

THE IMPACT OF MEDICAL DEVICE USABILITY AND ACCESSIBILITY INFORMATION ON  
PURCHASING DECISIONS OF PEOPLE WITHOUT DISABILITIES

by

Maysam M. Ardehali

A Dissertation Submitted in  
Partial Fulfillment of the  
Requirements for the Degree of

Doctor of Philosophy  
in Health Sciences

at

The University of Wisconsin-Milwaukee

December 2023

## ABSTRACT

### THE IMPACT OF MEDICAL DEVICE USABILITY AND ACCESSIBILITY INFORMATION ON PURCHASING DECISIONS OF PEOPLE WITHOUT DISABILITIES

by

Maysam M. Ardehali

The University of Wisconsin-Milwaukee, 2023  
Under the Supervision of Professor Roger O. Smith

The disparity between people with and without disabilities in timely diagnosis and appropriate treatment is exacerbated by inaccessible medical devices. The industry appears reluctant to address the shortcomings or provide usability and accessibility information to enhance the purchasing decisions of people with disabilities. This reluctance is mainly caused by the lack of market motivation based on perceptions of market size and the profitability of investments in accessibility evaluation and improvement. This study aims to challenge this notion by demonstrating that usability and accessibility information do not solely benefit what may be considered a “niche market,” and the purchasing decisions of people without disabilities may also hinge on medical device usability and accessibility information. A total of 194 people without disabilities participated in this study, and their preferences in purchasing blood pressure monitors were examined through a Discrete Choice Experiment. The findings revealed that usability and accessibility information not only appeal to a broader audience than only people with disabilities but also significantly impact the likelihood of purchase for people with no disabilities. These insights have the potential to provide the additional market motivation necessary to instigate a paradigm shift in the industry, thereby curtailing the

differential in health outcomes between people with and without disabilities and fostering a more equitable healthcare landscape for all.

*Keywords:* Medical Devices, Accessibility, Usability, Information, Decision Analysis, Stated Preferences, Discrete Choice Experiment

*To my beloved family,  
dearest friends,  
and those who brave the darkness  
to brighten the lives of others.*

© Copyright by Maysam M. Ardehali, 2023  
All Rights Reserved

# TABLE OF CONTENTS

<b>Abstract .....</b>	<b>ii</b>
<b>List of Figures .....</b>	<b>xiii</b>
<b>List of Tables .....</b>	<b>xv</b>
<b>List of Abbreviations.....</b>	<b>xvi</b>
<b>Acknowledgments.....</b>	<b>xvii</b>
<b>I. Introduction and Background .....</b>	<b>1</b>
Statement of the Problem .....	2
Definition of Accessibility and Usability .....	2
Disability and Medical Device Market Statistics .....	5
Academic and Regulatory Efforts .....	6
Accessible Medical Devices or Medical Device Accessibility Information .....	9
Industry Adoption of Accessible Design Principles .....	11
Approaches to Examining Preferences.....	12
Purpose of the Study .....	14
Significance of the Study .....	15
<b>II. Prior Discoveries.....</b>	<b>18</b>
Overview .....	18
Study 1 .....	19
Study Design .....	19

Data Collection .....	20
Data Analysis .....	20
Results .....	21
Study 2 .....	28
Study Design .....	29
Data Collection .....	29
Data Analysis .....	30
Results .....	30
Collective Discussion and Emergent Needs from Prior Research .....	35
<b>III. Methods.....</b>	<b>37</b>
Theoretical Foundation and Conceptual Congruence.....	39
Luce’s Choice Axiom .....	39
Random Utility Theory.....	41
Multi-Attribute Utility Theory .....	41
Quantitative Experimental Design Principles.....	42
Marginal Rates of Substitution and Tradeoffs .....	42
Willingness to Pay and Market Simulation.....	42
Critical Prerequisite .....	43
Research Questions .....	44
Research Design.....	45
DCE Design.....	45
Study Phases & Variables .....	77
Data Collection.....	80
Sample .....	80

Instrumentation.....	86
Data Analysis.....	96
Fraud Detection and Data Quality Assessment.....	96
Analysis Preparation and Internal Validity Evaluation .....	106
Multinomial Logit Analysis .....	108
Methods Summary Review.....	109
<b>IV. Results .....</b>	<b>113</b>
Feasibility Study.....	114
Objectives .....	114
Recruitment Characteristics .....	114
Design Report .....	115
Survey Results.....	115
Feedback Results and Potential Adjustment Needs .....	116
Phase I.....	117
Objectives .....	117
Design Report and Recruitment .....	117
Fraud Detection and Data Quality Assessment.....	118
Sociodemographic Characteristics .....	119
Potential for Context Resonance.....	119
Internal Validity Assessment .....	122
Multinomial Logit Analysis of Preferences.....	122
Participant Burden and Feedback .....	124
Potential Adjustment Needs .....	125
Phase II.....	126

Objectives .....	126
Design Report and Recruitment .....	126
Fraud Detection and Data Quality Assessment.....	127
Sociodemographic Characteristics .....	129
Potential for Context Resonance.....	130
Internal Validity Assessment .....	132
Preference Model Estimation.....	132
Participant Burden Evaluation.....	139
Key Findings Summary.....	141
Feasibility Study.....	141
Phase I.....	141
Phase II.....	142
<b>V. Discussion .....</b>	<b>145</b>
Outcomes Interpretation.....	146
Utilities and Tradeoffs .....	146
Sociodemographic Characteristics and Familiarity with Disability .....	148
Usability/Accessibility Score - Perceptions vs. Observation .....	152
Contextualization of Discoveries .....	155
Implications.....	156
Limitations .....	159
Experiment Design.....	159
On Blue Buses, Red Buses, and Homogeneity of Preferences .....	160
Scope .....	161
Future Directions .....	161

Multifaceted Accessibility and Usability Valuation Effect.....	161
<b>VI. Conclusion.....</b>	<b>164</b>
Personal Reflection .....	166
<b>VII. References .....</b>	<b>168</b>
<b>VII. Appendices .....</b>	<b>203</b>
Appendix A: Syntax – Choice Matrix Design in R.....	203
Appendix B: Syntax – Preparation for Survey Implementation in Python .....	207
Appendix C: Design – $D_z$ -Efficient Design for Feasibility Study .....	211
Appendix D: Design – $D_z$ -Efficient Design for Phase I .....	214
Appendix E: Design – $D_b$ -Efficient Design for Phase II .....	217
Appendix F: Phase I Survey .....	220
Appendix G: Phase II Survey .....	229
Appendix H: Syntax – Fraud Detection in Python .....	239
Appendix I: Syntax – Internal Validity Report & Preprocessing in Python .....	274
Appendix J: Syntax – Multinomial Logit Analysis in R.....	285
Appendix K: Equivalent Text Descriptions (EqTDs).....	288
Explanation of EqTDs.....	288
Figure 1 .....	289
Figure 2 .....	291
Figure 3 .....	292
Figure 4 .....	293
Figure 5 .....	295
Figure 6 .....	297

Figure 7 .....	298
Figure 8 .....	299
Figure 9 .....	302
Figure 10 .....	303
Figure 11 .....	305
Figure 12 .....	307
Figure 13 .....	309
Figure 14 .....	310
Figure 15 .....	312
Figure 16 .....	313
Figure 17 .....	314
Figure 18 .....	316
Figure 19 .....	318
Figure 20 .....	320
Figure 21 .....	321
Figure 22 .....	323
Figure 23 .....	326
Figure 24 .....	327
Table 1 .....	329
Table 2 .....	330
Table 3 .....	331
Table 4 .....	332
Table 5 .....	333
Table 6 .....	334

Table 7 .....	335
Table 8 .....	336
Table 9 .....	337
Table 10 .....	338
Table 11 .....	339
Table 12 .....	340
Table 13 .....	341
Table 14 .....	342
Table 15 .....	343

## LIST OF FIGURES

<b>Figure 1:</b> The Significance of This Study in Reducing Health Outcomes Disparities Between PwD and PwoD .....	16
<b>Figure 2:</b> Commonness of Medical Devices Used at Home .....	23
<b>Figure 3:</b> Commonness of Healthcare Facility Medical Equipment Categories .....	25
<b>Figure 4:</b> Stages of DCE Design .....	46
<b>Figure 5:</b> Most Prevalent Attributes of BPMs Across Four E-tailers.....	51
<b>Figure 6:</b> Variations of BPM Cost Across Four E-tailers .....	52
<b>Figure 7:</b> Cumulative Distribution of BPM User Ratings Across Four E-tailers.....	53
<b>Figure 8:</b> Comparison Between Total Eligible Respondents' sociodemographic Information from Prolific and the US Population Characteristics .....	86
<b>Figure 9:</b> Example of Manual Review of Data Quality After Automatic Fraud Detection .....	105
<b>Figure 10:</b> Example of Data Transformation from Wide to Long Format with the Addition of Choice Context.....	108
<b>Figure 11:</b> Summary of Data Collection and Analysis Pipeline.....	112
<b>Figure 12:</b> Parameter SE Simulation for Increments of Sample Size in Phase I .....	118
<b>Figure 13:</b> Phase I Violations Summary After Automatic and Manual Inspection .....	119
<b>Figure 14:</b> Sociodemographic Data and Geographical Distribution of Participants in Phase I ..	120
<b>Figure 15:</b> Distribution of Participants' Priority Ranking of the Study Attributes in Phase I.....	121
<b>Figure 16:</b> Time Requirement of Phase I Survey .....	124
<b>Figure 17:</b> Parameter SE Simulation for Increments of Sample Size in Phase II .....	127
<b>Figure 18:</b> Phase II Violations Summary After Automatic and Manual Inspection .....	128

<b>Figure 19:</b> Sociodemographic Data and Geographical Distribution of Participants in Phase II .	130
<b>Figure 20:</b> Distribution of Participant’s Priority Ranking for the Study Attributes in Phase II ...	131
<b>Figure 21:</b> Six-Class LC-MNL Parameter Estimate Variations Per Class .....	137
<b>Figure 22:</b> Survey Data that Differed Between Classes with Statistical Significance .....	138
<b>Figure 23:</b> Time Requirement of Phase II Survey.....	140
<b>Figure 24:</b> The MAUVE Hypothesis.....	163

## LIST OF TABLES

<b>Table 1:</b> Summary of Stated Preference Methods Comparison.....	14
<b>Table 2:</b> Number of BPMs found and catalogued from 4 e-tailer websites.....	50
<b>Table 3:</b> Summary of Dissertation Study Attributes and Levels.....	55
<b>Table 4:</b> Attribute Level Balance of Feasibility Study Choice Matrix.....	115
<b>Table 5:</b> Attribute Level Balance of Phase I Choice Matrix.....	117
<b>Table 6:</b> Phase I Preference Model Parameter Estimates.....	122
<b>Table 7:</b> Phase I MNL Model Performance Indicators.....	123
<b>Table 8:</b> Phase I Parameter Estimates Variance-Covariance Matrix.....	123
<b>Table 9:</b> Attribute Level Balance of Phase II Choice Matrix.....	127
<b>Table 10:</b> Phase II Preference Model Parameter Estimates.....	132
<b>Table 11:</b> Phase II MNL Model Performance Indicators.....	133
<b>Table 12:</b> Phase I Parameter Estimates Variance-Covariance Matrix.....	133
<b>Table 13:</b> Gradual Improvement of the LC-MNL Model Performance as the Number of Classes Increases.....	135
<b>Table 14:</b> Six-Class LC-MNL Model Parameter Estimates.....	136
<b>Table 15:</b> Usability and Accessibility Score Parameter Estimate Comparison Between PwD and PwoD.....	156

## LIST OF ABBREVIATIONS

PwoD	People without Disabilities
PwD	People with Disabilities
ISO	International Organization for Standardization
ACA	Affordable Care Act
RP	Revealed Preference
SP	Stated Preference
DCE	Discrete Choice Experiment
ADHD	Attention-Deficit/Hyperactivity Disorder
LCA	Luce's Choice Axiom
IIA	Independence of Irrelevant Alternatives
RUT	Random Utility Theory
MAU	Multi-Attribute Utility
WTP	Willingness to Pay
MRS	Marginal Rate of Substitution
BPM	Blood Pressure Monitor
SUS	System Usability Scale
MNL	Multinomial Logit
MLE	Maximum Likelihood Estimation
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
IID	Independent and Identical Distribution
R <sub>2</sub> D <sub>2</sub>	Rehabilitation Research Design & Disability
UWM	University of Wisconsin-Milwaukee
IQR	Interquartile Range
MXL	Mixed Logit
LC-MNL	Latent Class Multinomial Logit

## ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor and mentor, Dr. Roger O. Smith, for his unwavering support, wisdom, and guidance. Without his incessant enthusiasm to explore, disregarding all "*hic sunt dracones*," no part of this journey would have been possible.

I am deeply grateful to the members of my dissertation committee, Dr. Remi Daviet, Dr. Rochelle J. Mendonca, and Dr. Bhagwant S. Sindhu, for their continuous engagement and invaluable feedback. I treaded confidently into the unknown only because I relied on their support.

