., Bethge, S., Reed, S. D., & Schulman, K. A. (2016). Patient preferences for features of health care delivery systems: a discrete choice experiment. *Health Services Research, 51*(2), 704–727. https://doi.org/10.1111/1475-6773.12345

Mulhern, B. J., Bansback, N., Norman, R., & Brazier, J. (2020). Valuing the SF-6Dv2 Classification System in the United Kingdom Using a Discrete-choice Experiment With Duration. Medical Care, 58(6), 566-573. https://doi.org/10.1097/MLR.0000000000001324

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health, 16*(1), 104–113. https://doi.org/10.1016/j.jval.2012.09.001

Mulhern, B., Norman, R., Shah, K., Bansback, N., Longworth, L., & Viney, R. (2018). How should discrete choice experiments with duration choice sets be presented for the valuation of health states? *Medical Decision Making, 38*(3), 306–318. https://doi.org/10.1177/0272989x17738754

Mulhern, B., Norman, R., Street, D. J., & Viney, R. (2018). One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *Pharmacoeconomics, 37*, 29–43. https://doi.org/10.1007/s40273-018-0714-6

Munger, M. A., Walsh, M., Godin, J., & Feehan, M. (2017). Pharmacist's demand for optimal primary care service delivery in a community pharmacy: the OPTiPharm study. *Annals of Pharmacotherapy, 51*(12), 1069–1076. https://doi.org/10.1177/1060028017722795

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Najafzadeh, M., Schneeweiss, S., Choudhry, N. K., Avorn, J., & Gagne, J. J. (2019). General population vs. patient preferences in anticoagulant therapy: a discrete choice experiment. *The Patient – Patient-Centered Outcomes Research, 12*(2), 235–246. https://doi.org/10.1007/s40271-018-0329-1

NCSS Statistical Software. (2021). *Sample size for chi-square and other proportions tests in PASS*. https://www.ncss.com/software/pass/crosstabs-and-other-proportions-tests-in-pass/#CrossTabs
Ni, F., Arnott, D., & Gao, S. (2019). The anchoring effect in business intelligence supported decision-making. *Journal of Decision Systems, 28*(2), 67–81. https://doi.org/10.1080/12460125.2019.1620573
NICE, U. (2013). *Guide to the methods of technology appraisal.* National Institute for Health and Clinical Excellence (NICE).
Nickel, B., Howard, K., Brito, J. P., Barratt, A., Moynihan, R., & McCaffery, K. (2018). Association of preferences for papillary thyroid cancer treatment with disease terminology: a discrete choice experiment. *JAMA Otolaryngology – Head & Neck Surgery, 144*(10), 887–896. https://doi.org/10.1001/jamaoto.2018.1694
Norman, R., Cronin, P., & Viney, R. (2013). A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Applied Health Economics and Health Policy, 11*(3), 287–298. https://doi.org/10.1007/s40258-013-0035-z
Norman, R., Mulhern, B., & Viney, R. (2016). The impact of different DCE-based approaches when anchoring utility scores. *Pharmacoeconomics, 34*(8), 805–814. https://doi.org/10.1007/s40273-016-0399-7
Norman, R., Viney, R., Brazier, J., Burgess, L., Cronin, P., King, M., Ratcliffe, J., & Street, D. (2013). Valuing SF-6D health states using a discrete choice experiment. *Medical Decision Making, 34*(6), 773–786. https://doi.org/10.1177/0272989X13503499
Oehlmann, M., Meyerhoff, J., Mariel, P., & Weller, P. (2017). Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics & Management, 81*, 59–73. https://doi.org/10.1016/j.jeem.2016.09.002
Ogorevc, M., Murovec, N., Fernandez, N. B., & Rupel, V. P. (2019). Questioning the differences between general public vs. patient based preferences towards EQ-5D-5L defined hypothetical health states. *Health Policy, 123*(2), 166–172. https://doi.org/10.1016/j.healthpol.2017.03.011
Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J. M., & Luo, N. (2016). EuroQol Protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics, 34*(10), 993–1004. https://doi.org/10.1007/s40273-016-0404-1
Oppewal, H., Huybers, T., & Crouch, G. I. (2015). Tourist destination and experience choice: a choice experimental analysis of decision sequence effects. *Tourism Management, 48*, 467–476. https://doi.org/10.1016/j.tourman.2014.12.016
Parker, J. R., & Schrift, R. Y. (2011). Rejectable choice sets: how seemingly irrelevant no-choice options affect consumer decision processes. *Journal of Marketing Research, 48*(5), 840–854. https://doi.org/10.1509/jmkr.48.5.840
Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J., & Yazdizadeh, A. (2017). Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning, 157*, 63–74. https://doi.org/10.1016/j.landurbplan.2016.05.024
Peeters, Y., & Stiggelbout, A. M. (2010). Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value in Health*, *13*(2), 306–309.
Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research*, *28*(1), 206–224.
Purba, F. D., Hunfeld, J. A., Iskandarsyah, A., Fitriana, T. S., Sadarjoen, S. S., Passchier, J., & Busschbach, J. J. (2017). Employing quality control and feedback to the EQ-5D-5L valuation protocol to improve the quality of data collection. *Quality of Life Research*, *26*(5), 1197–1208.
R Development Core Team. (2020). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing.
Rakotonarivo, O. S., Schaafsma, M., & Hockley, N. (2016). A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods.

Journal of Environmental Management, 183, 98–109. https://doi.org/10.1016/j.jenvman.2016.08.032
Ramos-Goñi, J. M., Oppe, M., Slaap, B., Busschbach, J. J. V., & Stolk, E. (2017). Quality control process for EQ-5D-5L valuation studies. *Value In Health, 20*(3), 466–473. https://doi.org/10.1016/j.jval.2016.10.012
Rand-Hendriksen, K., Augestad, L. A., Kristiansen, I. S., & Stavem, K. (2012). Comparison of hypothetical and experienced EQ-5D valuations: relative weights of the five dimensions. *Quality of Life Research, 21*(6), 1005–1012. https://doi.org/10.1007/s11136-011-0016-3
Reed Johnson, F., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value In Health, 16*(1), 3–13. https://doi.org/10.1016/j.jval.2012.08.2223
Rencz, F., Brodszky, V., Gulácsi, L., Golicki, D., Ruzsa, G., Pickard, A. S., Law, E. H., & Péntek, M. (2020). Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value in Health*, *23*(9), 1235–1245.
Resano-Ezcaray, H., Sanjuán-López, A. I., & Albisu-Aguado, L. M. (2010). Combining stated and revealed preferences on typical food products: the case of dry-cured ham in Spain. *Journal of Agricultural Economics, 61*(3), 480–498. https://doi.org/10.1111/j.1477-9552.2010.00250.x
Rid, W., Haider, W., Ryffel, A., & Beardmore, B. (2018). Visualisations in choice experiments: comparing 3D film-sequences and still-images to analyse housing development alternatives. *Ecological Economics, 146*, 203–217. https://doi.org/10.1016/j.ecolecon.2017.10.019
Robinson, J., Clouston, B., Suh, J., & Chaloupka, M. (2008). Are citizens' juries a useful tool for assessing environmental value? *Environmental Conservation*, *35*(4), 351–360.
Rolfe, J., & Bennett, J. (2009). The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics, 68*(4), 1140–1148. https://doi.org/10.1016/j.ecolecon.2008.08.007
RStudio Team. (2020). *RStudio: integrated development environment for R. RStudio*. http://www.rstudio.com/
Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R., Khorrami, P. R., Kveder, J., Andersen, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., Pérez, J., Quail, S. K., … Folke, T. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour, 4*(6), 622–633. https://doi.org/10.1038/s41562-020-0886-x
Ryan, M., Scott, D. A., Reeves, C., Bate, A., van Teijlingen, E. R., Russell, E. M., Napper, M., & Robb, C. M. (2001). Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment*, *5*(5), 1–186.
Salloum, R. G., Abbyad, C. W., Kohler, R. E., Kratka, A. K., Oh, L., & Wood, K. A. (2015). Assessing preferences for a university-based smoking cessation program in Lebanon: a discrete choice experiment. *Nicotine & Tobacco Research*, *17*(5), 580–585. https://doi.org/10.1093/ntr/ntu188
Sarrias, M., & Daziano, R. (2017). Multinomial logit models with continuous and discrete individual heterogeneity in R: the gmnl package. *Journal of Statistical Software*, *79*(2), 46. https://doi.org/10.18637/jss.v079.i02
Schaafsma, M., Brouwer, R., Liekens, I., & De Nocker, L. (2014). Temporal stability of preferences and willingness to pay for natural areas in choice experiments: a test–retest. *Resource and Energy Economics, 38*, 243–260. https://doi.org/10.1016/j.reseneeco.2014.09.001
Scheufele, G., & Bennett, J. (2013). Effects of alternative elicitation formats in discrete choice experiments. *Australian Journal of Agricultural & Resource Economics, 57*(2), 214–233. https://doi.org/10.1111/j.1467-8489.2012.00612.x

Services Australia. (2021). *About Medicare*. https://www.servicesaustralia.gov.au/individuals/subjects/whats-covered-medicare/about-medicare

Shafrin, J., May, S. G., Zhao, L. M., Bognar, K., Yuan, Y., Penrod, J. R., & Romley, J. A. (2021). Measuring the value healthy individuals place on generous insurance coverage of severe diseases: a stated preference survey of adults diagnosed with and without lung cancer. *Value in Health, 24*(6), 855–861. https://doi.org/10.1016/j.jval.2020.06.021

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics, 41*(22), 2915–2924. https://doi.org/10.1080/00036840801964633

Sicsic, J., Krucien, N., & Franc, C. (2016). What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. *Social Science & Medicine, 167*, 116–127. https://doi.org/10.1016/j.socscimed.2016.08.050

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics, 37*(2), 201–226. https://doi.org/10.1007/s40273-018-0734-2

Spinks, J., & Mortimer, D. (2015). The effect of traffic lights and regulatory statements on the choice between complementary and conventional medicines in Australia: results from a discrete choice experiment. *Social Science & Medicine, 124*, 257–265. https://doi.org/10.1016/j.socscimed.2014.11.049

Stacey, D., Légaré, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., Holmes-Rovner, M., Llewellyn-Thomas, H., Lyddiatt, A., & Thomson, R. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews, 4*(4), CD001431. https://doi.org/10.1002/14651858.CD001431.pub5

Staff, N. P., Grisold, A., Grisold, W., & Windebank, A. J. (2017). Chemotherapy-induced peripheral neuropathy: a current review. *Annals of Neurology, 81*(6), 772–781. https://doi.org/10.1002/ana.24951

StataCorp. (2017). *Stata statistical software: release 15*.

Stolk, E., Ludwig, K., Rand, K., van Hout, B., & Ramos-Goñi, J. M. (2019). Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value in Health, 22*(1), 23–30. https://doi.org/10.1016/j.jval.2018.05.010

Street, D. J. (2019). *MNL simulation code* [Programming Code]. Centre for Health Economics Research and Evaluation (CHERE).

Street, D. J. (2020a). *Power calculations in Mathematica.* [Programming Code]. Centre for Health Economics Research and Evaluation (CHERE).

Street, D. J. (2020b). *Calculating utility of assessment tools in Mathematica* [Programming Code]. Centre for Health Economics Research and Evaluation (CHERE).

Street, D. J., & Burgess, L. (2007). *The construction of optimal stated choice experiments: theory and methods* (Vol. 647). John Wiley & Sons.

Street, D. J., Mulhern, B., Norman, R., & Viney, R. (2017). *Using simulations to compare DCE designs that could be used to value EQ-5D.* EuroQol Plenary meeting, Barcelona.

SurveyEngine GmbH. (2021). *SurveyEngine*. https://surveyengine.com/

Swedish Health Services. (2018). Peripheral neuropathy. YouTube, 17 August. https://www.youtube.com/watch?v=Zki82j3SmMo

Tirachini, A., Sun, L., Erath, A., & Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. *Transport Policy, 47*, 94–104. https://doi.org/10.1016/j.tranpol.2015.12.004

Torquati, B., Tempesta, T., Vecchiato, D., & Venanzi, S. (2018). Tasty or sustainable? The effect of product sensory experience on a sustainable new food product: an application of discrete choice experiments on Chianina tinned beef. *Sustainability, 10*(8), Article 2795. https://doi.org/10.3390/su10082795

Torrance, G. W. (1986). Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics, 5*(1), 1–30. https://doi.org/10.1016/0167-6296(86)90020-2
Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases, 40*(6), 593–600. https://doi.org/10.1016/0021-9681(87)90019-1
Torrance, G. W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions. *Pharmacoeconomics, 7*(6), 503–520. https://doi.org/10.2165/00019053-199507060-00005
Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research*, *7*(2), 118.
Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124–1131.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.
Vacca, A., Prato, C. G., & Meloni, I. (2019). Should I stay or should I go? Investigating route switching behavior from revealed preferences data. *Transportation, 46*(1), 75–93. https://doi.org/10.1007/s11116-017-9788-9
van Exel, N. J. A., Brouwer, W. B. F., van den Berg, B., & Koopmanschap, M. A. (2006). With a little help from an anchor: discussion and evidence of anchoring effects in contingent valuation. *The Journal of Socio-Economics*, *35*(5), 836–853. https://doi.org/10.1016/j.socec.2005.11.045
Vass, C. M., Rigby, D., & Payne, K. (2018). Investigating the heterogeneity in women's preferences for breast screening: does the communication of risk matter? *Value in Health, 21*(2), 219–228. https://doi.org/10.1016/j.jval.2017.07.010
Vass, C. M., Wright, S., Burton, M., & Payne, K. (2018). Scale heterogeneity in healthcare discrete choice experiments: a primer. *The Patient*, *11*(2), 167. https://doi.org/http://dx.doi.org/10.1007/s40271-017-0282-4
Veldwijk, J., Essers, B. A. B., Lambooij, M. S., Dirksen, C. D., Smit, H. A., & de Wit, G. A. (2016). Survival or mortality: does risk attribute framing influence decision-making behavior in a discrete choice experiment? *Value in Health, 19*(2), 202–209. https://doi.org/10.1016/j.jval.2015.11.004
Viney, R., Norman, R., Brazier, J., Cronin, P., King, M. T., Ratcliffe, J., & Street, D. (2014). An Australian discrete choice experiment to value EQ-5D health states. *Health Economics, 23*(6), 729–742. https://doi.org/10.1002/hec.2953
Viney, R., Norman, R., King, M. T., Cronin, P., Street, D. J., Knox, S., & Ratcliffe, J. (2011). Time trade-off derived EQ-5D weights for Australia. *Value in Health, 14*(6), 928–936. https://doi.org/10.1016/j.jval.2011.04.009
Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science, 42*(8), 1131. http://dx.doi.org/10.1287/mnsc.42.8.1131
Ware Jr, J. E. (2000). SF-36 health survey update. *Spine*, *25*(24), 3130–3139.
Waudby-Smith, I., Pickard, A. S., Xie, F., & Pullenayegum, E. M. (2020). Using both time tradeoff and discrete choice experiments in valuing the EQ-5D: impact of model misspecification on value sets. *Medical Decision Making, 40*(4), 483–497. https://doi.org/10.1177/0272989X20924019
Webb, E. J. D., Meads, D., Lynch, Y., Judge, S., Randall, N., Goldbart, J., Meredith, S., Moulam, L., Hess, S., & Murray, J. (2021). Attribute selection for a discrete choice experiment incorporating a best-worst scaling survey. *Value in Health, 24*(4), 575–584. https://doi.org/10.1016/j.jval.2020.10.025
Welie, A. G., Gebretekle, G. B., Stolk, E., Mukuria, C., Krahn, M. D., Enquoselassie, F., & Fenta, T. G. (2020). Valuing health state: an EQ-5D-5L value set for Ethiopians. *Value in Health Regional Issues*, *22*, 7–14.
Whitty, J. A., Kendall, E., Sav, A., Kelly, F., McMillan, S. S., King, M. A., & Wheeler, A. J. (2015). Preferences for the delivery of community pharmacy services to help manage chronic