My sincere appreciation goes to the University of Wisconsin-Milwaukee, College of Health Professions & Sciences, and specifically the School of Rehabilitation Sciences & Technology for their educational and financial support.

I am deeply indebted to my friends for sharing their love and kindness, joining me in celebrating the happiest times, and lending a supportive hand in the darkest. I am also grateful to the staff and students at the R<sub>2</sub>D<sub>2</sub> Center who directly or indirectly supported this research and those with whom I shared many joyous moments.

Lastly and most importantly, the limitless and unconditional support from my family is what made it all possible. My father's wisdom and guidance, my mother's empathy and kindness, and my sisters' support and perseverance elevate and guide me through life. To them, I owe a debt of gratitude I cannot possibly repay. Thank you for all your love, encouragement, empathy, and sacrifices.

## I. Introduction and Background

This study aimed to investigate the impact of medical device accessibility information availability on the purchasing decisions of People without Disabilities (PwoD). PwoD may not be conventionally considered the target audience for accessibility information but may ultimately benefit from it. In the context of this study, accessibility information is defined as “objective expert evaluation of usability and accessibility features.” For example, accessibility information for an accessible thermometer may be presented as a report that mentions features such as a large display, backlit display, auditory readout, step-by-step setup tutorial video, minimal setup, minimal maintenance, and automatic calibration. This information may also be presented as a score on a scale of 0 to 10, with 0 indicating a complete lack of accessibility and usability features and 10 indicating the inclusion of all reasonable accessibility and usability features. The features mentioned in the example above can assist People with Disabilities (PwD). However, PwoD may also receive additional value from these features, either constantly or circumstantially. This study empirically examines the preferences of PwoD while making medical device purchasing decisions based on the tradeoffs between Cost, Brand, User Ratings, and Usability/Accessibility information.

Chapter I identifies the problem through a literature review and describes the purpose and significance of this study. Chapter II offers an overview of prior scholarly investigations and the emergent needs leading to the hypotheses and research questions of the current study. Chapter III provides an in-depth discussion of the research design and methodology appropriate to examine the study hypotheses. Chapter IV presents the findings of this study in direct evaluation of the hypotheses, as well as exploratory analyses. Chapter V discusses discoveries,

their implications, and potential limitations. Ultimately, Chapter VI concludes the findings of this study and identifies possible directions for future investigation.

### **Statement of the Problem**

At its core, this study aims to answer the question: “Does providing usability and accessibility information for medical products impact the purchasing decisions of People without Disabilities?” This question should not be confused with “What medical device accessibility features are important to People without Disabilities?” or “What are the most important medical product features for People without Disabilities?” The main objective of this study is to establish whether usability and accessibility information can tangibly impact PwD's purchasing decisions. This chapter will discuss the importance of answering this question and the history of academic and regulatory efforts addressing the issues arising from the lack of medical device accessibility.

### ***Definition of Accessibility and Usability***

Two terms are commonly used in tandem throughout this dissertation: accessibility and usability. Sometimes presented as “Usability/Accessibility,” suggesting interchangeability, it becomes crucial to define these terms, identify their connections, and determine the operational definition of “Usability/Accessibility” in the context of this study. The International Organization for Standardization (ISO) defines accessibility as “usability of a product, service, environment or facility by people with the widest range of capabilities” (*ISO 9241-171*, 2008). Usability, on the other hand, is defined by ISO as “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (*ISO 9241-11*, 2018). Based on these definitions,

accessibility can be conceptualized as usability for all individuals regardless of their ability levels (Petrie & Bevan, 2009). In this interpretation, usability is viewed as a subset of accessibility. Interestingly, this interpretation coincides with the definition of Universal Design (Vanderheiden & Tobias, 1998).

Using different definitions for these terms, some sources have argued that accessibility is objective by nature and a prerequisite for usability, which is more subjective and individualistic (Iwarsson & Ståhl, 2003; Theofanos & Redish, 2003). In this interpretation, accessibility is viewed as a precondition for usability. Other sources have defined accessibility as usability for PwD, explicitly positioning disability as the differentiator between usability and accessibility (Powlik & Karshmer, 2002; Usability.gov, 2013; W3C Web Accessibility Initiative, 2022). In this interpretation, accessibility and usability evaluate the same construct for two mutually exclusive audiences. Despite the lack of consensus on the definition of these two terms, it can be argued that from a macro perspective, they both observe the quality of interactions between individuals and products, services, or environments.

All individuals are only Temporarily Able-Bodied (Breckenridge & Vogler, 2001), and able-bodied individuals have no guarantee that they will remain able-bodied in perpetuity. A highly accessible and usable product has a higher chance for longevity as the user's abilities change over time. In addition, abilities and disabilities exist on a continuum and are not dichotomous (Barnett & du Toit, 2018; De Schauwer et al., 2021; IO, 2018; Rabbitt, 2019). For example, some individuals may not be able to hear very well. However, they may not consider themselves deaf or hard of hearing, possibly due to social stigma (Spassiani & Friedman, 2014).

These individuals can often benefit from the same accessibility features intended for deaf or hard-of-hearing individuals.

Ability and disability can also be contextual and manifest only in specific environments, with certain product features, or under certain conditions. In fact, disability itself can be defined as a discrepancy between any person's abilities and their task requirements or environment design (Law et al., 1996; Ricee, 2020). A COVID-19 rapid test, which is designed for administration in as few steps as possible, has a short and intelligible step-by-step administration guide and provides access to an online video guide, may benefit an individual who has "brain fog" due to illness the same way it benefits an individual with cognitive disabilities such as ADHD. Evidence supports that User Experience design informed by the perspectives of PwD leads to a more inclusive User Experience for PwoD (Clarkson et al., 2013; Coleman et al., 2016).

The above statements are not meant to downplay the significant impact of disabilities on the lives of PwD or draw an ignorant analogy between "temporary/potential limits on abilities" and "disabilities." Instead, they aim to draw attention to the reasons PwoD can also gain improved usability from many features that traditionally may be known as accessibility features. Some disability-specific features (such as Braille) remain exclusively beneficial for PwD. However, many accessibility features are pertinent to PwoD, even those examined by assessments designed exclusively for PwD needs (more on page 158).

The definitions above suggest that "Usability/Accessibility" can be defined as the quality of interactions between PwoD or PwD and products, services, or environments. Subsequently, "Usability/Accessibility Information" refers to information that can apprise PwoD or PwD of the

presence or absence of features that improve the quality of their interactions with various products, services, or environments. In a more condensed format, “Usability/Accessibility Information” can educate all individuals on how easy a product, service, or environment will be to use.

### ***Disability and Medical Device Market Statistics***

In the United States, approximately 40.7 million People with Disabilities live in non-institutionalized settings (roughly 13% of the US population) (United States Census Bureau, 2019). For these individuals to be able to live independently, the ability to make informed choices while purchasing consumer products is of crucial importance. With a purchasing power of about half a trillion dollars, PwD still have limited access to accessibility information related to products and services (M. Lin et al., 2018; Scarborough-Kaufman, 2019). One major consumer product category critical to PwD participation in their healthcare is medical devices and equipment (Espicom, 2009; FDA, 2019; SelectUSA, 2019). However, there are tremendous disparities in healthcare access that prevent timely diagnosis and treatment of older adults and PwD, primarily due to the inaccessibility of medical equipment (Becker et al., 1997; Cheng et al., 2001; Coyle & Santiago, 2002; Grabois et al., 1999; Kailles, 2007; Markwalder, 2005; McClain et al., 2000; North Carolina Office on Disability and Health, 2007; Nosek et al., 1995; Nosek, 2000; Nosek & Howland, 1997; Sanchez et al., 2000; Schopp et al., 2002; Veltman et al., 2001).

Due to the advances in medicine and healthcare, PwD are living longer. This, along with the rapidly aging population, has increased the use of healthcare devices at home. These devices are often purchased either by individuals or their care partners (Gans et al., 1993; Kraus et al., 2017; Sade, 2012). These factors highlight the need for accessibility information for

medical devices and equipment to assist PwD, older adults, and their care partners in making informed purchasing decisions appropriate to the user's ability levels. More recently, the COVID-19 pandemic further highlighted this need due to the healthcare shift into home settings.

Medical devices are an approximately \$200 Billion industry in the United States (Donahoe, 2021). This makes the US the largest medical device industry and the world's largest medical device market (Espicom, 2009; SelectUSA, 2019). Hundreds of millions of dollars are spent on medical devices and equipment that may not even function as intended for most PwD and older adults, often due to the lack of accessibility information (McClain et al., 2000).

### ***Academic and Regulatory Efforts***

Significant strides have been made over the past few decades to address the disparity in healthcare access and outcomes relating to medical devices (ADA.gov, 2020; Center for Universal Design, 2002; Mendonca & Smith, 2007; Story et al., 2010, 2002). These efforts, although sparse, have focused on assessing the accessibility of medical devices and evaluating their usability for PwD. These assessments can help inform consumers on the various aspects of device accessibility. The direct benefit of providing medical device accessibility information is that the consumer will make an informed purchasing decision based on the device's features and their ability levels.

One of the most notable examples of such efforts is Dr. Rochelle J. Mendonca's dissertation, entitled "Accessibility Labeling and its Effects on Medical Device Purchase Decisions by Consumers with Disabilities." This dissertation investigated two key areas: 1) The impact of providing information on accessibility for medical devices to PwD and the tradeoffs

with cost and user ratings, and 2) The effect of sociodemographic characteristics on medical device purchasing decisions. Two types of medical devices were used as examples: blood pressure monitors and examination tables. The findings from 161 PwD demonstrated their preference towards devices with accessibility information, and sociodemographic characteristics have no statistically significant relationship with their decision to purchase the most accessible devices. The findings also demonstrated that the availability of accessibility information is just as important as, if not more than, other factors such as cost and user ratings (Mendonca, 2010).

The US regulatory bodies have also recognized the push for manufacturing accessible medical devices. In 2010, the US Patient Protection and Affordable Care Act (ACA) mandated the necessary improvements for the accessible design of medical products (Patient Protection and Affordable Care Act, 2010). Consequently, the US Access Board advanced the effort by developing accessible design guidelines for diagnostic equipment (United States Access Board, 2010). These guidelines were later adopted by the US Department of Veterans Affairs in 2017 (United States Access Board, 2017).

Sections 4302 and 5307 of the ACA focus on enhancing healthcare accessibility from a broader perspective. Section 4302 emphasizes the importance of collecting and reporting data on healthcare access and quality disparities to promote transparency and address accessibility barriers. Section 5307, on the other hand, primarily focuses on enhancing healthcare accessibility through increased funding and resources for expanding community health centers. This provision aims to improve access to quality healthcare services, particularly for underserved and vulnerable populations. These sections collectively reinforce the ACA's

commitment to promoting healthcare equity and ensuring equal access to healthcare services, particularly for underserved populations, including those with disabilities.

Sections 1557 and 4203 of the ACA narrow down the focus of the legislation on the importance of medical device accessibility within the healthcare system. Section 1557 prohibits discrimination based on race, color, national origin, sex, age, and disability in healthcare programs and activities receiving federal funding. Under Section 1557, individuals with disabilities are granted equal access to healthcare services, including access to medical equipment. Complementing this, Section 4203 explicitly addresses medical device accessibility by requiring medical device manufacturers to establish policies and procedures for accessible device design and distribution, promoting standards for the accessibility and usability of these devices. These sections of the ACA jointly serve to eliminate disparities and improve the accessibility of medical devices for individuals with disabilities, thereby reducing health outcomes disparities between PwD and PwoD.

Additionally, Section 4203 mandates the “establishment of standards for accessible medical diagnostic equipment” by the US Access Board (formerly known as the Architectural and Transportation Barriers Compliance Board) with input from the Food and Drug Administration. In compliance with this legal mandate, the US Access Board developed specific standards and guidelines for making medical diagnostic equipment accessible to individuals with disabilities. These guidelines outline design and technical requirements to ensure that diagnostic equipment (e.g., examination tables, scales, and medical imaging devices) can be used by individuals with various disabilities. Thus, the ACA enabled the Access Board to develop accessibility standards and requirements for medical equipment, establishing a framework that

helps reduce disparities in healthcare outcomes. Prior to the Access Board’s provision of its Final Rule on medical device accessibility in 2017, medical devices were not subject to any accessibility requirements (Gardner-Bonneau, 2019).

Despite all these regulatory and academic research efforts and the clear extant need for more accessible medical devices by PwD, the medical device industry has not yet fully adopted the accessible design model (along with many other responsible entities) (Glenn et al., 2023; Magasi & Marshall, 2021; Peacock et al., 2015; Raymaker et al., 2017). Our recent survey also confirmed that PwD are still very dissatisfied with their medical device selection and purchasing processes, view their medical devices as challenging to use, and indicate that having more accessibility information on the product prior to purchasing will help. In the second chapter, Study 1 will discuss this survey and its findings further.

### ***Accessible Medical Devices or Medical Device Accessibility Information***

Thus far, the need for accessible medical devices has been extensively detailed. However, some studies and references mentioned above demonstrated the need for “accessibility information” on medical devices. Examining the relationship between these two concepts can help identify their precedence more effectively.

To address healthcare outcomes disparities between PwD and PwoD due to medical equipment inaccessibility, improving the accessibility of medical devices should always remain the most significant objective. However, to improve medical device accessibility, identifying problem areas appears to be the first logical step. Problem discovery for medical device accessibility can be conducted using conventional methods such as Task Analysis or Usability Testing (Lewis, 2012; Story et al., 2010). Another way to approach this phase is by deploying

instruments designed to meticulously identify the tasks required to operate the medical device and then compare those tasks against a comprehensive matrix of functional disabilities to generate a report on the accessibility of the medical device for different categories of PwD. One of the prime examples of instruments suitable for the latter method is the Medical Devices & Equipment – Accessibility and Universal Design Information Tool (MED-AUDIT) (Mendonca et al., 2008; R. O. Smith et al., 2007).

Regardless of approach, a byproduct of problem discovery is extensive information about issues that prevent users from effectively operating the medical device to achieve the intended effect. Disabilities that are not impacted and disabilities that are already accommodated by the medical device features can also be identified in this process. This information is referred to as “accessibility information” and has been demonstrated to have significant value for PwD, as discussed earlier. Access to this information prior to purchase will help PwD or their care partners make informed purchasing decisions. As a result, even if most medical devices are inaccessible, PwD may find it easier to purchase medical devices that match their abilities and needs.

Designing fully accessible medical products may be a process rather than an immediately achievable objective. However, medical device accessibility evaluation can be conducted virtually for all medical devices currently available on the market by any individual with access to an evaluation instrument and the knowledge of how to apply it. The accessibility information produced from such evaluations can inform manufacturers about current needs and future design improvements, insurers about what medical devices to cover to remain in

compliance with regulatory requirements, and PwD about the medical devices they will be able to use effectively.

Accessibility of medical devices remains the ultimate goal for reducing health outcome disparities. However, the availability and dissemination of medical device accessibility information seem to take precedence due to the potential for widespread and immediate impact. Accessibility information obtained via accessibility evaluations of medical devices also paves the way for directing the industry's attention toward the populations for whom they design their products.

### ***Industry Adoption of Accessible Design Principles***

Two main factors contribute significantly to an industry's decision to adopt the accessible and universal design models: regulations and profit (Vanderheiden & Tobias, 1998, 2000). The shortcomings of the laws and regulations surrounding the topic, or their implementation and enforcement, are outside the purview of this dissertation. However, according to Vanderheiden and Tobias (1998, 2000), the profitability of designing and marketing more accessible medical devices can be further studied on two axes: perception and market motivations. In general, various industries perceive accessible design strategies as limited in scope and serving only a "niche audience" (i.e., PwD). Thus, adopting these strategies may seem to unjustifiably increase the design and manufacturing costs with minimal to no return. In addition, targeting PwD due to their larger share of the population appears to define the market motivation, especially for larger manufacturers seeking to maximize their profits.

The impact of accessible medical devices on the health outcomes of PwD is significant, as evidenced by academic research and regulatory efforts. However, the industry seems

reluctant to invest in designing and marketing more accessible medical devices and equipment. One immediately tangible manifestation of this issue is the unavailability of medical device accessibility information on major online marketplaces for medical devices. In the industry's view, investing in accessible design strategies increases design and manufacturing costs and only serves a comparatively small and niche portion of the population (Shinohara et al., 2018). Therefore, demonstrating that PwOD also prefer more accessible medical devices becomes the key to 1) adjusting perceptions on accessible design and 2) redefining market motivation for the medical device industry.

### ***Approaches to Examining Preferences***

Historically, in the context of healthcare economics (and economics in general), two approaches have been used to examine the preferences of consumers empirically: Revealed Preference (RP) and Stated Preference (SP) methods (de Corte et al., 2021; Lancsar & Louviere, 2008). RP methods allow for examining and modeling existing consumer preferences and market demand through direct or reported observation (K. Harris & Keane, 1998). These methods have high external validity when estimating demand models; however, one of the significant flaws of such methods is that they are useless in directly modeling the demand for products and features that do not yet exist (Kroes & Sheldon, 1988). On the other hand, SP methods allow for examining what the consumers report they would prefer rather than observing their preferences in action (Hall et al., 2004). As such, SP methods enable demand modeling for features, products, and services that do not exist and where direct market observations are not feasible (Carlsson, 2011). Since Usability/Accessibility Information is a

feature of medical devices not yet available, SP methods appear suitable for examining PwOD preferences toward this information.

SP methods consist of various approaches, such as Best-Worst Scaling, Conjoint Rank Analysis, and Discrete Choice Experiments (DCEs). In Best-Worst Scaling designs, participants are presented with a group of choices and asked to identify the most desirable and the least (Jaeger et al., 2008). In Conjoint Rank Analysis, participants are asked to rate or rank the choices presented in each group (Ryan, 2000). DCEs present the participants with a set of choices and ask them to decide on which product they would purchase in each set, with the underlying assumption being that consumers choose the option that grants them the highest benefit and utility (Viney et al., 2002).

Studies on the practicality of different SP methods have concluded that Conjoint Rank Analysis experiments place a high cognitive burden on participants, thus decreasing survey completion rates (Hauser, 2014; Huertas-García et al., 2016). Conjoint Rank Analysis methodologies stem solely from mathematical theories (Conjoint Measurement) and not human behavior models (such as Random Utility Theory) (J. J. Louviere et al., 2010). Additionally, interpreting the meaning of rankings and error components in Conjoint Rank Analysis is complex, and the results are usually of questionable validity and reliability (Flynn & Marley, 2014; J. J. Louviere et al., 2010).

Best-Worst Scaling and DCEs both rely on the Random Utility Framework (Flynn & Marley, 2014; J. J. Louviere et al., 2010) and have an established history in healthcare research (Cheung et al., 2016; E. de Bekker-Grob et al., 2012; Hollin et al., 2022; Soekhai et al., 2019). However, DCEs have seen an exponential growth in healthcare and health economics literature

**Table 1:***Summary of Stated Preference Methods Comparison*

	Participant Cognitive Burden	Theoretical Basis	Data Interpretation Difficulty	Validity & Reliability	Prevalence in Health Economics
Conjoint Rank Analysis	High	Mathematics	Difficult	Low	Low
Best-Worst Scaling	Medium	Human Behavior	Easy	High	Medium
Discrete Choice Experiment	Low	Human Behavior	Easy	High	High

over the past few decades (Soekhai et al., 2019; Wang et al., 2021). DCEs are also the least difficult to design and implement and impose the least cognitive burden on participants compared to other SP methods (Kjaer, 2005). This, along with the high validity of DCEs, has made this methodology particularly well-suited for eliciting preferences in the context of healthcare (E. W. de Bekker-Grob et al., 2019, 2020; F. R. Johnson et al., 2019; Linley & Hughes, 2013; Quaife et al., 2018; Telser & Zweifel, 2007). An application of DCE to examine medical device accessibility information preferences among PwD can be found in the Mendonca (2010) doctoral dissertation. A summary of the comparison between different SP methods is presented in Table 1.

### **Purpose of the Study**

This study aims to empirically investigate whether the availability of accessibility and usability information impacts PwD's medical device purchasing decisions and whether they prefer medical devices with higher accessibility and usability. This study projects that the accessibility and usability of medical devices can be marketed effectively to PwD and PwD. By demonstrating the value and importance of accessible medical devices and medical device

accessibility information for PwOD, manufacturers may have their concerns mitigated regarding the perception of accessible products. They may also find increased market motivation to justify the additional design and manufacturing costs.

### **Significance of the Study**

As discussed previously, health outcomes for PwD are severely impacted by inaccessible medical devices and equipment. Regulatory guidelines and the extant body of evidence mandate, advocate for, and support the need for more accessible medical devices. However, manufacturers seem reluctant to invest the additional time and resources to improve the accessibility of their products even marginally, or at least evaluate and market their products with accessibility information to assist purchasing decisions. This study will help demonstrate that PwOD prefer:

- the availability of accessibility and usability information for medical devices
- medical devices with higher accessibility and usability

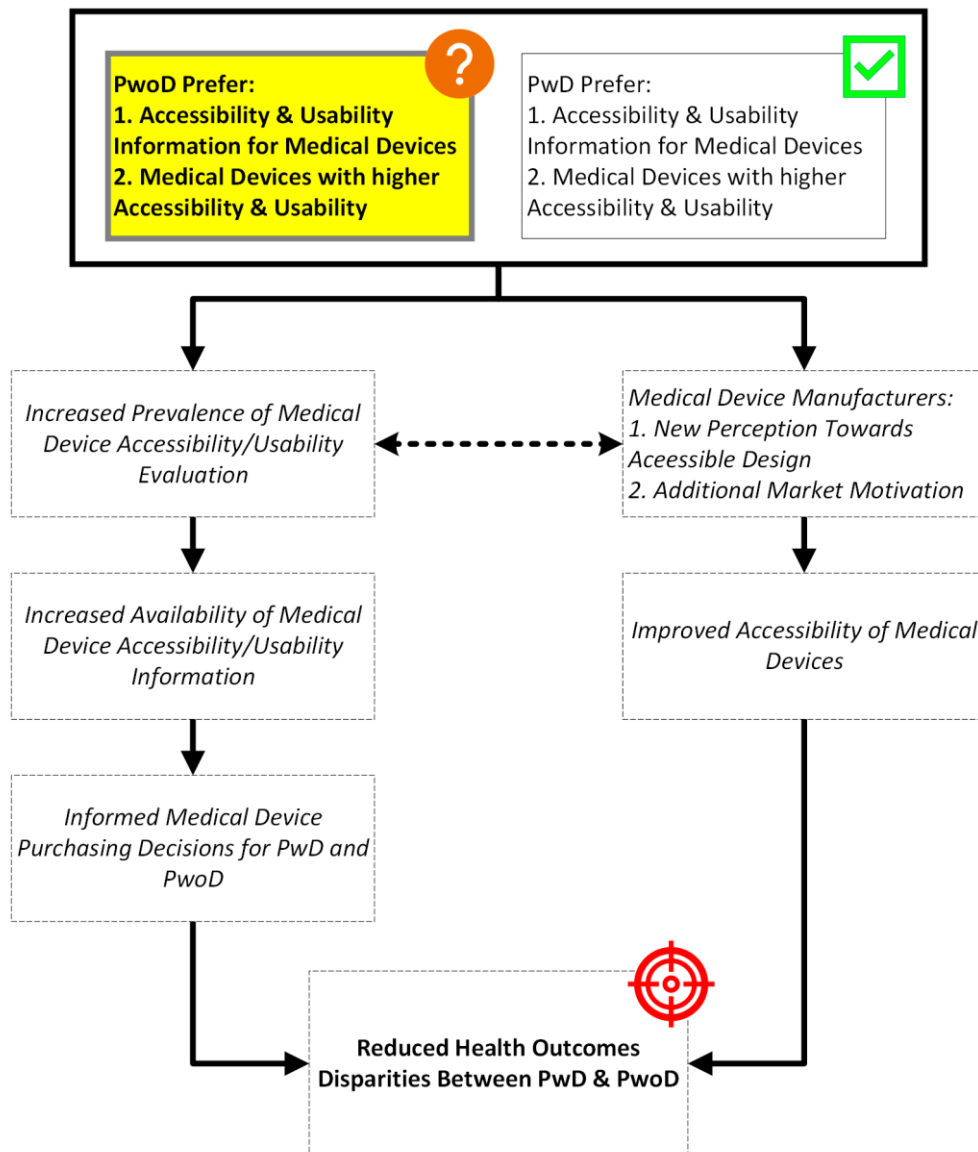
It has been established that PwD value accessibility and usability information for medical devices, and they prefer medical devices that are more accessible (Mendonca, 2010). However, to motivate medical device manufacturers and increase the prevalence of accessibility/usability evaluation for medical devices, a much larger demand must be demonstrated to justify the additional cost of evaluation, research, and product development. The significance of this study lies in demonstrating the potential for a more significant market demand for accessible medical devices and accessibility/usability information. By increasing the appeal of manufacturing accessible medical devices and the availability of medical device accessibility information, the disparities in health outcomes between PwD and PwOD may be

reduced to a significant degree. Figure 1 demonstrates the significance of this study due to the critical role it may play in reducing health outcome disparities between PwD and PwoD.

This study may also have a broader impact outside the specific scope of its hypotheses. By demonstrating the value of accessible design for medical devices for consumers with or without disabilities, this study may provide a blueprint for similar studies for different

**Figure 1:**

*The Significance of This Study in Reducing Health Outcomes Disparities Between PwD and PwoD*



industries. It also can potentially set or impact priorities in User Experience Research by demonstrating the overarching value of accessible design. Additionally, the methodology used in studying Stated Preferences appears to hold great potential for client-centered rehabilitation research and therapy plan development, in which it is currently severely underutilized.