conditions. *Research in Social & Administrative Pharmacy*, *11*(2), 197–215. https://doi.org/10.1016/j.sapharm.2014.06.007
Whitty, J. A., Ratcliffe, J., Kendall, E., Burton, P., Wilson, A., Littlejohns, P., Harris, P., Krinks, R., & Scuffham, P. A. (2015). Prioritising patients for bariatric surgery: building public preferences from a discrete choice experiment into public policy. *BMJ Open, 5*(10), e008919. https://doi.org/10.1136/bmjopen-2015-008919
Williamson, P. O., Lockshin, L., Francis, I. L., & Mueller Loose, S. (2016). Influencing consumer choice: short and medium term effect of country of origin information on wine choice. *Food Quality and Preference, 51*, 89–99. https://doi.org/10.1016/j.foodqual.2016.02.018
Willis, G. B. (1999). *Cognitive interviewing: a 'how to' guide*. Research Triangle Park, NC: Research Triangle Institute.
Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*(4), 387.
Wisløff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., & Olsen, J. A. (2014). Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics, 32*(4), 367–375. https://doi.org/10.1007/s40273-014-0136-z
Wolfram Research, I. (2016). *Mathematica*. Wolfram Research, Inc.
Wright, S. J., Ulph, F., Dharni, N., & Payne, K. (2017). Eliciting preferences for information provision in newborn bloodspot screening programs. *Value in Health*, *20*(4), 651–661. https://doi.org/10.1016/j.jval.2016.11.007
Yang, F., Devlin, N., & Luo, N. (2019). Cost-utility analysis using EQ-5D-5L data: does how the utilities are derived matter? *Value in Health, 22*(1), 45–49. https://doi.org/10.1016/j.jval.2018.05.008
Youngkong, S., Baltussen, R., Tantivess, S., Koolman, X., & Teerawattananon, Y. (2010). Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC Health Services Research, 10*, 197. https://doi.org/10.1186/1472-6963-10-197
Zhang, J., & Adamowicz, W. L. (2011). Unraveling the Choice format effect: a context-dependent random utility model. *Land Economics, 87*(4), 730–743. https://doi.org/http://le.uwpress.org/content/by/year
Zhao, M. J., Johnston, R. J., & Schultz, E. T. (2013). What to value and how? Ecological indicator choices in stated preference valuation. *Environmental & Resource Economics, 56*(1), 3–25. https://doi.org/10.1007/s10640-013-9636-0

Appendices

Appendix 3A: Studies included in Scoping Review

Alexander, D. L., Sailors, J. J., Al-Khatib, J. A., & Al-Habib, M. I. (2012). The effect of near- versus distant-future mindsets on socially responsible investors' mutual fund preferences. *Journal of Financial Services Marketing*, *17*(1), 67–79. https://doi.org/10.1057/fsm.2012.5

Amano, I., Kurisu, K., & Hanaki, K. (2016). Evaluation of consumers' preferences on drinking waters considering the information provision effect. *Water Science and Technology: Water Supply*, *16*(4), 1057–1067. https://doi.org/10.2166/ws.2016.019

Antoniou, C., & Kostovasilis, K. (2017). How may external information affect traffic risk perception? *Journal of Transportation Safety and Security*, *9*(3), 347–368. https://doi.org/10.1080/19439962.2016.1212444

Arentze, T., Feng, T., Timmermans, H., & Robroeks, J. (2012). Context-dependent influence of road attributes and pricing policies on route choice behavior of truck drivers: results of a conjoint choice experiment. *Transportation*, *39*(6), 1173–1188. https://doi.org/10.1007/s11116-012-9391-z

Arrua, A., Curutchet, M. R., Rey, N., Barreto, P., Golovchenko, N., Sellanes, A., Velazco, G., Winokur, M., Gimenez, A., & Ares, G. (2017). Impact of front-of-pack nutrition information and label design on children's choice of two snack foods: comparison of warnings and the traffic-light system. *Appetite*, *116*, 139–146. https://doi.org/10.1016/j.appet.2017.04.012

Baba, Y., Kallas, Z., Costa-Font, M., Gil, J. M., & Realini, C. E. (2016). Impact of hedonic evaluation on consumers' preferences for beef attributes including its enrichment with n-3 and CLA fatty acids. *Meat Science*, *111*, 9–17. https://doi.org/10.1016/j.meatsci.2015.08.005

Bansback, N., Harrison, M., & Marra, C. (2016). Does introducing imprecision around probabilities for benefit and harm influence the way people value treatments? *Medical Decision Making*, *36*(4), 490–502. https://doi.org/10.1177/0272989x15600708

Bech, M., Kjaer, T., & 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. https://doi.org/10.1002/hec.1587

Benning, T. M., Dellaert, B. G., Severens, J. L., & Dirksen, C. D. (2014). The effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: results from a discrete choice experiment. *Value in Health*, *17*(5), 578–587. https://doi.org/10.1016/j.jval.2014.04.007

Blake, M. R., Lancsar, E., Peeters, A., & Backholer, K. (2018). The effect of sugar-sweetened beverage price increases and educational messages on beverage purchasing behavior among adults. *Appetite*, *126*, 156–162. https://doi.org/10.1016/j.appet.2018.03.012

Burton, M., & Rigby, D. (2012). The self selection of complexity in choice experiments. *American Journal of Agricultural Economics*, *94*(3), 786–800. https://academic.oup.com/ajae/issue

Caputo, V., Lusk, J. L., & Jr.Nayga, R. M. (2018). Choice experiments are not conducted in a vacuum: the effects of external price information on choice behavior. *Journal of Economic Behavior & Organization*, *145*, 335–351. https://doi.org/10.1016/j.jebo.2017.11.018

Cole, A., Shah, K., Mulhern, B., Feng, Y., & Devlin, N. (2018). Valuing EQ-5D-5L health states 'in context' using a discrete choice experiment. *European Journal of Health Economics*, *19*(4), 595–605. https://doi.org/10.1007/s10198-017-0905-7

Dellaert, B. G. C., Donkers, B., & van Soest, A. (2012). Complexity effects in choice experiment-based bodels. *Journal of Marketing Research*, *49*(3), 424–434. https://doi.org/10.1509/jmr.09.0315

De-Magistris, T., Gracia, A., & Nayga Jr, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, *95*(5), 1136–1154. https://doi.org/10.1093/ajae/aat052

Devarasetty, P. C., Burris, M., & Douglass Shaw, W. (2012). The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation Research*, *46*(8), 1227–1240. https://doi.org/10.1016/j.tra.2012.05.002

Emberger-Klein, A., & Menrad, K. (2018). The effect of information provision on supermarket consumers' use of and preferences for carbon labels in Germany. *Journal of Cleaner Production*, *172*, 253–263. https://doi.org/10.1016/j.jclepro.2017.10.105

Gajic, A., Cameron, D., & Hurley, J. (2012). The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. *European Journal of Health Economics*, *13*(6), 789–799. https://doi.org/10.1007/s10198-011-0332-0

Grisolía, J. M., Longo, A., Hutchinson, G., & Kee, F. (2018). Comparing mortality risk reduction, life expectancy gains, and probability of achieving full life span, as alternatives for presenting CVD mortality risk reduction: a discrete choice study of framing risk and health behaviour change. *Social Science & Medicine*, *211*, 164–174. https://doi.org/10.1016/j.socscimed.2018.06.011

Harris, P., Whitty, J. A., Kendall, E., Ratcliffe, J., Wilson, A., Littlejohns, P., & Scuffham, P. A. (2015). The Australian public's preferences for emergency care alternatives and the influence of the

presenting context: a discrete choice experiment. *BMJ Open*, *5*(4), e006820. https://doi.org/10.1136/bmjopen-2014-006820

Heinzle, S. (2012). Disclosure of energy operating cost information: a silver bullet for overcoming the energy-efficiency gap? *Journal of Consumer Policy*, *35*(1), 43–64. https://doi.org/10.1007/s10603-012-9189-6

Heinzle, S. L., & Wustenhagen, R. (2012). Dynamic adjustment of eco-labeling schemes and consumer choice – the revision of the EU energy label as a missed opportunity? *Business Strategy and the Environment*, *21*(1), 60–70. https://doi.org/10.1002/bse.722

Hyland, M., Frei, C., Frei, A., & Mahmassani, H. S. (2018). Riders on the storm: exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behaviour and Society*, *13*, 44–60. https://doi.org/10.1016/j.tbs.2018.05.001

Jahn, S., Schuch-Haellmigk, J., Dannewald, T., & Boztug, Y. (2018). How category average reference points affect choice of sugary foods. *Appetite*, *126*, 201–209. https://doi.org/10.1016/j.appet.2018.04.001

Jarvis, S. L. (2012). *Stated preference methods and models: analyzing recreational angling in New England groundfisheries.* ProQuest Information & Learning]. psyh. https://www.lib.uts.edu.au/goto?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-99050-590&site=ehost-live

Jiang, C. C., & Fraenkel, L. (2017). The influence of varying cost formats on preferences. *Medical Decision Making*, *37*(1), 17–26. https://doi.org/10.1177/0272989x16677031

Jin, J., Wang, W., He, R., & Gong, H. (2017). Valuing health risk in agriculture: a choice experiment approach to pesticide use in China. *Environmental Science & Pollution Research*, *24*(21), 17526–17533. https://doi.org/10.1007/s11356-017-9418-2

Johnston, R. J., Holland, B. M., & Liuyang, Y. (2016). Individualized geocoding in stated preference questionnaires: implications for survey design and welfare estimation. *Land Economics*, *92*(4), 737–759. https://www.lib.uts.edu.au/goto?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=118463775&site=ehost-live

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in Health*, *21*(7), 767–771. https://doi.org/10.1016/j.jval.2017.10.002

Kenny, P., Goodall, S., Street, D. J., & Greene, J. (2017). Choosing a doctor: does presentation format affect the way consumers use health care performance information? *Patient*, *10*(6), 739–751. https://doi.org/10.1007/s40271-017-0245-9

Kerkman, K., Arentze, T., Borgers, A., & Kemperman, A. (2012). Car drivers' compliance with route advice and willingness to choose socially desirable routes. *Transportation Research Record, 2322*, 102–109.

Kim, D., & Park, B. J. R. (2017). The moderating role of context in the effects of choice attributes on hotel choice: a discrete choice experiment. *Tourism Management*, *63*, 439–451. https://doi.org/10.1016/j.tourman.2017.07.014

Klaiman, K., Ortega, D. L., & Garnache, C. (2016). Consumer preferences and demand for packaging material and recyclability. *Resources, Conservation and Recycling*, *115*, 1–8. https://doi.org/10.1016/j.resconrec.2016.08.021

Knox, S. A., Viney, R. C., Gu, Y., Hole, A. R., Fiebig, D. G., Street, D. J., Haas, M. R., Weisberg, E., & Bateson, D. (2013). The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products. *Social Science & Medicine*, *83*, 70–80. https://doi.org/10.1016/j.socscimed.2012.12.025

Koç, H., & van Kippersluis, H. (2017). Thought for food: nutritional information and educational disparities in diet. *Journal of Human Capital*, *11*(4), 508–552. https://doi.org/10.1086/694571

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing*, *33*(1), 11–26. https://doi.org/10.1016/j.ijresmar.2014.12.004

Krovetz, H., Taylor, R., & Villas-Boas, S. (2017). *Willingness to pay for low water footprint food choices during drought*. NBER Working Paper 23495, National Bureau of Economic Research. http://www.nber.org/papers/w23495.pdf

Lanz, B., & Provins, A. (2015). Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics*, *47*(3), 300–324. https://doi.org/10.1007/s11149-015-9272-4

Logar, I., & Brouwer, R. (2017). The effect of risk communication on choice behavior, welfare estimates and choice certainty. *Water Resources and Economics*, *18*, 34–50. https://doi.org/10.1016/j.wre.2016.11.004

Logar, I., & Brouwer, R. (2018). Substitution effects and spatial preference heterogeneity in single- and multiple-site choice experiments. *Land Economics*, *94*(2), 302–322. https://doi.org/10.3368/le.94.2.302

Meenakshi, J. V., Banerji, A., Manyong, V., Tomlins, K., Mittal, N., & Hamukwala, P. (2012). Using a discrete choice experiment to elicit the demand for a nutritious food: willingness-to-pay for orange maize in rural Zambia. *Journal of Health Economics*, *31*(1), 62–71. https://doi.org/10.1016/j.jhealeco.2012.01.002

Meyerhoff, J., & Glenk, K. (2015). Learning how to choose – effects of instructional choice sets in discrete choice experiments. *Resource and Energy Economics*, *41*, 122–142. https://doi.org/http://www.sciencedirect.com/science/journal/09287655

Min, J., Azevedo, I. L., Michalek, J., & de Bruin, W. B. (2014). Labeling energy cost on light bulbs lowers implicit discount rates. *Ecological Economics*, *97*, 42–50. https://doi.org/10.1016/j.ecolecon.2013.10.015

Mitrani, A. (2016). Full-scale controlled comparison of alternative information formats for stated choice exercises. *Transportation Research Record*(2565), 78–85. https://doi.org/10.3141/2565-09

Mühlbacher, A. C., Bethge, S., Reed, S. D., & Schulman, K. A. (2016). Patient preferences for features of health care delivery systems: a discrete choice experiment. *Health Services Research*, *51*(2), 704–727. https://doi.org/10.1111/1475-6773.12345

Mulhern, B., Norman, R., Shah, K., Bansback, N., Longworth, L., & Viney, R. (2018). How should discrete choice experiments with duration choice sets be presented for the valuation of health states? *Medical Decision Making*, *38*(3), 306–318. https://doi.org/10.1177/0272989x17738754

Nickel, B., Howard, K., Brito, J. P., Barratt, A., Moynihan, R., & McCaffery, K. (2018). Association of preferences for papillary thyroid cancer treatment with disease terminology: a discrete choice experiment. *JAMA Otolaryngology: Head & Neck Surgery*, *144*(10), 887–896. https://doi.org/10.1001/jamaoto.2018.1694

Oehlmann, M., Meyerhoff, J., Mariel, P., & Weller, P. (2017). Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics & Management*, *81*, 59–73. https://doi.org/10.1016/j.jeem.2016.09.002

Oppewal, H., Huybers, T., & Crouch, G. I. (2015). Tourist destination and experience choice: a choice experimental analysis of decision sequence effects. *Tourism Management*, *48*, 467–476. https://doi.org/10.1016/j.tourman.2014.12.016

Parker, J. R., & Schrift, R. Y. (2011). Rejectable choice sets: how seemingly irrelevant no-choice options affect consumer decision processes. *Journal of Marketing Research (JMR)*, *48*(5), 840–854. https://doi.org/10.1509/jmkr.48.5.840

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J., & Yazdizadeh, A. (2017). Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning*, *157*, 63–74. https://doi.org/10.1016/j.landurbplan.2016.05.024