## II. Prior Discoveries

### Overview

This chapter will discuss two prior studies in summary. Overall, these studies helped inform hypothesis design and evaluation for this dissertation. In Study 1, under the supervision of Drs. Roger O. Smith and Rochelle J. Mendonca, I designed and distributed a survey to examine the satisfaction of PwD (170 blind or low-vision individuals) with their medical devices and the significance of accessibility information in their purchasing decisions. Consequently, I analyzed the results, drafted a manuscript based on the findings, and submitted the manuscript for publication in the Journal of Blindness Innovation and Research after comments and revisions from Drs. Smith and Mendonca. The results indicated low satisfaction with the accessibility of medical devices and showed great support for the availability of accessibility information before purchase.

In Study 2, I conducted eight interviews with PwD and PwoD to identify how accessibility information may be relevant to PwoD. I asked the interviewees for their viewpoints on the significance of accessibility information for PwoD and examined how it needs to be presented to PwoD for maximum impact. This study was supported by the National Science Foundation Innovation Corps (I-Corps™ for short) program (National Science Foundation, 2020).

The findings of Studies 1 and 2 demonstrate (a) PwD's need for accessibility information; (b) acknowledgment from PwoD that accessibility information may relate to them; and (c) offering this information will assist PwoD in their purchasing decisions. Combined, these studies highlight the need for quantitative evaluation of the impact of medical device accessibility information on PwoD's purchasing decisions. These two studies highlighted that offering

accessibility information for medical devices improves decision-making and satisfaction for all consumers, regardless of their current levels of ability.

### **Study 1**

The first objective of this study was to examine how PwD perceive the accessibility of the medical devices and equipment they use at home or healthcare facilities. The second objective was to determine whether medical device accessibility information would be considered beneficial to this population. The results established the need to develop and adopt information technologies that present medical device accessibility information to help PwD make informed decisions.

### ***Study Design***

This cross-sectional study was conducted in a self-administered web survey format. The University of Wisconsin-Milwaukee Institutional Review Board reviewed and approved the survey. The participants received a survey link, allowing them to take the survey on any electronic device, such as a computer or smartphone. The survey was designed using Qualtrics (Qualtrics, 2023) and included four main sections: (a) general demographics and impairments; (b) experiences with medical devices at home; (c) experiences with medical devices at healthcare facilities; and (d) general medical equipment accessibility questions. Each section consisted of open-ended and closed-ended questions, with most of the questions requiring a response on a 5-point Likert-type scale for rating experiences or level of agreement with a statement. Overall, the survey consisted of 41 questions. However, some participants may have seen fewer questions depending on some of their answers (with 17 questions being the minimum).

### ***Data Collection***

Inclusion criteria were adults (over 18), having a disability or being a care partner for PwD, and English proficiency. Due to the exploratory nature of this study, we used purposive sampling by contacting 18 national and local disability advocacy organizations for participant solicitation. These organizations encompassed and supported a wide array of disabilities.

Informed consent was obtained electronically at the beginning of the survey, followed by screening questions. Participants were not under any time constraints as they could take the survey at the location and time of their choosing, using the devices with which they felt comfortable. The survey was reviewed and revised multiple times by the research team and occupational therapy-affiliated students at the Rehabilitation Research Design & Disability Center to review the flow and accessibility. At the beginning of the survey, we advised our participants to contact us with their questions or comments or if they required additional assistance in completing the survey. Data collection took place over five weeks in the winter of 2021.

### ***Data Analysis***

We reviewed all responses and excluded incomplete submissions. We then used Qualtrics built-in tools and Microsoft Excel (Microsoft Corporation, 2023a), SPSS (*IBM SPSS Statistics for Windows*, 2021), and NVivo (*NVivo*, 2020) for our mixed-methods data analysis. We coded the medical devices named by the participants into consistent categories to facilitate comparison (e.g., “blood pressure monitor” and “BPM” were coded as “Blood Pressure Monitor”).

Following the widely accepted procedure prescribed by Braun and Clarke, we performed a thematic analysis of the qualitative data using a semantic inductive approach (Braun & Clarke, 2006). Two experts in medical device accessibility and universal design coded the data. Conflicts in coding were negotiated between the two experts until they reached a consensus. When a consensus was not reached, a third expert in accessibility and universal design was consulted for the verdict. Coders first familiarized themselves with the responses and comments and then coded them with short descriptive labels. Next, they extracted the themes observed in the labels and reviewed them for generalizability and mutual exclusiveness.

## ***Results***

**General demographics and impairments.** Overall, from the 198 initial submissions and after removing incomplete records, 170 submissions remained from individuals who identified their primary functional impairment as vision-related (blindness or low vision). These individuals were aged between 18 and 93 years, of which 65% were 47 to 72. The ethnic backgrounds consisted of 82.6% White, 4.57% Black, 2.29% Native American, 2.29% Asian, and 8% selected Other (5 individuals chose to identify with more than one ethnicity). Our participants were 66.1% female and 32.7% male, from 39 States across the US, with California and Michigan having the highest number of participants in the study (13.6% and 6.8% respectively). Of these individuals, 81.2% identified blindness as their primary disability, and 28.2% selected low vision. Participants were allowed to identify more than one primary disability. The third most common primary disability was identified as partial hearing loss (at 12.9%), with other impairments such as difficulty reading, impairments in upper or lower extremities, and deafness also identified at lower rates.

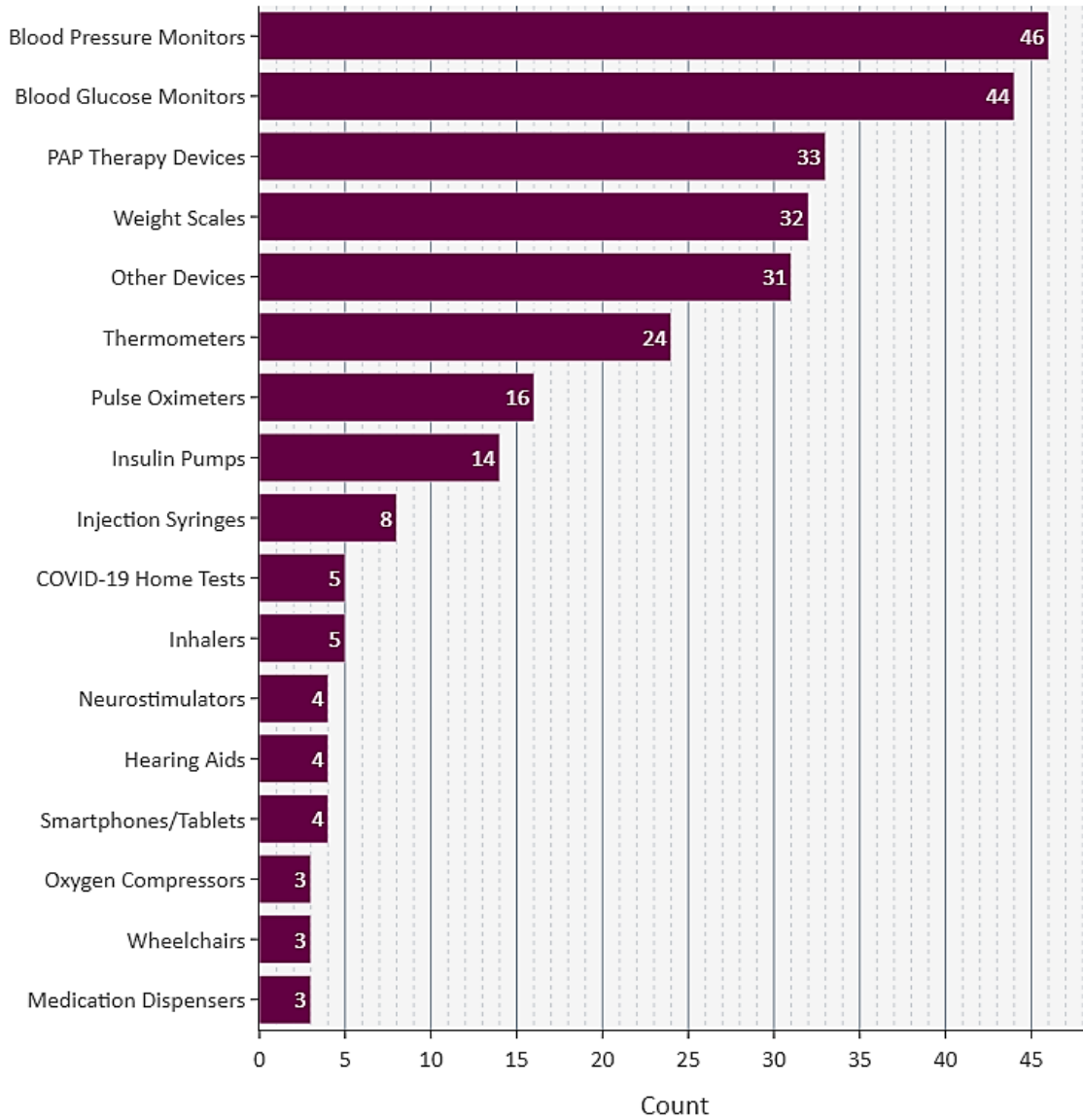
**Experiences with medical devices at home.** We asked our participants to identify up to 3 medical devices and equipment they use at home. We categorized these devices into 17 categories. Figure 2 shows these categories in descending order. Overall, blood pressure monitors and blood glucose monitors were the most common devices identified (with 16.49% and 15.77% respectively). For each home medical device identified, we asked five follow-up questions:

1. How challenging was the process of selecting the device? (5-point Likert-type scale from extremely challenging to not challenging at all)
2. How challenging was the process of purchasing the device? (5-point Likert-type scale from extremely challenging to not challenging at all)
3. How easy was the device to use? (5-point Likert-type scale from extremely difficult to extremely easy)
4. How satisfied are you with your purchase? (5-point Likert-type scale from extremely dissatisfied to extremely satisfied)
5. Did you have enough information about the accessibility of your device before purchase? (Yes/No)

Our participants had found selecting their devices challenging, with 67.31% of the ratings indicating a moderately challenging, very challenging, or extremely challenging selection process. By a relatively narrow margin, most of our participants indicated having a better experience with their purchasing process, with 55% rating the purchasing process as either slightly challenging or not challenging at all. The remaining 45% perceived the purchasing process as moderately to extremely challenging. Most participants found their home medical

**Figure 2:**

*Commonness of Medical Devices Used at Home*



devices extremely or somewhat difficult to use, with 72.76% of the total ratings. Only 16.34% of the responses rated the home medical device somewhat or extremely easy to use. 53.05% of the participants indicated they were somewhat or extremely dissatisfied with their purchase, compared to 35.50% who expressed satisfaction. In response to the last question, 61.42%

selected “No,” indicating that the majority did not have access to any information on the accessibility of their medical device before purchase.

**Experiences with medical devices at healthcare facilities.** In this survey section, we asked our participants to identify up to 3 medical devices and equipment they have used in a healthcare facility (such as the doctor’s office, clinics, and rehabilitation facilities). We categorized the types of devices and equipment identified into five general categories: Diagnostic, Positioning, Therapeutic, Information Technology, and Other Devices. Diagnostic equipment included devices such as X-ray scanners, MRI scanners, mammography equipment, blood pressure monitors, weight scales, and more. Positioning equipment included hospital beds, exam tables, dentist chairs, and other similar equipment. Therapeutic equipment included oxygen delivery systems, IVs and injection systems, pain management systems, and more. Information Technology devices included information kiosks, sign-in sheets, food selection systems, and other means through which the facility gathered or offered information to our participants. We bucketed these devices and equipment into more general categories since there was great variability between the devices our participants had mentioned and to reduce the number of devices that fell under “Other.” Any device or equipment that did not fit the above categories was placed into “Other Devices.”

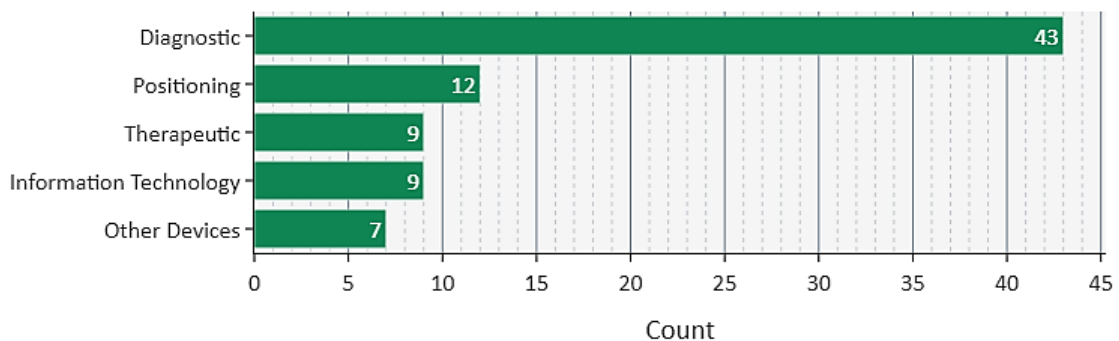
The most common equipment type mentioned was Diagnostic equipment, with 53.75% of the responses. Positioning equipment with 15%, Therapeutic equipment with 11.25%, Information Technology equipment with 11.25%, and Other Devices with 8.75% of the mentions came next. Figure 3 shows these categories in descending order. Following device identification, we asked our participants 3 questions per device:

1. How satisfied were you with the device you used? (5-point Likert-type scale from extremely dissatisfied to extremely satisfied)
2. How easy was the device to use? (5-point Likert-type scale from extremely difficult to extremely easy)
3. Did you have enough information about how easy or difficult the device will be to use? (Yes/ No)

In response to the first question, 56.58% expressed that they were somewhat or extremely dissatisfied with the device they used. Only 10.53% expressed being somewhat or extremely satisfied, and 32.89% of the ratings indicated neither satisfied nor dissatisfied. The responses to the second question on ease of use suggested that the majority found the equipment and devices at healthcare facilities challenging to use, with 56.58% of negative ratings (extremely or very challenging). Positive ratings (slightly challenging or not challenging at all) were only 17.11% of the ratings, and 26.32% indicated that medical devices were moderately challenging to use. In response to the last question, 58.97% of the respondents

**Figure 3:**

*Commonness of Healthcare Facility Medical Equipment Categories*



indicated having no knowledge of how easy or difficult the equipment will be to use. Only 41.03% of the responses indicated having knowledge of device accessibility before their visit.

**General medical equipment accessibility experiences.** In the last section of the survey, participants (a) rated their knowledge of the United States Access Board guidelines and standards for medical device accessibility; (b) rated their agreement with the benefits of having more personalized medical device accessibility information and how it impacts purchasing decisions; and (c) shared their worst experiences related to medical equipment accessibility.

First, participants were asked, “How familiar are you with the Access Board standards on the accessibility of medical equipment?” They rated their knowledge on a 5-point Likert-type scale from extremely familiar to not familiar at all. The overwhelming majority responded with not familiar at all, with 69.0% of the responses. 19.3% indicated they were slightly familiar, 9.7% indicated they were moderately familiar, and 2.1% indicated they were very familiar. None of the participants indicated being extremely familiar with the medical equipment accessibility standards and requirements.

The next question asked participants to rate their agreement with this statement: “Knowing about the accessibility ratings of medical devices as it relates to my specific impairments will be very helpful in informing my purchases.” On a 5-point Likert-type scale from strongly disagree to strongly agree, an overwhelming majority strongly agreed with this statement, with 69.4% of the responses. 18.8% somewhat agreed, 6.9% neither agreed nor disagreed, 2.1% somewhat disagreed, and 2.8 strongly disagreed.

The last survey question asked participants to share the worst experience they may have had relating to medical device accessibility. This question was open-ended and received 103 responses. Overall, seven themes emerged from the data:

1. **Lack of Accessibility Information and Features:** With 63 mentions, this theme was the most prominent among the collected responses. Responses that mentioned devices without accessibility features required for user operation or indicated a lack of information on the accessibility of products before purchasing were coded under this theme.
2. **Impact on Independence:** With 28 mentions, this was the second most common theme. Responses coded under this theme typically mentioned medical devices that required sighted assistance for operation or procedures and requirements that hindered independence.
3. **Systemic Issues:** This theme was the third most prominent, with 22 mentions. Responses coded under this theme typically mentioned the shortcomings of healthcare facilities' procedures, healthcare staff, medical device manufacturers, insurance companies, and issues related to the device prescription, delivery, or compensation for the cost.
4. **Complicated Instructions and Operation:** This theme emerged 15 times, making it the fourth most prominent theme. Responses that mentioned confusing operation procedures for devices, lack of instructions, inaccessibility of instructions, or challenges with operating devices properly were coded under this theme.

5. **Financial Impact:** With 14 mentions, this theme was the 5th most common theme in the responses. We included responses that highlighted the financial concerns of the respondent in purchasing, using, or maintaining their medical devices. Responses that referred to the higher cost of accessible medical devices, compared to their inaccessible counterparts, were also coded under this theme.
6. **Inferior Accessible Product:** With ten mentions, this theme was the sixth most prominent. Responses coded under this theme typically mentioned medical devices that are accessible but inferior in quality or features compared to their inaccessible counterparts.
7. **Potentially Life-threatening Consequences Due to Inaccessibility:** This theme emerged seven times and included responses that directly or indirectly mentioned a potentially life-threatening outcome solely due to the inaccessibility of medical devices. Almost all inaccessibility issues with medical devices, by nature, may lead to adverse and potentially life-threatening outcomes. However, under this theme, we only coded responses explicitly mentioning a life-threatening scenario. We made no inferences regarding the participant's intent or the connotations of their response.

## **Study 2**

The objective of this study was to gain a better understanding of accessibility information utility for PwOD via interviews. I conducted these interviews in the customer discovery phase of an NSF I-Corps™ program. The summary results were submitted as a part of the final report to fulfill the program requirements.

### ***Study Design***

The interviews were semi-structured and adhered to scientific interview best practices (Adeoye-Olatunde & Olenik, 2021; DeJonckheere & Vaughn, 2019) to examine perspectives on the potential utility of medical device accessibility information for PwoD. At the beginning of each interview, I explained the context and asked the interviewee to offer their views on the perceptions of PwoD on medical device accessibility information. Based on their response, I further navigated the interview to examine their comments and underlying thoughts.

### ***Data Collection***

All interviews were conducted online via Zoom (*Zoom Meetings, 2022*) and were recorded and transcribed. Data collection took place in January 2022. I summarized the key points from each interview by consulting the transcript and live interview notes.

Convenience sampling was used for recruitment. Overall, eight participants were interviewed: two undergraduate students in Occupational Sciences, two graduate students in Occupational Therapy, one doctoral student in Health Sciences, two Occupational Therapy professors, and one computer scientist with extensive exposure to accessibility concepts. Seven interviewees considered themselves as PwoD. While on the surface, it may appear that the sampling introduced bias to the study, this was intentional and by design. The interview explored a topic that was only familiar to some participants. While familiar with accessibility for PwD with varying degrees of mastery, most participants were not accustomed to considering accessibility and usability for PwD as a factor of utility for PwoD. In conducting these interviews, I attempted to facilitate looking at a familiar concept (accessibility for PwD) through a mostly unfamiliar and unexplored lens (PwoD perspective). This was done by encouraging participants

to assume the mindset of an individual they know who does not have a disability and is unfamiliar with accessibility concepts. In this way, the participants' knowledge of accessibility and accessible design became an asset in exploring how PwoD may perceive the topic.

### ***Data Analysis***

While the key findings are not quoted verbatim, I aimed to capture the essence of the participants' thoughts and ideas while generally remaining loyal to their wording and sentence structure, with adjustments for clarity. This was done to convey the significance and value of these insights while minimizing the potential for explicit confirmation bias in reporting. Of course, due to the nature of the interviews and their context, the findings may not be free of other sources of bias such as interviewer bias, response bias, Hawthorne effect, etc. Regardless, these interviews provided valuable insights and ideas for framing the central hypothesis of this dissertation. They also aided in the design and presentation of the study to the target audience.

### ***Results***

Below, each interviewee's five most significant points are listed, except for Interview 3. In Interview 3, the conversation mostly revolved around the impact of disability on different aspects of social life, which does not pertain directly to this dissertation.

#### **Key Findings from Interview 1.**

- If accessibility is associated with being more straightforward to use, that would be a draw for PwoD.
- Information can be provided on the product label or online.
- The accessibility features that are advertised need to benefit everyone, with or without disability, for PwoD to care.

- Information should not be presented in a way that makes PwoD think this product is only meant for PwD.
- Just merely having accessibility information can be the differentiating factor when there are multiple options.

### **Key Findings from Interview 2.**

- For care partners, the availability of accessibility information makes a significant difference.
- For a care partner, accessibility information will immediately create more trust in the product and the company. It shows they value the accessibility of their products and the care for the customer's experience.
- For PwoD who are purchasing a product for PwD or even PwoD, accessibility information will boost their confidence in the product if they think it will be usable by the person for whom they are making the purchase.
- With a prevalence of between 15 and 20%, there is a high chance that anyone knows someone with a disability.
- Accessibility must equate to ease of use when presented to PwoD.

### **Key Findings from Interview 3.**

- Not all PwoD are going to stay without disabilities forever.
- Raising awareness about the potential for disability is significant, the same as preventative care.

- The effort to provide product accessibility information can start only in the United States at first due to the significance of non-discriminatory treatment, but eventually, it will need to grow to raise awareness globally.

#### **Key Findings from Interview 4.**

- Most people may not know about accessibility, its meaning, and why they should care.
- Describing the accessibility features and why they should matter may increase the number of PwOD who care.
- PwOD tend to think that they will stay without disabilities forever.
- Raising awareness for PwOD about the possibility of disability will assist their appreciation for accessible products.
- The same strategy as organic food labeling can be followed to market accessible devices.

#### **Key Findings from Interview 5.**

- The idea of providing accessibility information for everything is not feasible currently because it will take a lot of investment from various parties to evaluate everything.
- For the idea of accessibility information on products and services to take off, the industry must be convinced that this is the best course of action.
- If large retailers or e-retailers require accessibility information from the manufacturers, that will create a significant incentive.
- Offering accessibility information on products online makes the information more accessible for machines (search engines), which in turn makes search results more

appropriate for the potential customers, impacting their purchasing decisions positively.

- Providing accessibility scores does not necessarily mean anything to an average consumer. Instead, the accessible features need to be highlighted and described.

#### **Key Findings from Interview 6.**

- PwOD may not be willing to initially go out of their way to look for accessibility information, but if it is presented upfront, they may consider it.
- Two distinct approaches to offering accessibility information must exist: online and in-store.
- Many people, when shopping online or in-store, do not take the time to read all the product features and descriptions. If they can, they will check the reviews and make their decision.
- If accessibility information is being offered to PwOD, it needs to be something that pertains to them, either immediately or in the future.
- The marketing method for accessibility information may significantly impact how PwOD view its relevance to their purchasing decisions.

#### **Key Findings from Interview 7.**

- Some PwOD will benefit more from accessibility information, and some will benefit less.
- One segment of PwOD who are going to benefit immensely from accessibility information are individuals who consider themselves as PwOD, but in fact, they do

have disabilities or impairments (such as needing glasses, having difficulty with their night vision, attention deficits, difficulty retaining information, etc.)

- Two categories of PwoD may value accessibility information greatly: 1) those who are altruistic and understand that accessibility information greatly benefits others, and 2) those who may have PwD in their family, friends, or colleagues.
- Some accessibility features can readily become product features, the same way ergonomic design features are product features.
- The presentation will impact how beneficial PwoD may perceive accessibility information.

#### **Key Findings from Interview 8.**

- Since there are more PwoD than PwD, it is worth investigating how a manufacturer's investment in measuring accessibility and providing information will impact the larger population.
- The participant mentioned they have a condition that affects one of their senses. They consider it an impairment and anticipate it will become more complicated over time but do not consider themselves a PwD.
- One way to increase awareness and help PwoD care more about accessibility information is to offer it on various products and platforms. Over time, people will start noticing the additional labeling and information and become more educated. Similar to health and safety ratings for restaurants.
- The launchpad for offering accessibility information should be online, and over time, it can be rolled out to brick-and-mortar stores.

- For products offered in brick-and-mortar stores, accessibility information can be summarized on a label, with a link or QR code for more in-depth and disability-specific information.

**Summary of Findings.** The consensus among the interviewees seems to be that PwOD may have difficulties inferring the benefits of accessibility information. Also, accessibility information needs to be presented to PwOD with emphasis on the added value of accessibility features. Nearly all interviewees highlighted that while accessibility information may not immediately pertain to some individuals, it is worth investigating how it may impact their purchasing decisions. Some interviewees mentioned that an appropriate place to start this investigation would be in online stores, where products with and without accessibility information can be presented to PwOD. Most interviewees also suggested that accessibility information needs to be presented with appropriate framing. Suppose the presentation is inappropriate and implicitly or explicitly suggests the product is designed for PwD. In that case, it may evoke the unfortunately persistent negative stigma around disability and lose its appeal to PwOD.

### **Collective Discussion and Emergent Needs from Prior Research**

Studies 1 and 2, despite their limited scope and examined populations, offer valuable insights into the significance of accessibility information. Study 1, in agreement with numerous other studies discussed under Statement of the Problem in the first chapter, highlights the overwhelming need for medical device accessibility information to assist the purchasing decisions of PwD and their care partners. The discoveries of Study 2 emphasize the need to

evaluate the impact of accessibility information on PwoD's purchasing decisions and provide considerations under which this impact may be accurately examined.