Rid, W., Haider, W., Ryffel, A., & Beardmore, B. (2018). Visualisations in choice experiments: comparing 3D film-sequences and still-images to analyse housing development alternatives. *Ecological Economics*, *146*, 203–217. https://doi.org/10.1016/j.ecolecon.2017.10.019

Salloum, R. G., Louviere, J. J., Getz, K. R., Islam, F., Anshari, D., Cho, Y., O'Connor, R. J., Hammond, D., & Thrasher, J. F. (2018). Evaluation of strategies to communicate harmful and potentially harmful constituent (HPHC) information through cigarette package inserts: a discrete choice experiment. *Tob Control*, *27*(6), 677–683. https://doi.org/10.1136/tobaccocontrol-2016-053579

Sandorf, E. D., Sourd, R. C. D., & Mahieu, P.-A. (2018). The effect of attribute-alternative matrix displays on preferences and processing strategies. *Journal of Choice Modelling*, *29*, 113–132. https://doi.org/10.1016/j.jocm.2018.01.001

Sever, I., & Verbič, M. (2018). Providing information to respondents in complex choice studies: a survey on recreational trail preferences in an urban nature park. *Landscape and Urban Planning*, *169*, 160–177. https://doi.org/10.1016/j.landurbplan.2017.09.003

Sicsic, J., Krucien, N., & Franc, C. (2016). What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. *Social Science & Medicine*, *167*, 116–127. https://doi.org/10.1016/j.socscimed.2016.08.050

Spinks, J., & Mortimer, D. (2015). The effect of traffic lights and regulatory statements on the choice between complementary and conventional medicines in Australia: results from a discrete choice experiment. *Social Science & Medicine*, *124*, 257–265. https://doi.org/10.1016/j.socscimed.2014.11.049

Torquati, B., Tempesta, T., Vecchiato, D., & Venanzi, S. (2018). Tasty or sustainable? The effect of product sensory experience on a sustainable new food product: an application of discrete choice experiments on Chianina tinned beef. *Sustainability*, *10*(8), Article 2795. https://doi.org/10.3390/su10082795

Uchida, H., Onozaka, Y., Morita, T., & Managi, S. (2014). Demand for ecolabeled seafood in the Japanese market: a conjoint analysis of the impact of information and interaction with other labels. *Food Policy*, *44*, 68–76. https://doi.org/10.1016/j.foodpol.2013.10.002

Vass, C. M., Rigby, D., & Payne, K. (2018). Investigating the heterogeneity in women's preferences for breast screening: does the communication of risk matter? *Value in Health*, *21*(2), 219–228. https://doi.org/10.1016/j.jval.2017.07.010

Veldwijk, J., Essers, B. A. B., Lambooij, M. S., Dirksen, C. D., Smit, H. A., & de Wit, G. A. (2016). Survival or mortality: does risk attribute framing influence decision-making behavior in a discrete choice experiment? *Value in Health*, *19*(2), 202–209. https://doi.org/10.1016/j.jval.2015.11.004

von Arx, L.-B., Johnson, F. R., Mørkbak, M. R., & Kjær, T. (2017). Be careful what you ask for: effects of benefit descriptions on diabetes patients' benefit-risk tradeoff preferences. *Value In Health*, *20*(4), 67–678. https://doi.org/10.1016/j.jval.2016.11.023

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal-Microeconomics*, *4*(4), 145–171. https://doi.org/10.1257/mic.4.4.145

Williamson, P. O., Lockshin, L., Francis, I. L., & Mueller Loose, S. (2016). Influencing consumer choice: short and medium term effect of country of origin information on wine choice. *Food Quality and Preference*, *51*, 89–99. https://doi.org/10.1016/j.foodqual.2016.02.018

Wright, S. J., Ulph, F., Dharni, N., & Payne, K. (2017). Eliciting preferences for information provision in newborn bloodspot screening programs. *Value in Health*, *20*(4), 651–661. https://doi.org/10.1016/j.jval.2016.11.007

Yeh, C.-H., Hartmann, M., & Hirsch, S. (2018). Does information on equivalence of standards direct choice? Evidence for organic labels from different countries-of-origin. *Food Quality and Preference*, *65*, 28–39. https://doi.org/10.1016/j.foodqual.2017.12.004

Zhang, J., & Adamowicz, W. L. (2011). Unraveling the choice format effect: a context-dependent random utility model. *Land Economics*, *87*(4), 730–743. https://doi.org/http://le.uwpress.org/content/by/year

Zhao, M., Johnston, R. J., & Schultz, E. T. (2013). What to value and how? Ecological indicator choices in stated preference valuation. *Environmental and Resource Economics*, *56*(1), 3–25. https://doi.org/10.1007/s10640-013-9636-0

Zhou, H., & Bukenya, J. O. (2016). Information inefficiency and willingness-to-pay for energy-efficient technology: a stated preference approach for China energy label. *Energy Policy*, *91*, 12–21. https://doi.org/10.1016/j.enpol.2015.12.040

Appendix 3B: Definitions and scope of general study details extracted

Country of study	Country where study was conducted
Delivery of study	How was survey administered? Online (required access to the internet and survey accessed via a link or website), mail survey (paper survey requiring use of postal services to deliver and/or return survey), specified location (specific location/room where study was conducted including laboratory setting, could be done over several sessions with multiple respondents in the room. It also includes recruitment at locations of interest such as shopping centres, fishing sites, walking sites), computer assisted personal interviews or CAPI (interviewer present and available for assistance while respondent completed survey on a computer), face to face interviews (a key distinguishing feature from CAPI is that the interviewer is the one that records answers)
Population of interest	General population, convenience samples: priority of convenience in recruitment e.g. university staff or university students, study specific population: samples were specific to the topic or study of interest e.g. patient population (required to have a particular condition to qualify for the study), health providers (doctors, nurses), users (a subcategory of general population where being a consumer or user of a particular product/service is required to be a respondent), children (those aged 17 years and under).
Study Sample Size	Total number of respondents: this is the number of respondents that are reported to have completed the study. 'Not reported' is used in the case where this information was not available or information provided was not clear in the study.

Appendix 3C: Definitions and scope of study development details

Were qualitative methods used?	Were qualitative methods used as part of the development of the study? Not reported if not mentioned in paper
No. of methods used?	Number of qualitative methods used to develop the study
Types of qualitative methods used	E.g. focus groups: usually involves an interviewer with multiple persons of interest in one session, could include the use different groups of people over multiple session; interviews: any type of qualitative interview e.g. includes think aloud/cognitive interviews, in-depth consultation one on one with interviewee; consultation with experts/stakeholders: where some sort of advice is sought from a person or multiple persons of interest
Rationale for any qualitative methods	Development of attributes/levels: contributed to the refinement of the choice, wording or presentation of attributes and/or levels used in the finalised DCE. Development of the survey: used to refine any part of the survey e.g. introductory information, concepts/descriptions used in the survey, other questions/tasks involved in the survey etc. Development of presentation difference/s: where methods have been used specifically to refine presentation difference/s used
Other methods used to develop study?	Apart from qualitative methods, were there other methods used to develop the survey or parts of the survey?
No. of methods used?	Number of methods (besides qualitative methods) used to develop the study
Types of methods used	Pilot study: pretesting survey or parts of survey often in the form of a draft version of the survey (could be done once or multiple times) in order to finalise survey used. Evidence from published literature: includes literature reviews, evidence from specific studies, systematic reviews. Reflection of the market: use of the market of interest to develop the survey, usually the attributes and levels, this is often done when the authors desire the survey to be as close as possible to what is seen in the market.
Rational for methods used	Development of attributes/levels: contributed to the refinement of the choice, wording or presentation of attributes and/or levels used in the finalised DCE. Development of the survey: used to refine any part of the survey e.g. introductory information, concepts/descriptions used in the survey, other questions/tasks involved in the survey etc. Development of presentation difference/s: where methods have been used specifically to refine presentation difference/s used. Priors to create choice sets: any piloting/pretesting work done to inform the construction of choice sets

Appendix 3D: Definitions and scope of validity check details

Were there any checks for validity?	Were there any checks used to test for validity as part of the survey? Not reported if not mentioned in study. This is only for pre-planned tests, not post-hoc analysis.
Types of validity check	Types: repeated choice sets (one or more choice sets repeated for respondents to complete, choice set could be slightly modified e.g. options/attributes in a different order), dominance test (where a choice set intentionally contains an option which is superior to the others), hold out task/s (where 1 or more tasks was not used as part of the modelling but used instead to test the predictive ability of the model itself), transitivity test (inclusion of test for the concept of transitivity e.g. if A preferred to B, and B preferred to C then A is preferred to C), other (description)

Appendix 3E: Definitions and scope of choice set presentation details

No. of attributes	Number of attributes that vary at least some of the time when seen by respondents, in some cases, this may be different from the total number of attributes included in the study. Attributes that are constant across all choice sets are not counted, since if they are the same across all choice sets, and they are part of the choice context/vignette/scenario.
How were choice sets assigned?	In blocks: respondents only see a portion of the total number of choice sets, the choice sets seen could be presented in random order. Random order: each respondents sees all choice sets, except in random order
Total no. of choice sets	Total number of choice sets constructed and included in the study, this may be different from the number of choice sets seen by each respondent
No. of choice sets per respondent	Number of choice sets shown and completed by respondents, this may include repeated choice sets that were included for interval validity checks
No. of options	Number of options, excluding opt out (separate data point)
Use of an opt out?	Opt out: defined as an option where respondents would not have to 'choose' e.g. status quo, none option. It would not count as an opt out if the option is not explicitly framed as such e.g. a 'no screening option' but shown as Option X in choice set with levels fixed to what they would be if respondents chose no screening. No words saying this is the 'no screening option'/status quo option
Was opt out part of choice set?	Was it part of the choice set? Or a separate question asked after the choice set?

Appendix 3F: Definitions and scope of choice set construction details

What is the design designed to detect	This has been separated into: main effects, main effects and 2 way interactions: can be all two way interaction effects or select 2 way interactions, main effects and interaction effects (if type of interaction not specified), main effects, 2 way interactions and higher order interaction effects
What was the design optimised for?	D- efficiency, A- efficiency, C-efficiency etc.
Software used	Software used to construct choice sets
Method to create choice sets	Technique used to create choice sets e.g. generator developed designs (e.g. Street and Burgess (2007)). Methods requiring changes to options: include random changes and based on improved value of function e.g. Modified Federov algorithm which involves exchanging profiles to improve design. Methods requiring changes to attributes/levels: include randomly or based on an improved value of a function e.g. coordinate-exchange algorithms which start with random collection of choice sets, change attributes and levels for the whole design by swapping, relabelling, cycling (Zwerina (Huber & Zwerina, 1996), Sándor and Wedel (2002)). Randomly chosen pairs: designs generated randomly (Oppe and Van Hout (2017)), designs may include restrictions on certain attribute/level combinations.
Were priors used?	Were priors used to inform the final choice sets used in the study? This could include point priors or distribution of priors (in the case of Bayesian design methods)
Bayesian designs: priors included in study?	If a Bayesian design method was used, did the study include information about what priors were used?

Appendix 3G: Definitions and scope of arm details

No. of arms	Number of arms in the study as reported in the methods or results section
How were arms assigned?	This is based on what is reported in the methods and results section. Random: random assignment to arm. Random by mail: this is for surveys that were conducted by mail, and usually involves a random component to the distribution of surveys, however response rate for each arm cannot be controlled. If the study does not report details of assignment to arm, then "Not reported" is used. If another method of arm assignment was used, this will be explained in further detail in the scoping review.
No. of part/s of DCE different between arms?	How many parts of the survey/DCE were different between arms as part of the experimental manipulation?
List part/s of DCE different between arms	Parts include: introductory section, vignette, choice sets, options, attributes. Definitions can be found in the introductory chapter of the thesis. Other aspects that could be different: can include incentives to participate: where incentives are used in some in the DCE and is systematically manipulated to be different between arms, incentives could be for respondent participation, other - consequences of study: manipulate consequences of study
No. of ways the DCE was presented differently to respondents between arms?	In what ways were the DCE presented differently to respondents in arms ?
List of ways the presentation of the DCE differed between arms	Framing of text: refers to manipulations that change the words/terminology used in the DCE e.g. positive vs. negative framing, concept/idea/term is represented differently in arms. Amount of information provided: each arm receives a different amount of information e.g. one group receives extra information about the topic whereas the other arm receives basic information. Presentation format of information: refers to any situation where information is presented differently between arms that is not text based e.g. one arm has an attribute with text whereas the other arm receives the text information and in addition receives a numeric explanation, presentation formats could include use of pictures, graphs, rating systems, videos/3D moving images etc. Information could also be different in terms of: number of choice sets/options/attributes as seen by respondents. Other: could include systematically manipulation of the experience respondent has with the topic e.g. one arm could receive a taste test of the product prior to completing the choice sets.

Appendix 3H: Definitions and scope of data analysis details

No. of estimation procedures used	This refers to the main models that were used to analyse the DCE data. It also includes estimation procedures used specifically to detect any potential differences between arms
Details of estimation procedures used	Common estimation procedures include: MNL/clogit, MXL, and gmnl. Variations of clogit: this refers to variations on the standard MNL or clogit models including heteroskedastic clogit, nested clogit, and alternative specific clogit. Latent class analysis (LCA): this can also include variations such as scale adjusted LCA, variance heterogeneity LCA. Hierarchical Bayes (HB) procedures: this includes variants of the MNL and MXL.
Were there differences between arms?	Based on the analyses in the results section of the study, were there *any* reported differences in results between arms?

Appendix 4A: Ethics approval for qualitative interviews

CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

UTS

1 August 2018

Dear Alice

RE: 'Patient preferences in the assessment of chemotherapy induced peripheral neuropathy (CIPN)': A discrete choice experiment

At the meeting on 10th July 2018, the CHERE management team agreed that this research project is appropriate to be conducted under CHERE's program ethics approval from the UTS Human Research Ethics Committee (UTS HREC REF NO. ETH18-2507).

Yours sincerely

Production Note:
Signature removed
prior to publication.

Kees van Gool

Deputy Director, CHERE

Appendix 4B: Cognitive interview protocol, Interviews 1-4

Author to bring

- Prompt sheet
- Probing Sheet
- Info sheet/Consent Form
- Laptop
- Link to survey and Excel spreadsheet
- Notepaper & pen
- Audio recorder
- Water

Introduction

- Settle them in the meeting room
- Give participant 5 minutes to sign ethics document
- Introduce myself
- We are doing a survey looking at patients prefer when it comes to being assessed for CIPN
- What we are after today is for you to help us test the survey, before we send it out.
- Will be looking at my sheet from time to time, questions I need to cover

Consent

- Confidentiality & recording [start recording once consented]
- Have you read the Participant Information Sheet/Consent Form? Do you have any questions?
- [SIGN CONSENT]

Instructions

- I'm not so interested in your actual answers, more in how you get to them
- As you go through the survey, I'd like you to 'think aloud'. This means I want you to tell me what you are thinking about as you answer the questions.
- I might sometimes also ask you some extra questions about your answers as you go
- [THINK ALOUD EXAMPLE] – For example, if one of the questions is "when was the last time you saw a doctor?" I might just write down "December", but if I was thinking aloud I might say "*Well, I saw my GP when I was sick the day of my office Christmas party. I'm not sure when that was, but it was probably December. I wonder if they want the year or just the month or the actual date? I guess a GP counts as a doctor?*"
- I didn't write these questions – don't worry about hurting my feelings

- If you need water/toilet, please let me know

Participant to look at information condition

- As you go through, don't forget to think out loud.

Participant to complete choice sets

- To load survey https://chere.surveyengine.com/survey/389/436
- In actual survey, underlined words in survey, can click on link to refer back to definition

Dummy choice sets and discussion

- Do you prefer version 1 or 2
- Is there anything you would change/add?
- What are your thoughts on the ordering of attributes?
- How did you find the use of bolding/underline?

Debrief

- Generally, what did you think about the survey?
- How helpful was the visual/written information provided?
- If didn't explain underlined words, would you have known to click on it?
- Is there anything about the survey you would change?
- Was there anything you found: boring, confusing, upsetting, embarrassing, annoying?
- How would you react if you got this survey by email?
- Did you have any questions?
- Wrap up

Close

- THANKYOU
- Anything else to say? [END RECORDING]

Appendix 4C: Cognitive interview protocol, Interviews 5-6

Author to bring

- Interview Plan sheet
- Probing Sheet
- Info sheet/Consent Form
- Laptop
- Notepaper & pen
- Audio recorder
- Water

Introduction

- Settle them in the meeting room
- Give participant 5 minutes to sign ethics document
- Introduce myself
- We are doing a survey looking at patients prefer when it comes to being assessed for CIPN
- What we are after today is for you to help us test the survey, before we send it out.
- Will be looking at my sheet from time to time, questions I need to cover

Consent

- Confidentiality & recording [start recording once consented]
- Have you read the PIC? Do you have any questions?
- [SIGN CONSENT]

Instructions

- I'm not so interested in your actual answers, more in how you get to them
- As you go through the survey, I'd like you to 'think aloud'. This means I want you to tell me what you are thinking about as you answer the questions.
- I might sometimes also ask you some extra questions about your answers as you go
- [THINK ALOUD EXAMPLE] – For example, if one of the questions is "when was the last time you saw a doctor?" I might just write down "December", but if I was thinking aloud I might say "*Well, I saw my GP when I was sick the day of my office Christmas party. I'm not sure when that was, but it was probably December. I wonder if they want the year or just the month or the actual date? I guess a GP counts as a doctor?*"
- I didn't write these questions – don't worry about hurting my feelings
- If you need water/toilet, please let me know

Participant to look at information condition

- As you go through, don't forget to think out loud.

Participant to complete choice sets

- To load survey https://chere.surveyengine.com/survey/389/436
- In actual survey, underlined words in survey, can click on link to refer back to definition

Attributes and levels discussion

- Symptoms
    - What does symptoms mean to you?
    - What does impact on usual activities mean to you?
- Level of detail
    - What does 'detect large changes' mean to you?
    - How is this different to 'detect small and large changes' for you?
- Mode of admin
    - What do you think of when it says 'fill in a survey'?
    - What does being asked questions mean to you?
    - What does being examined mean to you?
    - What do you imagine when it says 'doing a range of assessments by yourself'?
- How will results influence care/treatment
    - What does modifications to your treatment mean for you?
    - What do practical interventions in your lifestyle mean to you?

Discussion on proposed changes to attributes/levels (paper versions)

- Mode of admin
    - What do you imagine when it says 'you fill in a survey by yourself'?
    - What do you imagine when it says 'doing physical tests by yourself'?
    - How does this compare to the previous version? Do you have a preference and why?
- New attribute title: 'Assessment impact on clinic time'
    - What does this feature mean to you?
    - How does this compare to the previous version? Do you have a preference and why?
- What changes would you make? Anything to add to the attributes or levels?

Debrief

- Generally, what did you think about the survey?
- How helpful was the visual/written information provided?
- Did you have to refer back to the info sheet at any point?
- Is there anything about the survey you would change? Why?
- Was there anything you found: boring, confusing, upsetting, embarrassing, annoying?
- How would you react if you got this survey by email?
- Did you have any questions?
- Wrap up

Close

- THANKYOU
- Anything else to say? [END RECORDING]

Appendix 4D: Prompting questions

If can't choose	What is it about these options that makes it difficult for you to choose between them?
If asks for definition	What does the term "XXX" mean to you / What would you assume if I wasn't here? Go with your interpretation of the definition/question
If asks for instructions	What do you think you are being asked to do in this question? / What would you do if I wasn't here
If discusses remembering	How did you remember that? How well do you remember that?
General	How did you arrive at that answer? I noticed that you XXX – tell me, what you were thinking?
Can't answer	What was going through your mind as you tried to answer the question?
Period of silence	You took a little while to answer that. What were you thinking about?
Expresses uncertainty	It sounds like you had some difficulty answering that question? Can you tell me why?
Changed answer	What occurred to you that caused you to change your answer?
Conditional answer	You seem a little unsure. Was there something unclear about the question?

Appendix 4E: Summary of cognitive interviews

Section/Question	Action	Feedback
Information sheet (paper format in first 3 interviews, in survey from 4th interview)	Revised	All participants would try and 'self-diagnose' themselves based on the info sheet. There was some confusion about what was meant by CIPN. It was found that participant understanding of the info sheet did affect their understanding of the choice sets. Action: For the 4th interview onwards, addition of a description and definition of CIPN to the info sheet. A brief introduction to the info sheet was added to provide more context to the survey. Changes were also made to headings of sections to create more continuity in the info sheet.
Information sheet: defining CIPN	Revised	A participant suggested the use of the term peripheral neuropathy instead of CIPN. The participant was able to understand what we meant by CIPN based on our definition, however, it was raised that this is not a common term that patients use. It was noted that other participants also had not heard of the term CIPN before. Action: For the finalised survey, changed terminology used to peripheral neuropathy instead of CIPN.
Information sheet: Pictures of physical tests	Kept	All participants were asked about their opinion on the addition of pictures to the verbal description of different physical tests. Majority of participants were positive about their inclusion, 1 participant was indifferent.
Information sheet: Small and large change	Revised	Significant revision was made to this section after the 3rd participant as it was recognised that this was particularly confusing (see Attribute: Level of detail). Three participants saw the revised version, 1 participant mentioned that the examples helped to explain the differences better. 1 of these participants still required clarification, although this particular participant only had a brief look through the information sheet section. Action: Finalised wording based on IN FOCUS patient representative suggested wording.
Choice sets General	Noted	It was observed that it took a while (2-3 choice sets) for participants to get used to the choice set format. 2/4 participants gave feedback that there were too many choice sets. 1 participant did not like the choice set format. 1 participant felt the number of choice sets was fine, and said could do more (although it should be noted this participant was highly motivated and intelligent, which is not necessarily a reflection of the potential sample)
Attribute: Level of Detail	Revised	Based on participant 'think aloud' it was clear that there was confusion about what was meant by 'small' versus 'large' change in CIPN. Action: From the 4th interview onwards, the wording of the two levels were altered to better reflect what we mean (the assessment detects large changes vs the assessment detects small AND large changes). The info sheet section explaining

		this attribute was also modified to provide better clarity, this included adding a table of examples.
Attribute: Frequency of Administration	Removed	There was confusion about the context of 'frequency' i.e. frequency during chemo? Before or after chemo? For example, one participant said would choose once a week during chemo but would prefer once a month after chemo. An issue is also raised about the levels, one participant suggested the addition of 3 weeks as a level, which is a common frequency for chemotherapy sessions. Action: For the 4th interview onwards, after a discussion to confirm the intended use of this assessment, this attribute was changed to 'Frequency of Administration during Chemotherapy'. Action: Attribute removed
Attribute: Time for Assessment	Revised	Some participants raised the issue that this was not really an important point of consideration for them. Participants of this opinion, felt an assessment should take as long as it needs to take. Action: Kept, of interest to clinicians, attribute changed to 'Impact on clinic time'
Attribute: How will results influence care/treatment	Kept	Question was asked about whether this attribute should be split into 2 attributes. ¾ participants viewed these as connected concepts and preferred to keep them together.
Attribute: How will results influence care/treatment	Revised	It was noted that all participants were always drawn back to level 'may lead to modification in treatment'. This level (despite our initial assumptions) was viewed as highly positive, with participants viewing level 'will not change your care/treatment' as negative. As this was discovered to be a level/attribute of concern, more focus was given to this in the last 2 cognitive interviews. 1 participant viewed the term 'modifications to your treatment' as a change to a different chemotherapy drug, the other viewed it as a reduction in dosage Action: wording of levels changed to what was suggested by IN FOCUS patient representatives.
Questions about Cancer/Chemo	Revised	All participants found these questions straight forward, with no trouble recalling the information. It was noted in the question about the year of diagnosis, one participant had had cancer twice in their lifetime. This participant just put down the year of first diagnosis. Action: Specified we would like year of first diagnosis
Demographic Questions	Removed	For the question about what is your sex, had 3 options of female, male, prefer not to disclose. One participant made the comment that this feels rather negative and is the only option if the person identifies as intersex or other. 2 participants found it was difficult to choose an appropriate response to 'what is your current employment status'. A participant suggested the inclusion of 'home duties' as an option for employment status. Action: For the 4th interview onwards, question about employment was revised to 'What best describes you current

		employment status', question about sex, 2 more options were added: intersex and indeterminate. Action: question removed as not relevant to aims
Difficulty Question	Revised	It was apparent that most participants did take a while to get used to the format Action: A rating question about the difficulty of the survey was added for the 4th participant onwards. Action: question changed rating of how easy it was to complete/tell the difference between options
Information Sheet Questions	Revised	Two questions were added from the 4th participant onwards about whether participant referred back to the info provided and their thoughts on the info provided. 4th participant was confused about what was meant by 'information', it was interpreted as information provided by the choice sets rather than the information provided prior to the choice sets Action: changed to one question to rate usefulness of background information given
Question: Stage of Cancer	Removed	2 participants were uncertain about the definition of the stages of cancer. 1 participant made the comment that cancer was not really discussed in stages, at least at the patient level. Action: removed as confusing and not relevant to aims
Formatting/style (below)		
Bolding/ underlining	Kept	In response to question about the participant's opinion on the use of bolding and underlining in the levels. All participants found it helpful and would choose to keep this in the survey.
Attribute: symptoms and quality of life	Revised	One participant made a comment about formatting, symptoms was not bolded in the 2nd level, which was a bit confusing. Action: For the 4th interview onwards, the word 'symptoms' was bolded in the 2nd level so both 'symptoms' and 'usual activities' were bolded.
Questions after choice sets	Revised	A participant made comment that a section to provide feedback would have been good. Action: For the 4th interview onwards, optional feedback section was also added.
Adding hyperlinks/ mouse overs	Noted	A participant gave feedback that length of information sheet prior to the survey was long, would have liked to see more 'pop up' definition that would appear if hover mouse over a key word. Action: A discussion was had, there needs to be a balance between convenience for participants and our desire to know if participants did refer back to the info sheet provided. Noted but no further action taken.
Error messages	Revised	It was noted for a couple of participants, that the most common error is accidently skipping a question. The error message given is not helpful in identifying the issue. Potential action: Error message altered to clarify issue.

Appendix 4F: Consultation session with health economists and DCE experts

Information sheet

- To rephrase wording to 'would experience' , to highlight that these are 'potential common symptoms'
    - Idea that this is not a diagnosis, but rather potentially what could happen if you had chemotherapy and developed peripheral neuropathy
- Need to explicitly mention that this is all hypothetical – Richard for wording
- Reminder that this is an explanation of key terms and conditions at the bottom of every page (some people won't read it at the top)
- Get rid of the grading system – not necessary and makes it complicated, to just have small or large changes
- Specify examples of 'common potential symptoms'
- Physical tests is too vague – maybe physical assessments
- For mode of administration, have a summary of the potential assessments, not just physical tests, otherwise give impression that assessments are all just ones where assessments will be done to you
- To specify which physical tests are done by physicians versus by yourself
- Practical interventions: link to small and large changes, what are the most common interventions?
- Add section on the rationale for testing, benefits for testing – why should someone do this assessment?
    - To find out what proportion of people end up with the peripheral neuropathy going away, what proportion end up with permanent symptoms?

Choice scenario

- Need to have one
- Maybe: Imagine you have been diagnosed with CIPN, you will be doing an assessment which looks at your symptoms and the impact of your symptoms on your usual activities. You are asked to choose between two assessments with different features. Which assessment would you choose?

Attributes and levels

- Do these attributes and levels match to the current 6 assessments available?
- Mode of administration
    - Examined is vague- does this mean assessments done to you?
    - Specify what types of tests- physical tests is too vague
    - In reality, unrealistic to only do survey or physical tests, would often be a combination, maybe better to have different combinations e.g. fill in a survey at home and do physical assessments at your apt and discuss your results during your apt, you fill in a survey while waiting for your apt and do physical assessment at your apt and discuss your results during your apt, you do some physical assessments at home, fill in a survey at home and you discuss the results during your appointment
- Frequency of administration
    - Fortnight and 3 weeks is too close together
    - 3 weeks may be optimal for patients that have to go to an apt to get the chemo but there are some that only require pills so don't need an apt
    - Need to think further about range: maybe look at requirements for 6 assessments
- Assessment impact on clinic time
    - Attribute contradicts 'by yourself' level in mode of admin
    - Maybe change back to time for assessment?
- How will results influence care/treatment
    - To split into two attributes
    - Modifications in your treatment: focus group thought it was too vague, need to be explicit about e.g. instead of modifications to your treatment have this may lead to a reduction in chemo dose, this may lead to changing the type of chemo drug

Additional notes from second note taker

- to modify –'common symptoms of peripheral neuropathy 'may/ can' include'
- to have mouse overs
- To change 'potential symptoms' to 'most common symptoms'
- To get rid of grading terms in the examples of small and large changes

- To change wording of examples of physical tests e.g. If you would undergo this assessment, you would be asked if you can……
- Group liked the pictures
- To indicate which tests are by yourself vs by GP
- Examples of practical interventions: to list most common
    - Example: 'rails and bathroom modifications' to use a different word beside modifications
- Permanent or can be resolved? (symptoms)
    - does this impact on their attitude towards the assessment
- To think about
    - What are the benefits of testing?
    - What is the rationale behind it?
- What is the vignette?
- Do attributes match/correlate with all 6? (6 assessments identified from the Delphi survey)
- Symptoms and quality of life
    - Use of phrase 'activities of daily living'
    - Could include other variables e.g. mental health, EQ-5D components
    - Comment: pointless, you won't pay (attention?) to symptoms without developing what impact is
- Level of detail
    - Use 'can/may' e.g. the assessment can detect large changes only
- Mode of admin
    - 'Examined' vague word, better to have asked and examined
    - Add 'at home' for you do some physical tests by yourself
    - More information about the tests in these levels
    - Also could do a mix, some at home, some at doctors
    - What is reasonable in clinical practice?
- Frequency of admin
    - Levels too close, use 1,3,6 and 12 weeks
- How will results influence care/treatment
    - To separate into two attributes

Appendix 4G: Sets of priors used in simulations

Zero Priors	all 13 parameters set to 0, probability of choosing one option over another is exactly 50%

Prior set 1

Parameter	Prior	Reasoning
1	1.4	respondents more concerned about symptoms impacting on their daily life
2	−1.1	respondents prefer to keep track of all changes
3	0.8	prefer to be asked more questions than less
4	0.6	prefer more questions
5	−0.9	prefer to have a questionnaire than not
6	1.3	prefer to have tests
7	0.8	prefer to have tests
8	−0.8	prefer to have tests
9	−0.1	don't mind an extra few minutes, but during usual clinic time preferable
10	−0.6	don't mind an extra few minutes, but 30 minutes too much
11	−1.0	a separate appointment seen as an inconvenience
12	1.25	seen as positive for treatment to be adjusted to minimise symptoms of CIPN
13	0.6	seen as positive for some help to cope with daily life

Prior set 2

Parameter	Prior	Reasoning
1	−1.3	respondents care about symptoms more than impact
2	0.9	respondents only care about changes that affect their condition significantly
3	−0.2	prefer less questions
4	−0.6	prefer less questions
5	0.75	prefer having no questionnaire to having to answer a brief questionnaire
6	−0.9	prefer to have tests done on them
7	−1.4	technical tests seen as intimidating
8	0.8	prefer to not have a test
9	0.7	like having extra time devoted to testing for CIPN
10	0.35	like having extra time devoted to testing for CIPN
11	0.9	having a separate appointment devoted to testing for CIPN most preferable
12	−0.9	prefer to make decision together
13	−0.8	prefer to make decision together

Prior set 3

Parameter	Prior	Reasoning
1	1.4	respondents more concerned about symptoms impacting on their daily life
2	0.9	respondents only care about changes that affect their condition significantly
3	0.8	don't mind 12 relative to 3 questions
4	−0.6	prefer less questions, 12 ok but 20 too many
5	−0.9	prefer to have a questionnaire than not
6	1.3	prefer to have some involvement in the tests
7	−1.4	technical tests seen as intimidating
8	0.8	prefer to not have a test
9	−0.5	prefer no extra time
10	−0.7	prefer no extra time
11	−0.9	prefer no extra time
12	−1.68	strongly resistant to any change in treatment
13	0.6	seen as positive for some help to cope with daily life

Appendix 4H: Ethics approval No. ETH19-3464: copy of email

Dear Applicant

Thank you for your response to the Committee's comments for your project titled, "Preferences for how Chemotherapy Induced Peripheral Neuropathy (CIPN) is assessed: A Discrete Choice Experiment.". The Committee agreed that this application now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application.