To examine the utility of medical device accessibility information for PwoD purchasing decisions, an online survey experiment needs to be created. The online presentation will minimize the experiment cost while providing a realistic environment to simulate purchasing medical devices. In this survey, participants should be presented with different medical devices and receive the product information typically available when shopping online. Accessibility and usability information must also be presented in a digestible format, among other prevalent product information. This information needs to be defined and framed for participants, so they understand the meaning of it accurately. The impact of other product information in conjunction with accessibility and usability information must be evaluated by examining the stated preferences. This can be achieved by studying the tradeoffs made by the participants and the priority and utility of accessibility information in these decisions. A successful methodology for this investigation needs to allow for explaining PwoD choices within a finite set of alternatives. The next chapter will introduce and discuss a research methodology with such potential.

### III. Methods

Discrete Choice Experiments have been used extensively in healthcare to evaluate stated preferences empirically (Soekhai et al., 2019). DCEs are beneficial for understanding individual preferences and predicting choices by systematically presenting participants with hypothetical alternatives, enabling researchers to inform better policy and product design (Naik Panvelkar et al., 2010). In Chapter I under Approaches to Examining Preferences the background and merits of different Stated Preference (SP) methodologies, including DCE, were reviewed in detail. To synopsise the previous review of DCE advantages, it is often preferred over other stated preference methods because it provides a more realistic decision-making context, allowing participants to make choices resembling real-life decisions. Additionally, it offers the ability to quantify the relative importance of different attributes and estimate the trade-offs individuals are willing to make, providing more robust and actionable insights for policy and product development. Compared to other SP methods, this approach yields richer and more realistic insights while imposing a lower cognitive burden on participants overall.

For instance, in a healthcare setting, a DCE could involve presenting patients with a choice between two treatment options for a specific medical condition. Each option may have varying attributes, such as treatment duration, potential side effects, and cost. Patients are then asked to choose between these options, allowing researchers to assess their preferences and the perceived importance of these attributes in the treatment process. This information can help healthcare providers and policymakers make more informed treatment strategies and resource allocation decisions.

In DCEs, several key terms are used to describe the experiment components. These key terms are defined below (F. R. Johnson et al., 2013; Lancsar & Louviere, 2008; Street et al., 2008; Szinay et al., 2021):

- **Choice Task:** A "choice task" in DCE refers to a scenario presented to participants asking them to choose among different options. In each choice task, respondents are required to select the option they prefer. In the example above, there is one choice task consisting of the two treatment options.
- **Alternative:** An "alternative" is one of the choices presented within a choice task. Alternatives represent the specific combinations of features from which the participants can choose. In the example above, each treatment option is one alternative.
- **Attributes:** In DCE, attributes are the characteristics or features of the alternatives being presented to participants. These attributes are the key factors that can influence the choices made by participants. These attributes are selected to represent the features of the products, services, or choices being evaluated. In the example above, treatment duration, potential side effects, and cost are the attributes.
- **Levels:** Levels represent the specific values or variations within each attribute. The levels provide the range of choices within each attribute that participants can compare when making their selections. In the example above, specifically for the treatment duration attribute, three levels may be identified as short duration, medium duration, and long duration.

In a DCE, researchers vary the levels of attributes to create multiple alternatives and create choice tasks from different combinations of these alternatives. Participants subsequently select their preferred alternative in each choice task based on the attribute levels they prefer. Systematically examining the outcomes helps researchers infer how participants prioritize and make decisions based on the attributes and their levels.

### **Theoretical Foundation and Conceptual Congruence**

DCE methodology is rooted in several theoretical foundations and is compatible with many core research methodology concepts. The scope of the central hypothesis in this dissertation may require only an applied understanding of the methodology for successful evaluation. However, defining these underpinnings will further help establish this methodology as a powerful tool in decision analysis and an appropriate approach for evaluating the central hypothesis.

#### ***Luce's Choice Axiom***

Thurstone pioneered the expansion of psychometric models of choice behavior to accommodate stochastic variations (Böckenholt, 2006). He posited that preferences result from cognitive processes that involve an unobservable and unquantifiable level of randomness (Thurstone, 1927). This work later became the foundation of Random Utility Theory (RUT) (Marschak, 1959). In his other influential work, Thurstone demonstrated the pertinence of experimental examination of preferences instead of observing and evaluating the manifestation of preferences based on real-life choices (Thurstone, 1931). These two approaches were later termed “Stated Preference” and “Revealed Preference” methods, as described previously in Chapter I under Approaches to Examining Preferences.

Luce's Choice Axiom (LCA) was later established based on Thurstone's research (Deng et al., 2023). According to Pleskac (2015) LCA is a probabilistic approach to the examination of choice behavior and is predicated on two grounds: (a) choice behavior is probabilistic rather than deterministic; and (b) there exists a relationship between the probability of choosing an alternative in a choice task and the probability of choosing the same alternative from a choice task with more alternatives. Pleskac explains (a) in a forced binary choice scenario (i.e., the participant has two alternatives and must pick one with no opt-out option). Based on Pleskac's explanation, a deterministic approach to decision analysis would suggest that either one alternative is more desirable than the other, or they are both equally desirable, and the participant is indifferent. However, according to Pleskac, a probabilistic approach would instead use a probability function to describe how likely the alternatives are to be chosen by the participant, and this approach is more likely to be compatible with human psychology. To explain (b) with an example, it stands to reason to assume the probability that a customer buys a shirt with a specific set of attribute levels from a clothing store would be related to the probability of that customer buying the same shirt from a different branch of the same clothing store, regardless of the differences between the inventories of these branches (Pleskac, 2015).

LCA consists of two parts (Luce, 1977, 2008):

- Part 1: In a binary choice scenario between  $x$  and  $y$ , if alternative  $x$  is never preferred over alternative  $y$ , removing alternative  $x$  from the choice task will not impact the choice probabilities in any meaningful way.

- Part 2: Alternatives that are never chosen are termed “irrelevant alternatives” (such as  $x$  in Part 1). In a choice task, the choice probability will not change regardless of how many irrelevant alternatives are added or removed.

Part 2 of LCA is also known as “Independence of Irrelevant Alternatives” (IIA), which became an important part of decision theory (Arrow, 2012; Ray, 1973). Most analytical approaches to model DCE data are based on the fundamental principles of LCA (Billot & Thisse, 1999; Mazzanti, 2003).

### ***Random Utility Theory***

Random Utility Theory (RUT) forms one of the core theoretical foundations for DCE (Vass et al., 2017). Developed by McFadden in 1974, RUT suggests that when presented with a choice-set, individuals are assumed to select the alternative that provides them with the highest utility, which is a combination of the inherent utility of the attributes and the random component representing unobservable factors affecting their choice (McFadden, 1974). In other words, according to RUT, consumer choice behavior consists of a deterministic factor and a stochastic factor (Timmermans, 2001). For his contributions to the development of theory and methodologies for discrete choice modeling, McFadden won the Economic Sciences Nobel Prize in 2000, shared with James J. Heckman (McFadden, 2001; NobelPrize.org, 2023).

### ***Multi-Attribute Utility Theory***

The Lancaster Model posits that consumers derive utility and satisfaction from the specific attributes or characteristics of products rather than the products themselves (Lancaster, 1966). According to Lancaster (1966), consumer preferences are based on these attributes, and consumers allocate their budget to maximize utility by selecting the

combination of attributes they value most. The Multi-Attribute Utility (MAU) Theory is an extension of the Lancaster Model that was developed to provide a structured framework for decision-making involving multiple attributes (Deng et al., 2023). In the MAU framework, preferences for various attributes are quantified to evaluate and compare alternatives with multiple characteristics (Jansen, 2011). MAU Theory is another significant cornerstone of DCE (Vass et al., 2017).

### ***Quantitative Experimental Design Principles***

DCEs are designed to accommodate the key principles of quantitative experimental design, such as randomization, efficient designs (often based on orthogonal arrays or D-Optimal designs), and robust statistical models for data analysis (F. R. Johnson et al., 2013; Kjaer, 2005; Lancsar & Louviere, 2008; Street et al., 2008; Szinay et al., 2021). This compliance with traditional research methods increases the credibility of the findings once the experiment is designed and conducted appropriately.

### ***Marginal Rates of Substitution and Tradeoffs***

The Marginal Rate of Substitution (MRS) refers to the rate at which an individual is willing to trade one attribute or characteristic of a choice alternative for another (Benjamin et al., 2014). In other words, it quantifies how much of one attribute a respondent is willing to give up to gain an additional unit of another attribute while remaining indifferent between two alternatives (Hayes, 2023). By evaluating the choices made in a DCE, one can infer the MRS, which informs stakeholders about the trade-offs that consumers are willing to make when selecting products or services (Lancsar et al., 2017; Mott et al., 2020; Van Der Pol et al., 2014).

### ***Willingness to Pay and Market Simulation***

DCEs can facilitate market simulations and estimate consumers' Willingness to Pay (WTP). Market simulations involve using the preference data collected from DCEs to model and predict consumer behavior in various market scenarios (Mariel et al., 2013; Philips et al., 2012). These simulations allow businesses and policymakers to assess the potential impact of different product features, pricing strategies, or policy changes (Rouvinen & Matero, 2013; Tsafarakis et al., 2011). By incorporating the estimated WTP, which quantifies the monetary value consumers place on specific attributes or improvements, decision-makers can adjust their strategies to better align with consumer preferences and optimize pricing strategies (Nieboer et al., 2010; Ryan & Wordsworth, 2000; Schlereth et al., 2012; Woo et al., 2022). Enabling market simulations and WTP estimation adds to the versatility and utility of DCEs in choice analysis.

### **Critical Prerequisite**

A critical yet occasionally overlooked or underreported component of DCE in healthcare research are supplementary studies to inform the experimental design and the broader context for interpreting the findings (Bridges et al., 2011; E. de Bekker-Grob et al., 2012; Vass et al., 2017). As discussed previously, DCEs are powerful methodologies for understanding preferences, decision-making, and tradeoffs among various attributes of healthcare interventions or services. However, the success of a DCE hinges on the quality and appropriateness of its design.

In general guidelines and best practices for developing DCEs in healthcare, the importance of conducting quantitative or qualitative market studies before the formulation of attributes and levels is consistently emphasized (Bridges et al., 2011; F. R. Johnson et al., 2013). These initial studies serve as a foundation for understanding the unique characteristics of the

market in question, the specific population or stakeholder group under investigation, and the prevailing preferences and priorities within that context. Neglecting this significant preparatory step can have detrimental consequences, as it may lead to the inappropriate development of attributes and their associated levels within the DCE (Coast & Horrocks, 2007; E. de Bekker-Grob et al., 2012; Vass et al., 2017).

While DCEs are valuable tools in healthcare research, their effectiveness and credibility depend on rigorous preparations such as preliminary qualitative or quantitative market studies. These studies help ensure that the attributes and levels in the experiment align with realistic scenarios, fostering greater validity and reliability. In the context of this dissertation, a portion of these preparatory steps were carried out as discussed earlier in Chapter II - Prior Discoveries. Additional market research in preparation for experimental design will be discussed later in this chapter under DCE Design.

### **Research Questions**

This study aims to assess the impact of medical device Usability/Accessibility Information on the purchasing decisions of PwOD. This can be evaluated by examining the stated preferences of PwOD when presented with medical devices differing in attribute levels. DCE methodology seemed to be the most appropriate approach to examine PwOD stated preferences. The central hypothesis of this study, as mentioned earlier, is “Usability/Accessibility Score impacts purchasing decisions of PwOD.” This central hypothesis was examined by investigating the following questions:

- While making purchasing decisions, how much utility do PwOD assign to the most commonly available attribute levels for medical devices in the presence of

Usability/Accessibility Information? This is the main question this study aimed to answer.

- What tradeoffs are PwoD willing to make between the most commonly available attribute levels for medical devices and Usability/Accessibility Information?
- What is the role of sociodemographic characteristics in utility assignment to various attribute levels of medical devices?
- How much does familiarity with disability and accessibility concepts impact the utility assignment to Usability/Accessibility Information?
- How much do PwoD claim they value Usability/Accessibility Information, and how much is this claim confirmed by their stated preferences as determined by the DCE?