You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all UTS policies and guidelines including the Research Management Policy (http://www.gsu.uts.edu.au/policies/research-management-policy.html).

Your approval number is UTS HREC REF NO. ETH19-3464.

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.

The following standard conditions apply to your approval:

• Your approval number must be included in all participant material and advertisements. Any advertisements on Staff Connect without an approval number will be removed.

• The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat (Research.Ethics@uts.edu.au).

• The Principal Investigator will notify the UTS HREC of any event that requires a modification to the protocol or other project documents, and submit any required amendments prior to implementation. Instructions can be found at https://staff.uts.edu.au/topichub/Pages/Researching/Research%20Ethics%20and%20Integrity/Human%20research%20ethics/Post-approval/post-approval.aspx#tab2.

• The Principal Investigator will promptly report adverse events to the Ethics Secretariat (Research.Ethics@uts.edu.au). An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.

• The Principal Investigator will report to the UTS HREC annually and notify the HREC when the project is completed at all sites. The Principal Investigator will notify the UTS HREC of any plan to extend the duration of the project past the approval period listed above through the progress report.

• The Principal Investigator will obtain any additional approvals or authorisations as required (e.g. from other ethics committees, collaborating institutions, supporting organisations).

• The Principal Investigator will notify the UTS HREC of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

I also refer you to the AVCC guidelines relating to the storage of data, which require that data be kept for a minimum of 5 years after publication of research. However, in NSW, longer retention requirements are required for research on human subjects with potential long-term effects, research with long-term environmental effects, or research considered of national or international significance, importance, or controversy. If the data from this research project falls into one of these categories, contact University Records for advice on long-term retention.

You should consider this your official letter of approval. If you require a hardcopy please contact Research.Ethics@uts.edu.au.

If you have any queries about your ethics approval, or require any amendments to your research in the future, please do not hesitate to contact Research.Ethics@uts.edu.au.

Yours sincerely,

A/Prof Beata Bajorek
Chairperson
UTS Human Research Ethics Committee
C/- Research Office
University of Technology Sydney
E: Research.Ethics@uts.edu.au

REF: E38

Appendix 4I: Copy of full survey

Page 1: Introduction/ consent page

UTS HREC REF NO. ETH19-3464 – What features are important to include in an assessment for peripheral neuropathy induced by chemotherapy?

What is the research study about?
The purpose of this research/online survey is to understand what features are important to include in an assessment for peripheral neuropathy resulting from chemotherapy treatment.

As part of the survey, you will be asked about your preferences for different peripheral neuropathy assessment tools. You will also be asked some questions relating to your cancer/chemotherapy treatment and questions relating to general demographics.

Who is conducting this research?
My name is Alice Yu and I am a PhD student at UTS. My supervisor is Prof. Deborah Street.

How long will this survey take?
We expect the survey to take between 15 to 20 minutes to complete.

Do I have to take part in this research study?
Participation in this study is voluntary. You can change your mind at any time and stop completing the survey without consequences.

Are there any risks/inconvenience?
We don't expect this questionnaire to cause any harm or discomfort, however if you experience feelings of distress as a result of participation in this study you can change your mind at any time and stop completing the survey without consequences.

What will happen to information about me?
Submission of the online questionnaire is an indication of your consent. By continuing to answer the survey questions you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. No identifying information will be collected as part of study nor will the researchers have access to identifying information. The data will be stored securely.
Data collected will be used for the purpose of this research project and potentially in future research projects that aim to improve on the methods used to design questionnaires of this general type. In all instances, results from the survey will be published in a form that does not identify you
We plan to discuss the results among the researchers involved in this project and with our partner organisation, the IN FOCUS Study. IN FOCUS are interested in using these results to inform the development of a potential peripheral neuropathy assessment tool. As mentioned previously, no identifying information will be collected and in any publication, results will only be in the form of summary quantitative information.

What if I have concerns or a complaint?
If you have concerns about the research that you think or my supervisor or myself can help you with, contact me on +61 2 9514 4768 or via email: Alice.Yu@chere.uts.edu.au.
Alternatively, you can contact my supervisor Prof. Deborah Street via email: Deborah.Street@uts.edu.au
If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number ETH19-3464.

What support is available?
Some people may find some of the issues raised in the survey confronting. The Cancer Council offers an Information and Support Line. If you need to, please call 13 11 20 or visit their website at https://www.cancer.org.au/about-cancer/patient-support/131120.html. Your local doctor or other health care provider may also be able to offer you advice.
If you agree to be part of the research and to research data gathered from this survey to be published in a form that does not identify you, please continue with answering the survey questions.

If you agree to be part of the research and to research data gathered from this survey to be published in a form that does not identify you, please continue with answering the survey questions.

Page 2: Notice of potential display issues (added at relaunch on 1st of November 2019)

Notice: potential display issues when completing survey on a mobile phone

We welcome the use of any device when completing this survey.

However, we have noted that there are some display issues when this survey is viewed on some mobile phones.

We recommend that you use other devices such as a computer or tablet. Thank you.

Please continue to the next page when you are ready.

Page 3

Have you ever been diagnosed with cancer?

Select only one answer

- ○ yes
- ○ no

Page 4

Have you ever received chemotherapy as a treatment for cancer?

Select only one answer

- ○ yes
- ○ no

Page 5

On this page and the next few pages, some background information is provided to define what is meant by peripheral neuropathy and to explain some key definitions and terms.

Please read through them.

Please note: the following information *is not* intended to reflect your personal circumstances. They *are not* intended to have relevance for your personal decisions

What is peripheral neuropathy and how is it related to chemotherapy?

Peripheral neuropathy is a major side effect experienced as a result of chemotherapy treatment, affecting up to 40% of cancer survivors.

Peripheral neuropathy occurs when drugs used to treat cancer cause damage to the peripheral nerves (i.e. nerves in the hands and feet, sometimes extending to the arms and legs).

Some common symptoms of peripheral neuropathy can include:
- Numbness
- Tingling, 'pins and needles' or electric shock-like sensations
- Burning sensations
- Balance problems
- Muscle weakness
- Constipation
- Decreased reflexes

These symptoms can lead to problems with completing everyday activities, such as:
- Trouble using your fingers to pick up or hold things; dropping things
- Trouble with buttoning clothes
- Tripping or stumbling while walking

Page 6

Assessing for peripheral neuropathy

There are different ways in which a patient may be assessed for peripheral neuropathy while they undergo chemotherapy treatment for cancer.

Please continue to the next page.

Page 7

One of the ways in which a peripheral neuropathy assessment can be different, is in terms of the level of detail of the assessment.

What do we mean by this?

The peripheral neuropathy assessment could pick up:

• *minor and major* nerve damage, including *small changes* in your condition, whether it is important or not OR

• only *major* nerve damage and *large changes* in your condition

What is the difference between an assessment that picks up small versus a large change in peripheral neuropathy?

Potential Symptom	Examples of Assessment Detecting Small Change	Examples of Assessment Detecting Large Change
Numbness in hands	The assessment detects when the numbness becomes worse, although you are still able to perform the same tasks e.g. using knives or dressing yourself.	The assessment detects when the numbness that did not interfere with daily activities worsens such that chopping vegetables and carrying pots becomes difficult.
Numbness in feet	The assessment detects when the numbness becomes worse, but it does not affect your ability to walk or maintain balance.	The assessment detects when the numbness in the feet progresses to being unsteady on your feet, especially at night.
Pain/tingling in hands or feet	The assessment detects when increasing pain or tingling occurs.	The assessment detects when pain or tingling that makes cooking difficult worsens to being unable to button clothing or use keys.

Page 8:

Another way in which a peripheral neuropathy assessment can be different is in terms of the type of physical tests involved.

What do we mean by this?

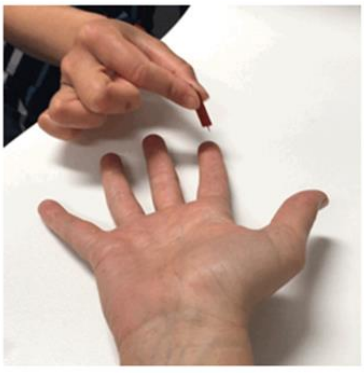
The peripheral neuropathy assessment may involve different types of physical tests on your body.

There are three main types of physical tests:

• Clinician administered tests
• Patient activity based tests
• Technical tests

Examples of clinician administered tests

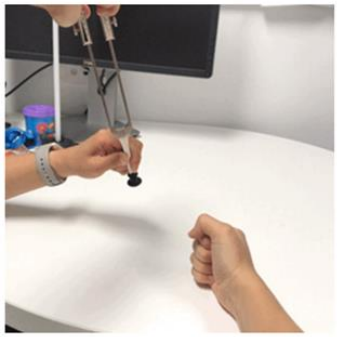
Sharp and dull test
This is a pain perception test where you will be asked to perceive the difference between a sharp or dull stimulus.
An example of a sharp and dull test is given below.



Page 8 (continued)

Tuning fork test

This is a vibration sensation test. A tuning fork is placed against different parts of your body and you are asked to say when you feel the vibration stop.
An example of a tuning fork test is given below.



Examples of patient activity based tests

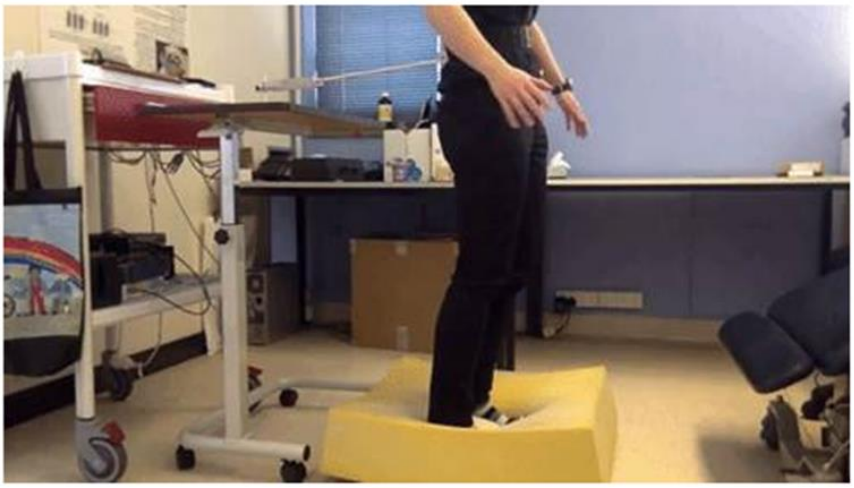
Peg Board Test

This is to test the manual dexterity of the hands.
An example of a peg board test is the grooved pegboard test, which consists of holes that are randomly positioned.
The pegs need to be rotated to match the hole before they can be inserted.
An example of a grooved pegboard test is given below.

Page 8 (continued)

Sway Test

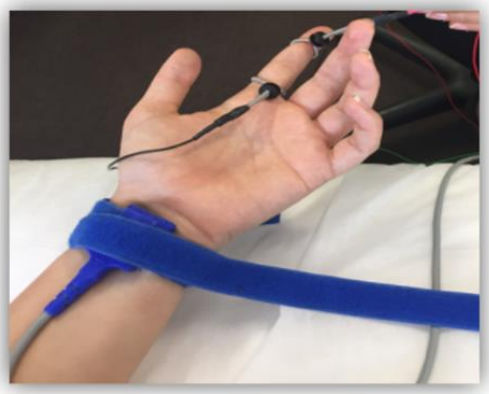
An example of a sway test is the Romberg's test, which is a test of the body's sense of positioning.
You will be asked to close your eyes and you will be assessed on your sense of balance.
An example of a sway test is given below.



Page 8 (continued)

Example of a technical test

Nerve conduction studies
These tests record the properties of electrical impulses travelling along your nerves using stick-on-electrodes and impulses which feel a little like tapping. An example of a nerve conduction study is given below.



Page 9

Other ways in which a peripheral neuropathy assessment can be different include:

- Whether the assessment focuses on symptoms or how symptoms impact on usual activities
- The length of questionnaire that the patient may have to complete
- The impact of assessment on clinic time
- How will results from the assessment influence care/treatment

Page 10

Choice Questions: Instructions

On the next few pages, you will be asked to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect.

You will then be presented information describing the features of two potential peripheral neuropathy assessment tools.

You will be asked to decide which assessment tool you prefer.

You will be asked this question 8 times. For each of the 8 questions, the values for the features may differ e.g. the number of questions on the questionnaire may be different.

*These questions are hypothetical. The values used for the features **are not** intended to reflect your personal circumstances. They have no relevance for your personal decisions. There are no right or wrong answers to these questions; we are simply interested in your views.*

When you are ready, please press next to continue to the choice questions.

Page 11 Choice Set Example 1 (each respondent randomly assigned 1 block)

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about your *symptoms*
Level of detail	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	12 questions to answer	20 questions to answer
Physical test/s	No physical test	Patient activity based test e.g. peg board test, sway test
Impact on clinic time	Usual clinic time plus 30 minutes extra	Usual clinic time plus 10 minutes extra
How will results influence care/treatment	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

Page 12 Choice Set Example 2

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about your *symptoms*	The assessment asks about your *symptoms*
Level of detail	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition
Questionnaire	3 questions to answer	20 questions to answer
Physical test/s	Clinician administered test e.g. sharp and dull test, tuning fork test	Patient activity based test e.g. peg board test, sway test
Impact on clinic time	During usual clinic time	Usual clinic time plus 30 minutes extra
How will results influence care/treatment	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment	The *doctor may change your general care* (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time
Which would you choose?	✓ Assessment A	○ Assessment B

Page 13 Choice Set Example 3

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about your *symptoms*
Level of detail	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition
Questionnaire	20 questions to answer	12 questions to answer
Physical test/s	Patient activity based test e.g. peg board test, sway test	No physical test
Impact on clinic time	You require a separate appointment, which can take up to 60 minutes	Usual clinic time plus 30 minutes extra
How will results influence care/treatment	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

Page 14 Choice Set Example 4

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about how your *symptoms impact on your usual activities*
Level of detail	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	No questionnaire	12 questions to answer
Physical test/s	Clinician administered test e.g. sharp and dull test, tuning fork test	Patient activity based test e.g. peg board test, sway test
Impact on clinic time	Usual clinic time plus 30 minutes extra	During usual clinic time
How will results influence care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment
Which would you choose?	○ Assessment A	○ Assessment B

Page 15 Choice Set Example 5

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about your *symptoms*	The assessment asks about how your *symptoms impact on your usual activities*
Level of detail	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	3 questions to answer	12 questions to answer
Physical test/s	Clinician administered test e.g. sharp and dull test, tuning fork test	Technical test e.g. nerve conduction studies
Impact on clinic time	During usual clinic time	You require a separate appointment, which can take up to 60 minutes
How will results influence care/treatment	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

Page 16 Choice Set Example 6

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about how your *symptoms impact on your usual activities*
Level of detail	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	3 questions to answer	20 questions to answer
Physical test/s	No physical test	Clinician administered test e.g. sharp and dull test, tuning fork test
Impact on clinic time	Usual clinic time plus 10 minutes extra	You require a separate appointment, which can take up to 60 minutes
How will results influence care/treatment	The *doctor may change your general care* (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

Page 17 Choice Set Example 7

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about your *symptoms*	The assessment asks about your *symptoms*
Level of detail	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition	The assessment will pick up *minor and major nerve damage*, including *small changes* in your condition whether it is important or not
Questionnaire	20 questions to answer	3 questions to answer
Physical test/s	No physical test	Clinician administered test e.g. sharp and dull test, tuning fork test
Impact on clinic time	During usual clinic time	Usual clinic time plus 30 minutes extra
How will results influence care/treatment	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment
Which would you choose?	○ Assessment A	○ Assessment B

Page 18 Choice Set Example 8

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you prefer? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about your *symptoms*
Level of detail	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition
Questionnaire	3 questions to answer	12 questions to answer
Physical test/s	No physical test	Patient activity based test e.g. peg board test, sway test
Impact on clinic time	Usual clinic time plus 10 minutes extra	During usual clinic time
How will results influence care/treatment	The *doctor may change your general care* (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time	The *doctor will discuss* the results with you, *and together* you can decide what they mean for you and your care/treatment
Which would you choose?	○ Assessment A	○ Assessment B

Pages 19 – 23: Choice set evaluation questions

The next few questions are about how you found answering the choice questions

Prior to the choice questions, some information was provided to define peripheral neuropathy and to explain some key definitions and terms.

Did this information help you understand the choice questions?

Select only one answer

- ○ Yes
- ○ No
- ○ No, I already knew this information prior to this survey
- ○ No, it was confusing
- ○ I don't remember
- ○ Other. Please specify: _____

How did you decide whether Assessment A or B was better in each choice question?

Select only one answer

- ○ I considered all features for each assessment option
- ○ I considered all features which were different between the two assessment options
- ○ I considered only the features which were most important to me
- ○ Other strategy (please explain): _____

Thinking about the features of a peripheral neuropathy assessment tool. Please select which feature is the *most important* and *least important* to you:

	Whether the assessment focuses on symptoms or how symptoms impact on usual activities	Level of detail of the assessment	Length of questionnaire	Type of physical test/s	Impact of assessment on clinic time	How will results influence care/treatment
MOST Important to you	○	○	○	○	○	○
LEAST Important to you	○	○	○	○	○	○

Are there any other features of a peripheral neuropathy assessment tool, *not included* in the choice questions, that you think are important?

Select only one answer

- ○ No
- ○ Yes. Please specify: _____

Please indicate the extent to which you agree or disagree with each of the following statements about the choice questions

Select one response from each row

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Most of the time, I could easily identify the differences between assessment options	○	○	○	○	○
Most of the time, I could easily choose between the assessment options	○	○	○	○	○

Pages 24 – 31: Cancer demographic questions

The next few questions are some background questions about you

In what year did you receive your cancer diagnosis? If more than once, please indicate your initial year of cancer diagnosis.

Select only one answer

-- select one --

What type of cancer/s have you been diagnosed with? Please choose one or more from the list below.

Select all that apply

☐ Anal cancer	☐ Bladder cancer	☐ Bone cancer
☐ Bowel cancer	☐ Brain cancer	☐ Breast cancer
☐ Cancer of unknown primary	☐ Cervical cancer	☐ Head and neck cancer
☐ Kidney cancer	☐ Leukaemia	☐ Liver cancer
☐ Lung cancer	☐ Lymphoma	☐ Mesothelioma
☐ Myeloma	☐ Ovarian cancer	☐ Pancreatic cancer
☐ Skin cancer	☐ Stomach cancer	☐ Testicular cancer
☐ Thyroid cancer	☐ Uterine cancer	☐ Other type of cancer

Are you currently receiving chemotherapy ?

Select only one answer

- ○ Yes
- ○ No
- ○ I don't know

In what year, did you complete your chemotherapy treatment? If more than once, please indicate most recent year of completion.

Select only one answer

-- select one --

Have you received, or are you currently receiving, any of the following types of chemotherapy? Please choose one or more types from the following list.

Select all that apply

☐ Oxaliplatin (Eloxatin, Oxalatin, Oxaliccord, Xalox, FOLFOX, XELOX)	☐ Cisplatin (cisplatinum, Platinol)	☐ Carboplatin (Carbaccord)
☐ Paclitaxel (Taxol, Anzatax, Plaxel, Abraxane)	☐ Docetaxel (Taxotere, Dotax, Oncotaxel)	☐ Cabazitaxel (Jevtana)
☐ Vincristine	☐ Vinblastine	☐ Vinorelbine (Navelbine)
☐ Vinflunine (Javlor)	☐ Vindesine (Eldisine)	☐ Thalidomide (Thalomid)
☐ Bortezomib (Velcade)	☐ Ixabepilone	☐ Lenalidomide (Revlimid)
☐ Pomalidomide (Pomalyst)	☐ Eribulin (Halaven)	☐ I have not received any of these types of chemotherapy
☐ I don't know the name of the chemotherapy that I have received		

Have you ever been assessed for peripheral neuropathy ?

Select only one answer

- ○ Yes
- ○ No
- ○ I don't remember

Page 29 conditional on page 28 = yes

Which peripheral neuropathy assessment tool/s have you been tested with? Please choose one or more types from the following list.

Select all that apply

- ☐ Questionnaire
- ☐ Clinician administered test e.g. sharp and dull test, tuning fork test
- ☐ Patient activity based test e.g. peg board test, sway test
- ☐ Technical test e.g. nerve conduction studies
- ☐ I don't know
- ☐ Other. Please specify: _____

Have you ever had peripheral neuropathy as a result of chemotherapy treatment?

Select only one answer

- ○ Yes
- ○ No
- ○ I don't know

Page 31 conditional on page 30 = yes

Please briefly explain how you determined you had peripheral neuropathy as a result of chemotherapy treatment.

Enter text below

Page 32 Sociodemographics

What is your year of birth ?

Select only one answer

-- select one --

What is your sex?

Select only one answer

- Female
- Male
- Intersex/indeterminate/other
- Prefer not to disclose

What is the highest level of education that you have completed ?

Select only one answer

- No school certificate or other qualifications
- Secondary school
- Trade or apprenticeship
- TAFE or vocational college
- Bachelor's degree
- Postgraduate degree

Page 33 Feedback/comments section

If you have any other feedback or comments, please enter it below.

Enter text below

Page 34 Conclusion to survey

Thank You! Please press 'submit answers and finish' to end the survey.

Contact Details

If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact me (us) on +61 2 9514 4768 or via email: Alice.Yu@chere.uts.edu.au

If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number ETH19-3464.

What support is available?

Some people may find some of the issues raised in the survey confronting. The Cancer Council offers an Information and Support Line. If you need to, please call 13 11 20 or visit their website at https://www.cancer.org.au/about-cancer/patient-support/131120.html. Your local doctor or other health care provider may also be able to offer you advice.

Appendix 4J: Copy of email newsletter by IN FOCUS

Thank you for taking part in the IN FOCUS National Survey of Cancer Survivors. You are receiving this email because you indicated in the survey that you would be interested in learning about opportunities to take part in future research.

We are currently looking for people interested in taking part in a number of studies. Please see below for information on how you can take part.

Online survey: what do you value when being assessed for CIPN?

What is this survey about?

In this survey, we are working with researchers from University of Technology Sydney to focus on understanding what features you most value when being assessed for peripheral neuropathy (tingling and numbness in hands and feet) following chemotherapy treatment.

Why is this survey important?

Your opinion will assist in the development of an assessment tool for peripheral neuropathy that is meaningful to both the patient and the clinician.

How do I participate?

If you are interested, please click on the link below:

https://chere.surveyengine.com/survey/488/567

Sydney-based study: Nerve testing study

What is the purpose of this study?

The purpose of this study is to improve our understanding of nerve problems following chemotherapy treatment, which may lead to better treatment and management of these symptoms.

Sydney-based study: Exercise rehabilitation for CIPN

What is the purpose of this study?

This study looks at the effect of exercise training on balance, mobility, quality of life and nerve function in people who have nerve symptoms which affect their daily functioning after chemotherapy (at least 3 months post-treatment).

What does the study involve?

Our exercise program consists of 8 weeks of 1 hour exercise sessions, 3 times per week. You will be randomly assigned to either an individualized supervised exercise program in a clinic, or an individualised exercise program at home. Each session will involve cardiovascular, resistance/strength, stretching and balance exercises.

What does the study involve?

We are conducting one-off research sessions with people who have received chemotherapy that can damage the nerves, and who completed chemotherapy between 3 months - 5 years ago.

You will be asked to complete a comprehensive nerve assessment including questions relating to nerve symptoms, tests of your hand function, balance and nerve function. The assessments will take around 1.5 - 2 hours to complete.

How do I participate?

If you are interested in participating or would like any further information, please contact Eva Battaglini at in_focus@unsw.edu.au

You will also have assessments of hand function, balance and gait, as well as questionnaires relating to nerve symptoms and how they affect your daily life.

How do I participate?

If you are interested in participating or would like any further information, please contact Kimberley Au at k.au@unsw.edu.au

IN FOCUS

The IN FOCUS research program is working towards increasing our understanding of chemotherapy-induced peripheral neuropathy (CIPN). Our collaborative research program aims to better understand the impact of CIPN, and to find effective assessment and treatment strategies for this condition. For further information, email us at in_focus@unsw.edu.au

Appendix 4K: Copy of the wording of email to relaunch survey on 1st of November 2019

Good afternoon,

We recently launched a survey in order to better understand preferences for how peripheral neuropathy is assessed during chemotherapy treatment.

After launch, we realised that there was a problem with the survey display on some browsers/devices that made it difficult to click next on the survey. We apologise for any inconvenience or frustration this may have caused.

This issue has now been fixed.

We would like to encourage anyone that is interested to give this survey another try; preferably on a computer, laptop or smart tablet.

The link to the survey is here: https://chere.surveyengine.com/survey/488/567

Thank you.

Appendix 4L: Respondent comments

Free text responses
Oncologists should advise this could occur with certain drugs and not wait till you get symptoms 5 years after your chemo
Questionnaires should include areas for patient self assessment, even take home dairy type assessments. Patients will be nervous and not think of all symptoms in a clinical interview. Also, elderly patients do not want to hold the busy doctor up, and may think some things not important enough to mention until they become acute.
Emotional wellbeing
Lose my Balance When Bathing
The level of impact peripheral neuropathy has on daily living
Assessing the health of the lower legs and feet and soles of feet
Greater stress on how to cope with numbness and pins and needles.
Would you stop or change chemotherapy if the PN is permanent ?
take people to childrens playground where air is being pumped through the "floor" and with different air pressures get candidates to walk from A to B across. If they fall over they will not hurt themselves! YOu can measure pressure and give some validity to PN of feet. It would be fun too, lots of laughts! I had tio sell my boat as I couls no longer stand up in it without falling over! With fingers agaiin you need some validity and I think you need to test dexterirty against size. Maybe you could add in time too. What is wrong say doing up a shirt?
Choice questions are too clinical/. You need to invent day to day tests to measure physical constraint of PN in real life.
How to care for these symptoms and ongoing effects in the future, and what to watch out for. And potential for referral for assistance from an excercise physiologist or similar allied health professional
Whilst this is a purely scientific assessment, given the intention of administering this test during a very stressful period of time it may be worth considering the inclusion of a psychological question re potential mental health impact.
Information about peripheral neuropathy potentially worsening after chemotherapy is completed.

Longevity of symptoms
Would you like a family member to be with you at appointment
Mental health issues due to having peripheral neuropathy.
acknowledging that chemo causes PN, my oncologist still doesn't believe it does!
Consider that when people first come out of treatment they are very forcosed on their body. Once they get back itno real life their symptoms dont seem as important. AND, your body does recover to a certain degree. Patients should be aware that their symptoms may not be forever. Easier to handle then.
Why wasn’t there mention of decreasing chemotherapy as well as managing pain or other needs such as Physio?
Ongoing issues – how to deal the condition years after treatment has finished

Appendix 5A: CIPN DCE general population sample copy of survey

UTS HREC REF NO. ETH18-2507 – What features are important to include in an assessment for peripheral neuropathy induced by chemotherapy?

What is the research study about?
This purpose of this research is to help us better understand some aspects of how treatment for cancer is managed. When people have chemotherapy for cancer, they can develop a condition called peripheral neuropathy as a side effect. This can lead to symptoms such as numbness or 'pins and needles' in the hands or feet or problems with balance, leading to increased risk of falls.

Therefore assessment tools are needed to assess whether patients are experiencing peripheral neuropathy and how it affects them. This online survey will help us to understand what features are important to include in an assessment tool for peripheral neuropathy.

As part of the survey, you will be asked to imagine that you have peripheral neuropathy and then we will ask you to consider assessment tools for different peripheral neuropathy, and tell us which you prefer. We will also ask for some information about you to help us to understand your answers – including your age, gender and some questions about your experience of health and health care.

How to complete this survey
This survey ***cannot be completed on a mobile phone*** due to display issues.
Please use other devices such as a PC, laptop or tablet. Thank you.

Who is conducting this research?
My name is Alice Yu and I am a PhD student at the University of Technology Sydney (UTS). My supervisor is Prof. Deborah Street.

How long will this survey take?
We expect the survey to take around 15 minutes to complete.

Do I have to take part in this research study?
Participation in this study is voluntary. You can change your mind at any time and stop completing the survey.

Are there any risks/inconvenience?
We don't expect this questionnaire to cause any distress or inconvenience, however if you do feel distressed you can change your mind at any time and stop completing the survey without consequences.

How do I consent to be part of the study?
Submission of the online questionnaire is an indication of your consent. By continuing to answer the survey questions you consent to the research team collecting and using personal your answers information about you for the research project.

What will happen to my information?
All of your answers will be treated confidentially, and are anonymous. There is no identifying information that is collected as part of the study and the researchers will not have access to any information that can identify you. All data will be stored securely. Data collected will be used for the purpose of this research project and potentially in future research projects that aim to improve on the methods used to design questionnaires of this general type. In all instances, results from the survey will be published in a form that does not identify you.