## **Research Design**

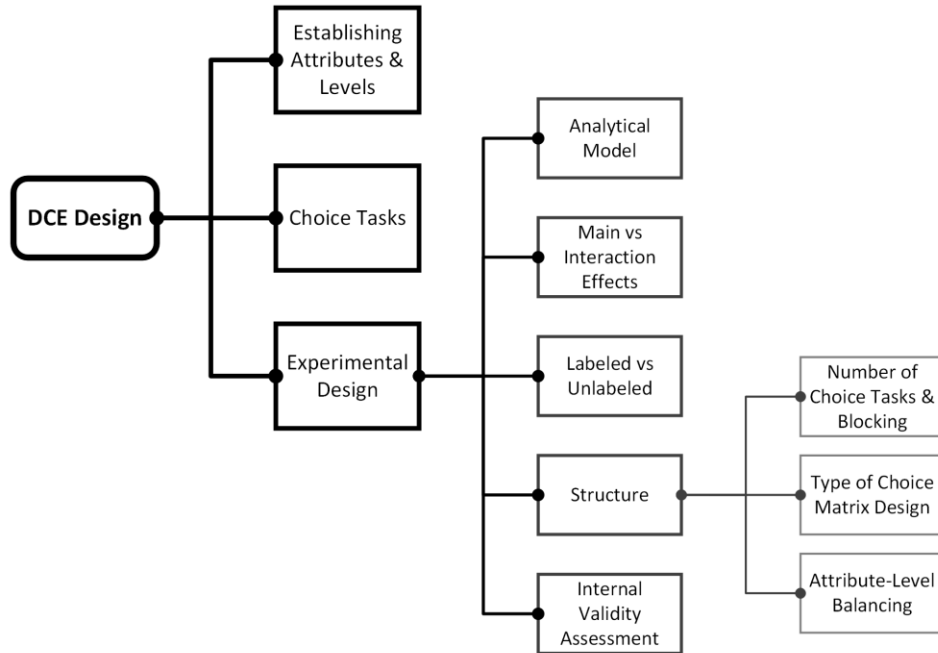
This section will provide an in-depth overview of the meticulous procedures involved in the experimental design of this study. The design of DCE will be discussed in detail, the variables of interest will be presented and discussed, and finally, the different stages of this study will be identified with justification.

### ***DCE Design***

According to Szinay et al. (2021), DCE design comprises of three main components: Establishing Attributes and Levels, Choice Task Design, and Experimental Design. Experimental Design further breaks down into 5 subcomponents: Analytical Model, Main Vs. Interaction Effects, Labeled Vs. Unlabeled Experiments, Structure, and Internal Validity Assessment (Szinay et al., 2021) (see Figure 4). This methodology guideline was referenced due to its recency and

**Figure 4:**

*Stages of DCE Design*



clear breakdown of various DCE components. Other guidelines and best practices in DCE design support the same approach (Bridges et al., 2011; Hess & Daly, 2014; F. R. Johnson et al., 2013; Lancsar & Louviere, 2008; Street et al., 2008).

To remove abstractness from this study and provide participants with a tangible product with realistic attributes and levels, I first needed to select a medical device for which to establish attributes and levels. Based on the findings of Study 1 discussed in Chapter II, blood pressure monitors (BPMs) were the most identified medical device used at home by the specific study population. Research also shows that BPMs are one of the most prevalent home medical devices (Cifter, 2017). Additionally, about half of US adults have hypertension, only 22.5% of these individuals have their hypertension under control, and hypertension was the leading or contributing cause of about 700,000 deaths in the US in 2021 (CDC, 2023). The prevalence of

hypertension increases substantially above the age of 50 (Ostchega et al., 2020). Monitoring blood pressure is a key aspect of managing hypertension (Glynn et al., 2010; Uhlig et al., 2013). Considering the factors above, BPMs appeared as promising devices, likely to be used by PwOD and PwD, and prevalent enough to serve as a proxy for medical devices in general.

**Establishing Attributes & Levels.** While designing a DCE, one key factor that determines the complexity of the design and the subsequent cognitive burden for the participants, is the number of attributes and levels. Typically, in healthcare research, between four to six attributes are identified, and it is generally advised to define between two to four levels per attribute (Bridges et al., 2011; Soekhai et al., 2019). Other DCE guidelines suggest including between five to seven attributes (Marshall et al., 2010), and no more than 10 (Mangham et al., 2009). In a systematic review of the literature, de Bekker-Grob et al. found that 71% of the reviewed studies included between four to six attributes (E. de Bekker-Grob et al., 2012). It is also recommended that attributes should have about three levels (Szinay et al., 2021). The total number of unique alternatives that can be produced based on the number of attributes and levels is equal to the product of level counts for all attributes (Lancsar & Louviere, 2008; Szinay et al., 2021).

For example, if a DCE has two attributes each with five levels, and three attributes each with four levels, the total number of unique alternatives for this DCE is calculated as:  $5^2 \times 4^3 = 1600$ . Presenting 1600 alternatives in this hypothetical DCE would place a significant cognitive and time burden on the participants. Thus, while identifying attributes and levels must follow a logical and evidence-based procedure, guidelines advise researchers to exercise caution and avoid unnecessarily inflating the number of attributes and levels they introduce in their DCE.

Several methods exist to efficiently reduce the cognitive burden of DCEs with many unique alternatives (such as fractional-factorial D-optimal designs) with minimal tradeoffs. These methods were utilized in this study, and they will be discussed in the continuation of this chapter. However, the first step was to establish attributes and levels effectively and appropriately.

To the best of my knowledge, no prior studies have yet evaluated the central hypothesis of this dissertation even in approximation, especially using a DCE. In fact, my extensive literature review did not reveal any use of this methodology in the context of product accessibility. While this methodology has been used in eliciting stated preferences towards access to healthcare services (Baji et al., 2012; Dubov et al., 2019; Oliver et al., 2019; Radley et al., 2019; Ratcliffe et al., 2009; Zickafoose et al., 2015) and rehabilitation delivery programs (Kjær & Gyrd-Hansen, 2008; Laver et al., 2013; Milte et al., 2013; Noto et al., 2023; Shields et al., 2021), the attributes and levels for these studies are substantially different from what is appropriate for medical devices and equipment. Consequently, consulting the existing literature to inform the selection of attributes and levels did not appear to provide a solid foundation for establishing the attributes and levels for this dissertation study.

I aimed to establish the attributes and levels based on the previously referenced guidelines and best practices, as well as the Mendonca (2010) dissertation, while attempting to retain only the core attributes and levels necessary for hypothesis evaluation. It must be noted that Dr. Mendonca's dissertation evaluates a similar hypothesis with a fundamentally different population. While the methodology and research questions of this dissertation may bear some resemblance to Mendonca (2010), the target population of that study entirely changes the

context of its hypothesis. As such, if used, the attributes and levels established by Mendonca (2010) must be re-evaluated for appropriateness and relevance to the context of this study.

I conducted extensive market research to identify the most prevalent attributes available for BPMs on e-tailer websites. After identifying these attributes, I examined their levels across e-tailers to determine a realistic and reasonable range for the attribute levels of this study. The purpose of this market research was to create realistic alternatives for the DCE, without overwhelming the participant with additional unnecessary or unrealistic information.

I visited four prominent e-tailers: Amazon, Best Buy, Walmart, and Walgreens, and searched the term “blood pressure monitor.” I did not log into any of these websites and maintained my visit anonymous to avoid personalized advertisements or search results. Without changing the sorting criteria, the first 100 devices from each e-tailer were catalogued (if more than 100 devices were available). BPMs that had additional functionalities (e.g., electrocardiogram or smart watch) were ignored to only include devices that are exclusively designed to measure blood pressure. Including these devices with additional features would have unreasonably skewed the levels for attributes such as cost. Also, only the original cost was catalogued if the device was on sale, as sales are unpredictable, and the discount amount may vary based on factors that are not typically disclosed by e-tailers. Some e-tailers presented different colors of blood pressure monitors individually, as opposed to presenting one product with the option of changing the color. In these scenarios, if the color variations had different attributes (cost, user ratings, etc.), I catalogued them as separate products since they would reasonably appear as different products to consumers. However, if the attributes were the same, only one instance of the product was catalogued, and the rest were omitted. This market

research took place on May 9, 2023. Table 2 shows the number of BPMs catalogued from each e-tailer, compared to the total number of search results reported by their website.

The findings of this market research showed that the most common attributes presented for BPMs are cost, brand, user rating, product images, product description, number of raters, and shipping information. To select an appropriate and relevant number of attributes, I attempted to narrow the list of prominent BPM features.

Product images are highly subjective in their appeal to different consumers. Generic product images can be constructed and presented to participants for each product. However, I decided against using images as an attribute in this study to avoid scenarios in which a participant favors one alternative but selects another solely due to aesthetics. Due to the novelty of this study’s hypothesis, it appeared logical to limit attributes to ones with objectively measurable qualities, and refrain from presenting highly subjective attributes which may potentially override preferences for other factors. Product descriptions can be very long and including them in the study may unnecessarily increase the cognitive burden for participants. The number of raters also varied greatly both between and within e-tailers and some devices

**Table 2:**

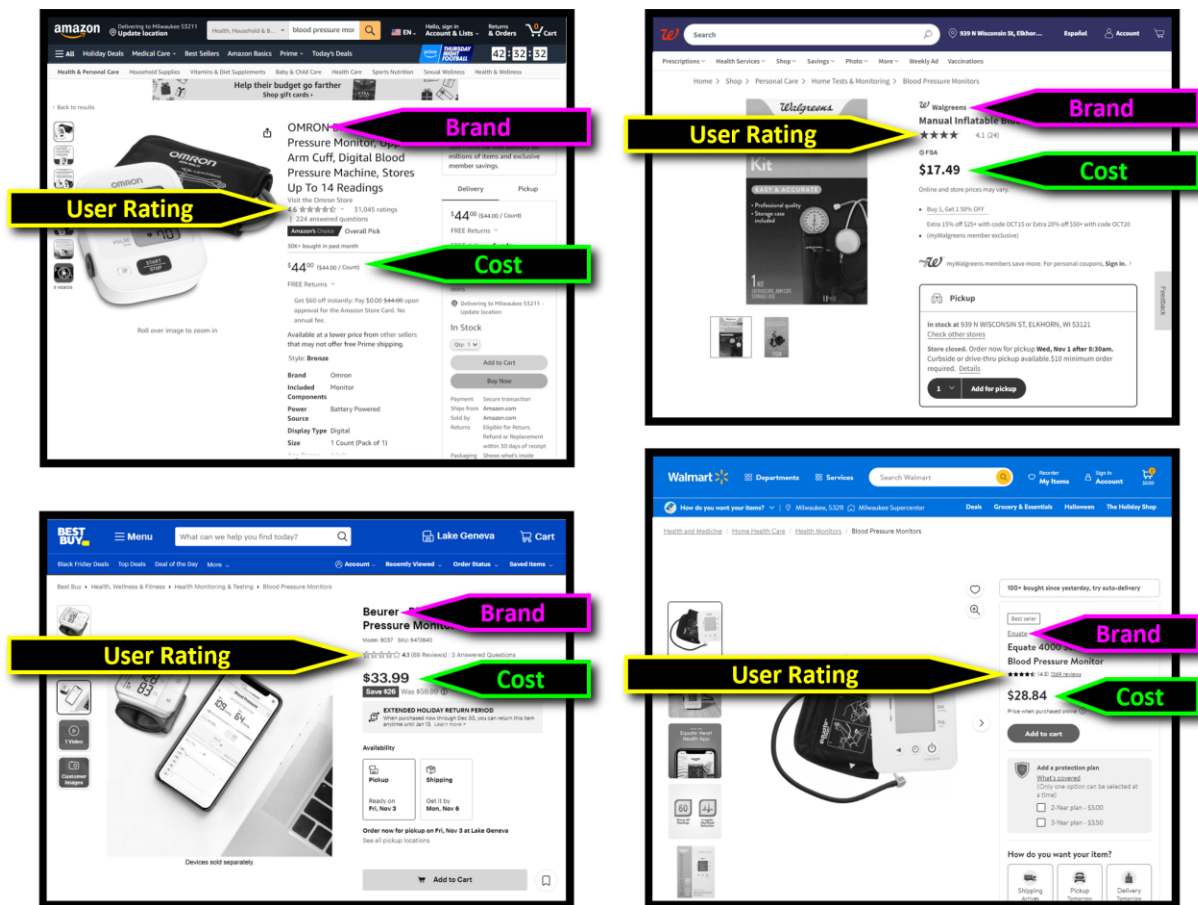
*Number of BPMs found and catalogued from 4 e-tailer websites*

E-tailer	Total BPMs Found	Unique BPMs Catalogued
Amazon	861	97
Best Buy	28	16
Walgreens	79	47
Walmart	>1000	91

*Note.* Data accurate as of 4/9/2023

had orders of magnitude more raters than others. Shipping information (such as speed and method of shipping) are subjective to the location of consumers. Additionally, some e-tailers also have brick-and-mortar stores (e.g., Best Buy) with pick-up options available, while others may not provide such an option. It appeared unreasonably difficult to establish a realistic range for shipping speed and delivery method across different e-tailers for various rural and urban locations in the US. Thus, I decided to only include Cost, Brand, and User Rating as the existing attributes of blood pressure monitors (see Figure 5).