How will information collected be used?
The results of this study will be shared with clinicians who are interested in developing tools to assess peripheral neuropathy.

What if I have concerns or a complaint?
If you have concerns about the research that you think my supervisor or myself can help you with, contact me on +61 2 9514 4768 or via email: Alice.Yu@chere.uts.edu.au.
Alternatively, you can contact my supervisor Prof. Deborah Street via email: Deborah.Street@uts.edu.au
If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number ETH18-2507.

Do you agree to be part of the research and for the research data gathered from this survey to be published in a form that does not identify you?

Select only one answer

- ○ yes
- ○ no

Please press next to continue to the survey.

next

Please indicate your age group:

Select only one answer

-- select one --

What is your gender?

Select only one answer

- Female
- Male
- Other

What is the highest level of education that you have completed ?

Select only one answer

- No school certificate or other qualifications
- Secondary school
- Trade or apprenticeship
- TAFE or vocational college
- Bachelor's degree
- Postgraduate degree

Introduction to the survey

In this survey, we will ask you to think about peripheral neuropathy, which is a particular side effect to chemotherapy drugs.

In the next few pages, there is some background information provided to help you understand peripheral neuropathy and to explain the terms we use in the survey.

Arm 1 Introduction section

Background Information

What is peripheral neuropathy and how is it related to chemotherapy?

Peripheral neuropathy is a major side effect experienced as a result of chemotherapy treatment, affecting up to 40% of cancer survivors.

Peripheral neuropathy occurs when drugs used to treat cancer cause damage to the peripheral nerves (i.e. nerves in the hands and feet, sometimes extending to the arms and legs).

Some *common symptoms* of peripheral neuropathy can include:

- Numbness
- Tingling, 'pins and needles' or electric shock-like sensations
- Burning sensations
- Balance problems
- Muscle weakness
- Constipation
- Decreased reflexes

These symptoms can lead to problems with completing usual activities such as:

- Trouble using your fingers to pick up or hold things; dropping things
- Trouble with buttoning clothes
- Tripping or stumbling while walking

Assessing for peripheral neuropathy

There are different ways in which a patient may be assessed for peripheral neuropathy while they undergo chemotherapy treatment for cancer.

Please continue to the next page.

Another way in which a peripheral neuropathy assessment can be different is in terms of the type of physical tests involved.

What do we mean by this?

The peripheral neuropathy assessment may involve different types of physical tests on your body.

There are three main types of physical tests:

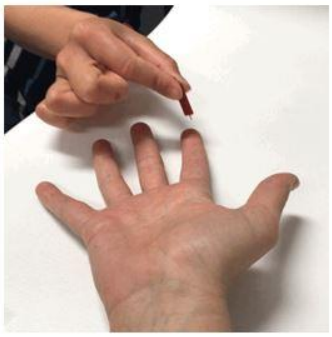
- Clinician administered tests
- Patient activity based tests
- Technical tests

In the following pages, we will show you some examples of each type of physical test.

Examples of clinician administered tests

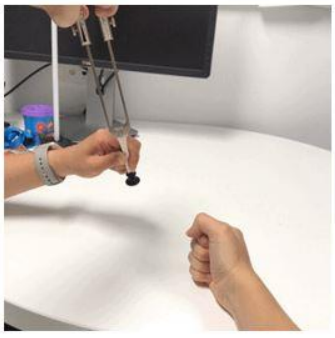
Sharp and dull test

This is a pain perception test where you will be asked to perceive the difference between a sharp or dull stimulus.
An example of a sharp and dull test is given below.



Tuning fork test

This is a vibration sensation test. A tuning fork is placed against different parts of your body and you are asked to say when you feel the vibration stop.
An example of a tuning fork test is given below.



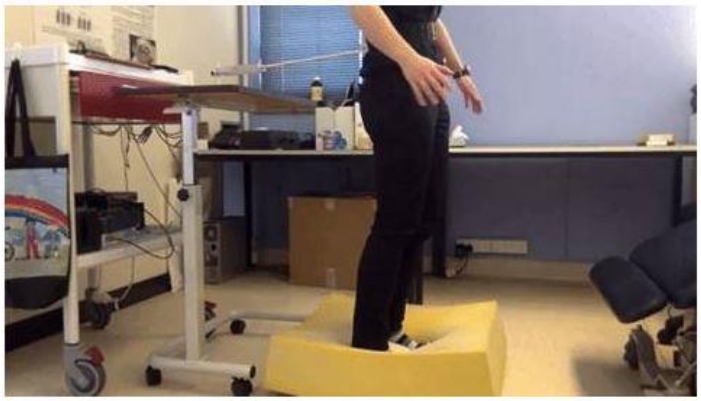
Examples of patient activity based tests

Peg Board Test

This is to test the manual dexterity of the hands.
An example of a peg board test is the grooved pegboard test, which consists of holes that are randomly positioned.
The pegs need to be rotated to match the hole before they can be inserted.
An example of a grooved pegboard test is given below.

Sway Test

An example of a sway test is the Romberg's test, which is a test of the body's sense of positioning.
You will be asked to close your eyes and you will be assessed on your sense of balance.
An example of a sway test is given below.



Example of a technical test

Nerve conduction studies

These tests record the properties of electrical impulses travelling along your nerves using stick-on-electrodes and impulses which feel a little like tapping. An example of a nerve conduction study is given below.

Introduction section for Arm 2

Background information

What is peripheral neuropathy and how is it related to chemotherapy?

Peripheral neuropathy is a major side effect experienced as a result of chemotherapy treatment, affecting up to 40% of cancer survivors.

Peripheral neuropathy occurs when drugs used to treat cancer cause damage to the peripheral nerves (i.e. nerves in the hands and feet, sometimes extending to the arms and legs).

Some *common symptoms* of peripheral neuropathy can include:

- Numbness
- Tingling, 'pins and needles' or electric shock-like sensations
- Burning sensations
- Balance problems
- Muscle weakness
- Constipation
- Decreased reflexes

These symptoms can lead to problems with completing usual activities such as:

- Trouble using your fingers to pick up or hold things; dropping things
- Trouble with buttoning clothes
- Tripping or stumbling while walking

Assessing for peripheral neuropathy

There are different ways in which a patient may be assessed for peripheral neuropathy while they undergo chemotherapy treatment for cancer.

Assessment for peripheral neuropathy can vary in terms of the detail of the assessment.

What do we mean by this?

The peripheral neuropathy assessment could pick up:

- *minor and major* nerve damage, including *small changes* in your condition, whether it is important or not

OR

- only *major* nerve damage and *large changes* in your condition

What is the difference between an assessment that picks up small versus a large change in peripheral neuropathy?

Potential Symptom	Examples of Assessment Detecting Small Change	Examples of Assessment Detecting Large Change
Numbness in hands	The assessment detects when the numbness becomes worse, although you are still able to perform the same tasks e.g. using knives or dressing yourself.	The assessment detects when the numbness that did not interfere with daily activities worsens such that chopping vegetables and carrying pots becomes difficult.
Numbness in feet	The assessment detects when the numbness becomes worse, but it does not affect your ability to walk or maintain balance.	The assessment detects when the numbness in the feet progresses to being unsteady on your feet, especially at night.
Pain/tingling in hands or feet	The assessment detects when increasing pain or tingling occurs.	The assessment detects when pain or tingling that makes cooking difficult worsens to being unable to button clothing or use keys.

On the next page is a short video (1 minute and 35 seconds long).
The video was developed by, and used with the permission of, the Swedish Health Services. More information is provided at the end of this survey.

Please be patient with the video, as it may require time to load.

When you are ready, please press next to view the video.

Please ensure the volume of your device turned on and functioning as expected.

Another way in which a peripheral neuropathy assessment can be different is in terms of the type of physical tests involved.

What do we mean by this?

The peripheral neuropathy assessment may involve different types of physical tests on your body.

There are three main types of physical tests:

- Clinician administered tests
- Patient activity based tests
- Technical tests

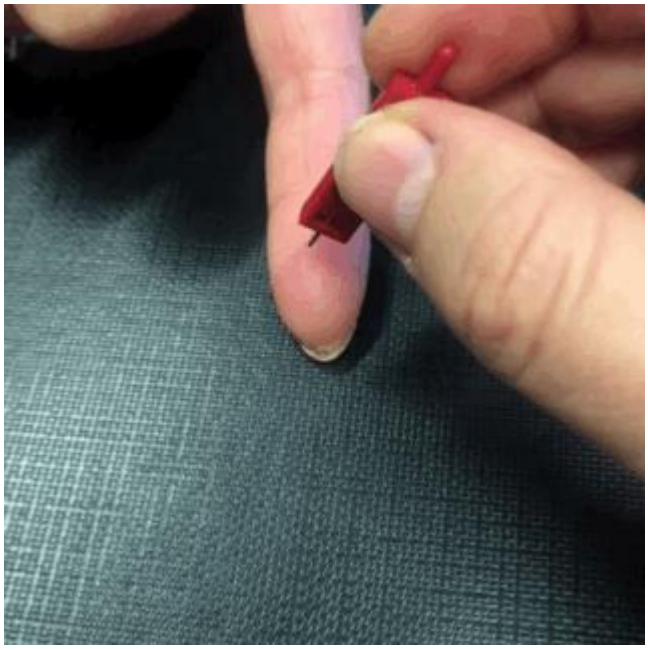
In the following pages, we will show you some examples of each type of physical test.

Examples of clinician administered tests

Sharp and dull test

This is a pain perception test where you will be asked to perceive the difference between a sharp or dull stimulus.
An example of a sharp and dull test is given below.

Pictures shown as moving in survey



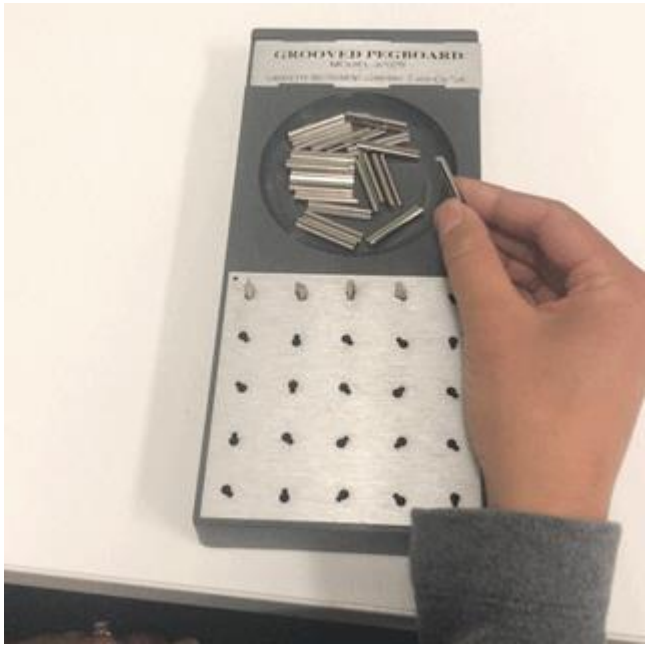
Tuning fork test

This is a vibration sensation test. A tuning fork is placed against different parts of your body and you are asked to say when you feel the vibration stop.
An example of a tuning fork test is given below.

Examples of patient activity based tests

Peg Board Test

This is to test the manual dexterity of the hands.
An example of a peg board test is the grooved pegboard test, which consists of holes that are randomly positioned.
The pegs need to be rotated to match the hole before they can be inserted.
An example of a grooved pegboard test is given below.



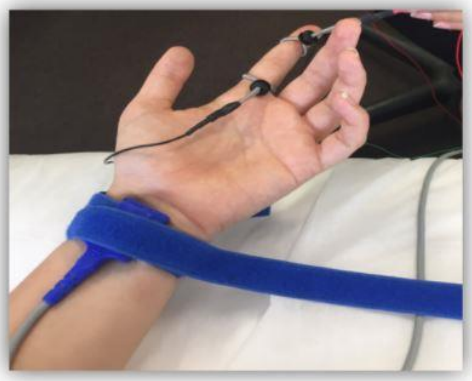
Sway Test

An example of a sway test is the Romberg's test, which is a test of the body's sense of positioning.
You will be asked to close your eyes and you will be assessed on your sense of balance.
An example of a sway test is given below.

Example of a technical test

Nerve conduction studies

These tests record the properties of electrical impulses travelling along your nerves using stick-on-electrodes and impulses which feel a little like tapping. An example of a nerve conduction study is given below.



Other ways in which a peripheral neuropathy assessment can be different include:

- Whether the assessment focuses on symptoms or how symptoms impact on usual activities
- The length of questionnaire that the patient may have to complete
- The impact of assessment on clinic time
- How will results from the assessment influence care/treatment

Choice Questions: Instructions

On the next few pages, you will be asked to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect.

You will then be presented information describing the features of two potential peripheral neuropathy assessment tools.

You will be asked to decide which assessment tool you prefer.

You will be asked this question 8 times. For each of the 8 questions, the values for the features may differ e.g. the number of questions on the questionnaire may be different.

These questions are hypothetical. There are no right or wrong answers to these questions; we are simply interested in your views.

When you are ready, please press next to continue to the choice questions.

Example choice set, respondents each randomly assigned 8 choice sets

We would like you to imagine that you have been given a chemotherapy drug where peripheral neuropathy is a known side effect. Imagine that during the time you are undergoing chemotherapy, you will also be assessed for peripheral neuropathy. In addition to talking with your clinician, you may be asked to complete a questionnaire prior to your appointment and/or undergo some physical test/s during your appointment.

If these were your only options, which peripheral neuropathy assessment tool would you choose? A or B?

	Assessment A	Assessment B
Symptoms and usual activities	The assessment asks about how your *symptoms impact on your usual activities*	The assessment asks about how your *symptoms impact on your usual activities*
Level of detail	The assessment will pick up *minor and major nerve damage,* including *small changes* in your condition whether it is important or not	The assessment will *only* pick up *major nerve damage* and *large changes* in your condition
Questionnaire	No questionnaire	12 questions to answer
Physical test/s	Patient activity based test e.g. peg board test, sway test	Technical test e.g. nerve conduction studies
Impact on clinic time	During usual clinic time	Usual clinic time plus 30 minutes extra
How will results influence care/treatment	The *doctor may change your general care* (e.g. medications to help relieve symptoms, physiotherapy, walking aids) if there are significant changes in your condition over time	The *doctor may change your chemotherapy/cancer treatment* if there are significant changes in your condition over time
Which would you choose?	○ Assessment A	○ Assessment B

The next few questions are about how you found answering the choice questions

Prior to the choice questions, some information was provided to define peripheral neuropathy and to explain some key definitions and terms.

Did this information help you understand the choice questions?

Select only one answer

- ○ Yes
- ○ No
- ○ No, I already knew this information prior to this survey
- ○ No, it was confusing
- ○ I don't remember
- ○ Other. Please specify: [ ]

Questions only for Arm 2

The background information also included a short video. Were you able to view the full video?

Select only one answer

- ○ yes
- ○ no

If no, asked:

Why could you not view the full length of the video?

Select only one answer

- ○ Video was unable to load
- ○ Video was laggy
- ○ No sound
- ○ Other. Please specify: _____

If yes, asked:

How would you rate the quality of the video?

Select only one answer

very poor	poor	okay	good	excellent
○	○	○	○	○

Link back to questions for all respondents

For each of the following types of physical tests, would you agree that the information provided prior to the choice questions helped you understand what they were?

Select one response from each row

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
sharp and dull test	○	○	○	○	○
tuning fork test	○	○	○	○	○
peg board	○	○	○	○	○
sway test	○	○	○	○	○
nerve conduction study	○	○	○	○	○

In the context of peripheral neuropathy, I could imagine what it would feel like to live with:

Select one response from each row

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
numbness	○	○	○	○	○
pins and needles	○	○	○	○	○
burning sensations	○	○	○	○	○
balance problems	○	○	○	○	○
muscle weakness	○	○	○	○	○
constipation	○	○	○	○	○
decreased reflexes	○	○	○	○	○

In the context of peripheral neuropathy, I could imagine what it would feel like to live with:

Select one response from each row

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
trouble picking up/holding things	○	○	○	○	○
trouble buttoning clothes	○	○	○	○	○
trouble while walking e.g. tripping or stumbling	○	○	○	○	○

How did you decide whether Assessment A or B was better in each choice question?

Select only one answer

- ○ I considered *all features* for each assessment option
- ○ I considered *all features which were different* between the two assessment options
- ○ I considered *only the features* which were most important to me
- ○ Other strategy (please explain):

Thinking about the features of a peripheral neuropathy assessment tool. Please select which feature is the *most important* and *least important* to you:

	Whether the assessment focuses on symptoms or how symptoms impact on usual activities	Level of detail of the assessment	Length of questionnaire	Type of physical test/s	Impact of assessment on clinic time	How will results influence care/treatment
MOST Important to you	○	○	○	○	○	○
LEAST Important to you	○	○	○	○	○	○

If selected cancer, then asked:

The next few questions are some background questions about you.

Have you ever suffered from any of the following conditions (if more than one, please select your primary condition):

Select only one answer

- ○ Asthma
- ○ Arthritis
- ○ Cancer
- ○ Cardiovascular disease
- ○ Diabetes and high sugar levels
- ○ Kidney disease
- ○ Osteoporosis
- ○ Mental, behavioural and cognitive conditions
- ○ None of these

What type of cancer/s have you been diagnosed with?

Select all that apply

☐ Colon/rectum/bowel cancer (colorectal)	☐ Breast cancer	☐ Prostate cancer
☐ Lung cancer (including trachea, pleura and bronchus)	☐ Cervical cancer	☐ Cancer of other female reproductive organs (including uterus, ovary)
☐ Bladder/kidney cancer	☐ Stomach cancer	☐ Leukaemia
☐ Non-Hodgkin lymphoma	☐ Other type of lymphoma	☐ Cancer of unknown primary site
☐ Other cancer (please specify):		

Please indicate the extent to which you agree or disagree with each of the following statements about the choice questions

Select one response from each row

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
Most of the time, I could easily *identify the differences* between assessment options	○	○	○	○	○
Most of the time, I could easily *choose between* the assessment options	○	○	○	○	○

Have you ever been diagnosed with *skin* cancer?

Select only one answer

- ○ yes
- ○ no

If yes, asked:

What type of skin cancer have you been diagnosed with? Please choose one or more.

Select all that apply

- ☐ Melanoma
- ☐ Basal cell carcinoma (BCC)
- ☐ I don't know
- ☐ Squamous cell carcinoma (SCC)
- ☐ Other form of skin cancer: _____

Have you ever been diagnosed with other types of cancer?

Select only one answer

- ○ yes
- ○ no

What other types of cancer have you been diagnosed with?

Select all that apply

☐ Colon/rectum/bowel cancer (colorectal)	☐ Breast cancer	☐ Prostate cancer
☐ Lung cancer (including trachea, pleura and bronchus)	☐ Cervical cancer	☐ Cancer of other female reproductive organs (including uterus, ovary)
☐ Bladder/kidney cancer	☐ Stomach cancer	☐ Leukaemia
☐ Non-Hodgkin lymphoma	☐ Other type of lymphoma	☐ Cancer of unknown primary site
☐ Other cancer (please specify): _____		

If not, asked:

What type of cancer/s have you been diagnosed with?

Select all that apply

☐ Colon/rectum/bowel cancer (colorectal)	☐ Breast cancer	☐ Prostate cancer
☐ Lung cancer (including trachea, pleura and bronchus)	☐ Cervical cancer	☐ Cancer of other female reproductive organs (including uterus, ovary)
☐ Bladder/kidney cancer	☐ Stomach cancer	☐ Leukaemia
☐ Non-Hodgkin lymphoma	☐ Other type of lymphoma	☐ Cancer of unknown primary site
☐ Other cancer (please specify):		

Link back to next question:

Has someone close to you ever received a diagnosis for cancer?

Select only one answer

- ○ yes
- ○ no

If yes, asked:

Has someone close to you ever been diagnosed with *skin* cancer?

Select only one answer

- ○ yes
- ○ no

What type of skin cancer have they been diagnosed with? Please choose one or more.

Select all that apply

- ☐ Melanoma
- ☐ Basal cell carcinoma (BCC)
- ☐ I don't know
- ☐ Squamous cell carcinoma (SCC)
- ☐ Other form of skin cancer:

Have they ever been diagnosed with other types of cancer?

Select only one answer

- ○ yes
- ○ no

If no, asked:

What type of cancer/s have they been diagnosed with?

Select all that apply

☐ Colon/rectum/bowel cancer (colorectal)	☐ Breast cancer	☐ Prostate cancer
☐ Lung cancer (including trachea, pleura and bronchus)	☐ Cervical cancer	☐ Cancer of other female reproductive organs (including uterus, ovary)
☐ Bladder/kidney cancer	☐ Stomach cancer	☐ Leukaemia
☐ Non-Hodgkin lymphoma	☐ Other type of lymphoma	☐ Cancer of unknown primary site
☐ Other cancer (please specify):		

Link back to final question:

If you have any other feedback or comments, please enter it below.

Enter text below

Thank You! Please Press 'Submit Answers And Finish' To End The Survey.

Contact Details

If you have concerns about the research that you think I or my supervisor can help you with, please feel free to contact me (us) on +61 2 9514 4768 or via email: Alice.Yu@chere.uts.edu.au

If you would like to talk to someone who is not connected with the research, you may contact the Research Ethics Officer on 02 9514 9772 or Research.ethics@uts.edu.au and quote this number ETH18-2507).

Video used in the survey

The video extract presented in this questionnaire was developed by, and used with the permission of, the Swedish Health Services (https://www.swedish.org/).
The full video can be seen at (https://www.youtube.com/watch?v=Zki82j3SmMo).

submit answers and finish

Appendix 5B: CIPN DCE general population sample ethics approval (copy of letter)

2nd July

Dear Alice

RE: General population preferences for the features of an assessment tool for CIPN

At the meeting on 2nd July, the CHERE management team agreed that this research project is appropriate to be conducted under CHERE's program ethics approval from the UTS Human Research Ethics Committee (UTS HREC REF NO. ETH18-2507).

Yours sincerely

Production Note:
Signature removed
prior to publication.

Jane Hall
Director of Strategy

Appendix 5C: Excluded respondents comment box text

ID	Comment Box Text
122842	bvj,bvjgbvmjgbnvjhnb
123341	dftgshydg dfgsthgdg dfghs dfgsh dgjsh
123611	dghdf dghsdg dfghs fghsgd dghsxdg hjxh s
124205	dhsgyewysbjsgd e7aygdyasgdhuñ cxcjkfyogsazki hhxszdosihf
124353	iuyad eiufh oirehf

Appendix 5D: Sensitivity analysis: MNL model results with fastest 10% removed by arm

	Arm 1	**Arm 2**
Log likelihood	−779	−773
Obs	1200	1208
Iterations	4	4
AIC	1584	1573
BIC	1650	1639

MNL by arm	Arm 1		Arm 2	
	Estimate	P-value	Estimate	P-value
S&Q 2 (symptoms & usual activities)	−0.053	0.539	0.035	0.681
Det 2 (minor and major changes)	0.535	0.000***	0.722	0.000***
Q 2 (3 questions to answer)	−0.104	0.409	0.183	0.144
Q 3 (12 questions to answer)	−0.055	0.626	0.256	0.028*
Q 4 (20 questions to answer)	−0.022	0.862	0.238	0.058
PhyT 2 (clinician administered test)	0.463	0.000***	0.547	0.000***
PhyT 3 (patient activity based test)	0.547	0.000***	0.535	0.000***
PhyT 4 (technical test)	0.414	0.000***	0.564	0.000***
CT 2 (usual clinic time + 10 mins)	0.004	0.968	0.028	0.797
CT 3 (usual clinic time + 30 mins)	−0.063	0.537	−0.105	0.303
CT 4 (separate appointment, takes up to 60 mins)	−0.35	0.002**	−0.187	0.09
Res 2 (doctor may change your general care)	−0.224	0.007**	−0.169	0.043*
Res 3 (doctor may change your chemo/cancer treatment)	−0.315	0.000***	−0.28	0.001**

*** p < 0.001; **p < 0.01; *p < 0.05

Appendix 5E: Sensitivity analysis: Arm 2 with 4 respondents that could not view the video removed

Log likelihood	−850
Obs	1312
Iterations	4
AIC	1727
BIC	1794

MNL	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	0.033	0.080	0.683
Det 2 (minor and major changes)	0.663	0.086	0.000***
Q 2 (3 questions to answer)	0.085	0.119	0.477
Q 3 (12 questions to answer)	0.117	0.110	0.285
Q 4 (20 questions to answer)	0.061	0.120	0.611
PhyT 2 (clinician administered test)	0.516	0.110	0.000***
PhyT 3 (patient activity based test)	0.421	0.097	0.000***
PhyT 4 (technical test)	0.573	0.111	0.000***
CT 2 (usual clinic time + 10 mins)	0.042	0.104	0.685
CT 3 (usual clinic time + 30 mins)	−0.052	0.096	0.589
CT 4 (separate appointment, takes up to 60 mins)	−0.156	0.106	0.139
Res 2 (doctor may change your general care)	−0.170	0.080	0.034*
Res 3 (doctor may change your chemo/cancer treatment)	−0.240	0.080	0.003**

*** p < 0.001; **p < 0.01; *p < 0.05

Appendix 5F: S-MNL results general population arms combined

Log likelihood	−2290
Obs	3616
Iterations	35
Draws	10000
AIC	4618
BIC	4711

S-MNL Model	Estimate	S.E.	P-value
S&Q 2 (symptoms & usual activities)	0.217	0.083	0.009**
Det 2 (minor and major changes)	1.229	0.172	0.000***
Q 2 (3 questions to answer)	0.11	0.116	0.342
Q 3 (12 questions to answer)	0.23	0.116	0.048*
Q 4 (20 questions to answer)	0.152	0.128	0.233
PhyT 2 (clinician administered test)	0.889	0.167	0.000***
PhyT 3 (patient activity based test)	0.869	0.134	0.000***
PhyT 4 (technical test)	0.903	0.155	0.000***
CT 2 (usual clinic time + 10 mins)	0.11	0.109	0.312
CT 3 (usual clinic time + 30 mins)	−0.029	0.106	0.784
CT 4 (separate appointment, takes up to 60 mins)	−0.304	0.107	0.005**
Res 2 (doctor may change your general care)	−0.556	0.104	0.000***
Res 3 (doctor may change your chemo/cancer treatment)	−0.626	0.108	0.000***
τ	0.993	0.161	0.000***
$\delta_{general\ population\ sample}$	−0.526	0.138	0.000***

*** p < 0.001; **p < 0.01; *p < 0.05

Appendix 6A: Durations used in choice sets

In this section, the duration combinations used for health states A and B, and for health states B and C (where applicable) have been summarised by arm. Duration combinations for the health states were the same in the Peruvian and Danish data sets Durations used in choice sets are in years (e.g. 0.25 years = 3 months).

Appendix Table 1 Arm death_arm duration combinations for health state A and B

Duration	Duration	N	%
1	4	25	20%
4	7	25	20%
7	10	25	20%
10	15	25	20%
15	1	25	20%

Note: Duration is in years

Appendix Table 2 Arm fullhealth_arm duration combinations for health states A and B

Duration	Duration	N	%
1	1	19	18.1%
4	4	23	21.9%
7	7	19	18.1%
10	10	21	20.0%
15	15	23	21.9%

Note: Duration is in years

Appendix Table 3 Arm fullhealth_arm frequency of durations for health states B vs C (full health)

Duration B	Duration C	N	%
1	0.25	8	7.6%
1	0.5	11	10.5%
4	0.5	13	12.4%
4	1	10	9.5%
7	0.5	10	9.5%
7	1	7	6.7%
7	4	2	1.9%
10	1	6	5.7%
10	4	9	8.6%
10	7	6	5.7%
15	1	4	3.8%
15	4	6	5.7%
15	7	4	3.8%
15	10	9	8.6%

Note: Duration is in years

Appendix 6B: Test for poolability of arms by country

Level	Combined death_arm	Combined fullhealth_arm
MO2	−0.024	0.047*
MO3	−0.042	−0.139***
MO4	−0.116***	−0.151***
MO5	−0.251***	−0.348***
SC2	−0.022	−0.001
SC3	−0.05*	−0.017
SC4	−0.138***	−0.198***
SC5	−0.22***	−0.278***
UA2	0.024	−0.053**
UA3	−0.141***	−0.11***
UA4	−0.22***	−0.268***
UA5	−0.31***	−0.416***
PD2	0.014	−0.077***
PD3	−0.059*	−0.127***
PD4	−0.269***	−0.423***
PD5	−0.495***	−0.653***
AD2	0.019	−0.027
AD3	−0.015	−0.146***
AD4	−0.273***	−0.411***
AD5	−0.433***	−0.682***
τ	0.865***	2.824***
$\delta_{dummy\_country}$	0.49***	2.172***

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute

Appendix 6C: MNL parameter estimates on latent scale

Appendix Table 4 MNL parameter estimates on latent scale

MNL results on latent scale	Peru		Denmark	
	death_arm	fullhealth_arm	death_arm	fullhealth_arm
Duration	0.249***	0.258***	0.389***	0.498***
MO2 x Duration	−0.005	−0.001	−0.011	0.021
MO3 x Duration	−0.006	−0.031***	−0.033*	−0.069***
MO4 x Duration	−0.043***	−0.047***	−0.066***	−0.067***
MO5 x Duration	−0.115***	−0.149***	−0.09***	−0.118***
SC2 x Duration	−0.001	−0.01	−0.005	−0.028*
SC3 x Duration	−0.015	−0.003	−0.024*	−0.033*
SC4 x Duration	−0.063***	−0.048***	−0.044***	−0.092***
SC5 x Duration	−0.084***	−0.1***	−0.073***	−0.126***
UA2 x Duration	−0.002	−0.019*	0.001	−0.008
UA3 x Duration	−0.03**	−0.023*	−0.047***	−0.04***
UA4 x Duration	−0.074***	−0.085***	−0.089***	−0.103***
UA5 x Duration	−0.118***	−0.157***	−0.106***	−0.15***
PD2 x Duration	−0.009	−0.017*	0.013	−0.039***
PD3 x Duration	−0.013	−0.02*	−0.012	−0.063***
PD4 x Duration	−0.054***	−0.083***	−0.129***	−0.194***
PD5 x Duration	−0.13***	−0.154***	−0.229***	−0.283***
AD2 x Duration	−0.001	0.004	0.008	−0.005
AD3 x Duration	0.008	−0.025**	−0.012	−0.058***
AD4 x Duration	−0.033***	−0.047***	−0.148***	−0.206***
AD5 x Duration	−0.078***	−0.101***	−0.22***	−0.31***

*p < 0.05; ** p < 0.01; *** p < 0.001 Note: base (omitted level) is level 1 of each attribute