**Figure 5:**  
*Most Prevalent Attributes of BPMs Across Four E-tailers*



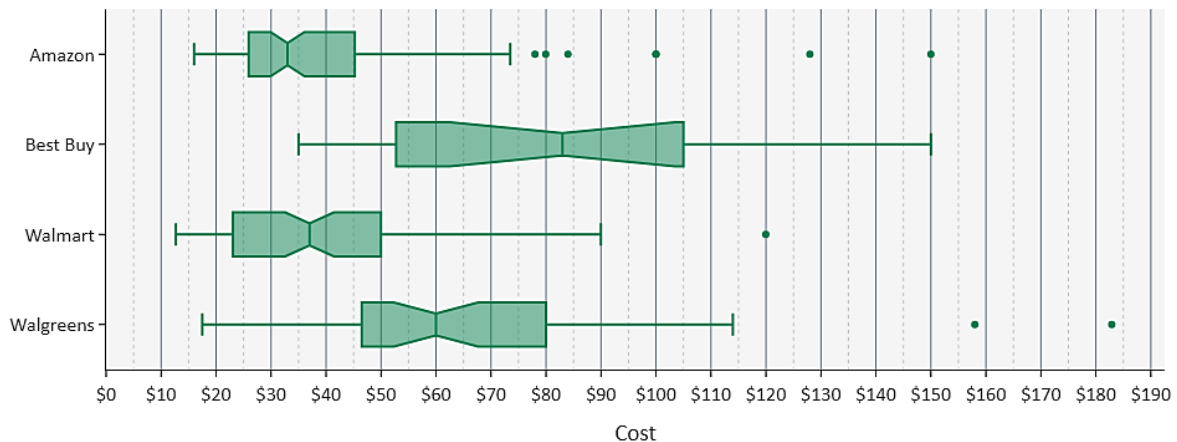
Analyzing the cost range for BPMs showed large variability across different e-tailers (see Figure 6). To select a reasonable range without causing a floor/ceiling effect, I decided to target between the \$20 and \$30 budget range. The average of this range was also used as the medium cost level, and the numerical levels for Cost were defined as \$20, \$25, and \$30.

The market research data showed that the most commonly available brand between e-tailers was Omron. This brand was consistently available on all four e-tailer websites. It also had the highest number of BPMs available, with the exception of Walmart where it placed second with the difference of only one product. To study how this attribute impacts the purchasing decisions of PwOD, I included “HomeHealth,” a fictitious brand name, as the second level of this attribute. By creating two categorical levels for Brand, one introduced as a “trusted” brand (Omron), and one as a “newly established” brand (HomeHealth), I aimed to examine the impact of brand.

The distribution of user rating on all e-tailers showed that 3.8 stars can be considered as a low rating and 4.8 stars is high. The average of these two ratings, 4.0 stars, was considered a

**Figure 6:**

*Variations of BPM Cost Across Four E-tailers*

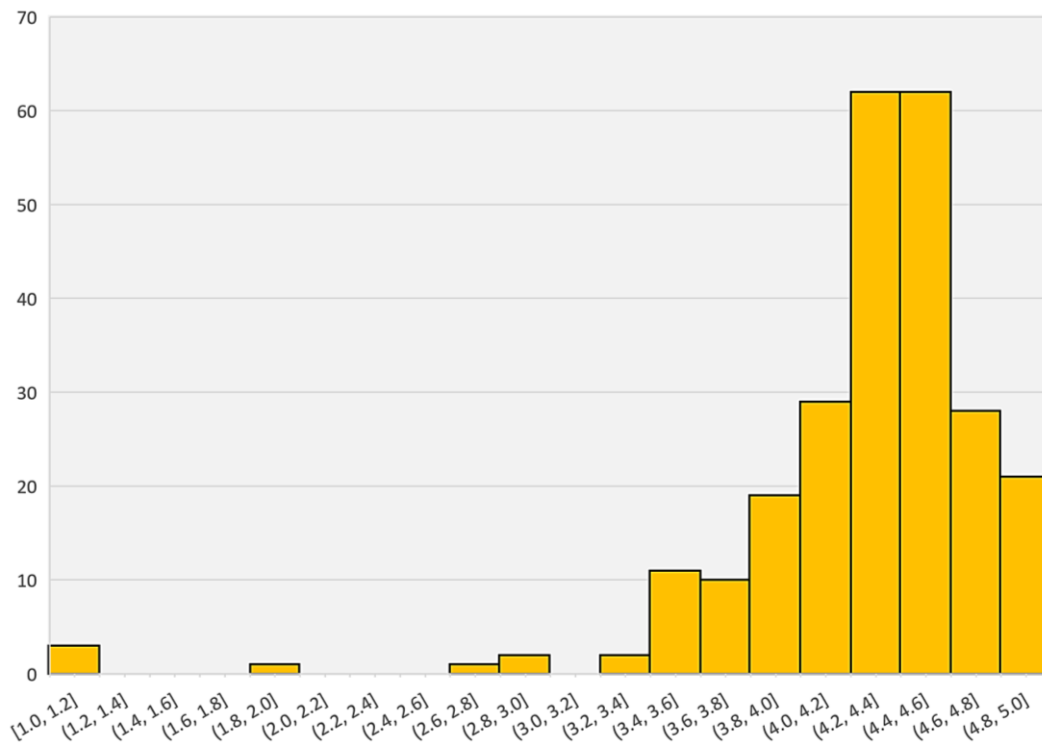


medium rating. These levels for User Rating also correspond to the levels used by Szinay et al. (2021). Based on the data, the numerical levels for User Rating were defined as 3.2 stars, 4.0 stars, and 4.8 stars. Figure 7 shows the distribution of user ratings cumulatively across all four e-tailers.

To evaluate the central hypothesis of this study, Usability/Accessibility Information was also included in the attributes. However, to simplify the presentation and reduce cognitive burden on participants, I opted to use a Usability/Accessibility Score. This attribute was introduced to the participants as “an independent expert evaluation of how easy the device will be to use.” By definition, higher usability and accessibility result in easier device operation (as discussed previously under Definition of Accessibility and Usability). I selected this wording to

**Figure 7:**

*Cumulative Distribution of BPM User Ratings Across Four E-tailers*



associate usability and accessibility with ease of use, which was also informed by Study 2 in Chapter II.

To establish the levels for Usability/Accessibility Score, I decided to adopt the benchmarks for System Usability Scale (SUS) (Brooke, 1995). SUS is the most prominent standardized assessment for evaluating perceived usability and has been used and validated extensively for decades (Lewis, 2018). It typically consists of ten questions presented to users, each rated on a Likert-type scale from 1 to 5, where 1 represents strong disagreement and five represents strong agreement. These questions are typically a mix of positive and negative statements, and users provide their ratings based on their experience with the system. After collecting responses, the SUS scores are calculated by combining and averaging the user's ratings and then multiplying by 2.5 to convert the scores to a scale of 0 to 100. Higher SUS scores indicate better perceived usability, while lower scores suggest more usability issues.

A benchmarking study conducted by examining the data collected from SUS over a decade, provided invaluable insights into the interpretation of different scores and their implications (Bangor et al., 2008). According to Bangor et al. (2008), SUS scores around 40 represent poor usability, while scores around 60 represent medium usability, and scores around 80 represent high usability. By adopting these benchmarks, I defined 4/10 as a low Usability/Accessibility Score, 6/10 as medium, and 8/10 as high. To evaluate the impact of information availability or unavailability, I also added an "Unavailable" level to this attribute. As such, the categorical level for Usability/Accessibility Score were identified as Unavailable, Low (4/10), Medium (6/10), and High (8/10).

As established above, a summary of attributes and levels is presented in Table 3. These attributes and levels were established based on extensive market research and literature review. However, as discussed previously under “A Critical Prerequisite for DCE Design,” I still needed to confirm that these values are meaningful to participants. Validating the appropriateness of these attribute levels was one of the foci of the Phase I study, detailed later in this chapter. The structure and findings of these studies will be thoroughly discussed in the continuation of this chapter, as well as the next.

**Choice Task Design.** According to Szinay et al. (2021) once the attributes and levels are established, the next step is to decide on a full-profile or partial-profile design and the opt-out option. In a full-profile design, all attributes will be presented in each alternative in a choice task. In contrast, a partial-profile design only presents select attributes in different alternatives. The opt-out option allows participants to refrain from making a choice in a choice task.

Since there are less than five attributes in this study, it is appropriate to use a full-profile design as it will not impose an undue cognitive load on the participants and will maximize the extracted choice information (Mühlbacher & Johnson, 2016). Considering that the purpose of

**Table 3:**

*Summary of Dissertation Study Attributes and Levels*

Attribute	Levels	Data Type
Cost	\$20, \$25, and \$30	Numerical
Brand	Omron, HomeHealth	Categorical
User Rating	3.2, 4.0, and 4.8 Stars	Numerical
Usability/Accessibility Score	Unavailable, Low (4/10), Medium (6/10), and High (8/10)	Categorical

this study is to compare attributes and levels of different alternatives, providing an opt-out option is unnecessary (Milte et al., 2013; Nieboer et al., 2010). While presenting an opt-out function may be realistic in some scenarios, it limits the amount of information that can be collected by examining tradeoffs (Mangham et al., 2009), and there is no consensus on how the choice model should be constructed in the presence of opt-out options (Campbell & Erdem, 2019; Veldwijk et al., 2014). Therefore, the choice task design in this study will be full-profile with no opt-out option.

**Experimental Design.** In DCEs, experimental design is the systematic approach that will ultimately result in creating the choice tasks presented to respondents (F. R. Johnson et al., 2013; Mangham et al., 2009). According to Szinay et al. (2021), experimental design steps in DCEs include identifying the appropriate statistical model that will be used to analyze the choice data, determining whether main effects will be estimated with or without interaction effects, and deciding whether the design is labeled or unlabeled. Other considerations include number of choice tasks, blocking, choice matrix design strategy, attribute level balance, and internal validity assessment procedures (Szinay et al., 2021). These experimental design components for DCEs are corroborated strongly by a multitude of other reputable references (Bridges et al., 2011; Hess & Daly, 2014; F. R. Johnson et al., 2013; Lancsar & Louviere, 2008; Mariel et al., 2021; Pérez-Troncoso, 2020; Street et al., 2008; van den Broek-Altenburg & Atherly, 2020). Deliberate and accurate configuration of DCE design parameters will enhance data quality and improve the validity and reliability of insights obtained from the choice model (Mariel et al., 2021; Pérez-Troncoso, 2020). The experimental design approach I employed in this study, in correspondence with DCE design guidelines, is detailed below.

**Analytical Model.** Random Utility Theory provides the framework for analytical choice modeling in DCEs (McFadden, 1974; Watson et al., 2020). As described previously, this theory views the utility ( $U$ ) of the choice  $j$  for individual  $i$  as the aggregation of a deterministic component ( $V_{ij}$ ) and a stochastic component ( $\varepsilon_{ij}$ ) (Lancsar & Louviere, 2008). Equation (1) shows how this utility function is calculated, based on the Lancsar & Louviere (2008) presentation.

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1)$$

In (1),  $\varepsilon_{ij}$  represents random choice variability due to unmeasured, inexplicable, or unobservable factors, endogenous or exogenous, within or between individuals (Adamowicz et al., 2008; Ben-Akiva & Lerman, 1985; Lancsar & Louviere, 2008). On the other hand,  $V_{ij}$  as the deterministic component can be estimated using an appropriate analytical model. By far the most common analytical model to estimate this deterministic component, especially in health-related fields, is the Multinomial Logit (MNL) model (E. de Bekker-Grob et al., 2012; Hensher et al., 2005; Kjaer, 2005; Soekhai et al., 2019; Szinay et al., 2021; Train, 2009).

Using the MNL model to describe the deterministic component, the overall utility function ( $U$ ) for alternative  $j$  will be a function of the attributes vector ( $X_j$ ) and part-worth utilities vector ( $\beta$ ). With the overall stochastic variabilities consolidated as  $\varepsilon$ ,  $U_j$  can be estimated as:

$$U_j = \beta X_j + \varepsilon \quad (2)$$

Considering the attributes established previously (see Table 3), (2) can be further specified for alternative  $j$  from this study:

$$U_j = \beta_{Cost}x_{Cost,j} + \beta_{Brand}x_{Brand,j} + \beta_{UserRating}x_{UserRating,j} + \beta_{UsabilityAccessibilityScore}x_{UsabilityAccessibilityScore,j} + \varepsilon \quad (3)$$

In (2), and by extension (3), vector  $\beta$  is estimated via MNL analysis of DCE choice data. In (3),  $x_{Cost,j}$  represents the specific value of the Cost attribute in alternative  $j$ ,  $x_{Brand,j}$  specifies the Brand for alternative  $j$ , and so forth. Once vector  $\beta$  is estimated from the analysis, the utility of alternative  $j$  (in this study a BPM with specific attribute levels) can be estimated from (3) and, for example, compared to the utility of alternative  $k$  to determine which BPM is likely to be perceived as more appealing by the consumers. Vector  $\beta$  holds coefficients that determine the magnitude and direction of impact for each attribute. These coefficients help estimate which attributes impact decisions more than others, and whether they increase or decrease the likelihood of selection as attribute levels change.

MNL models are estimated using Maximum Likelihood Estimation (MLE) (Train, 2009). MLE finds the values of model parameters that make the observed data most probable (Rossi, 2018). In other words, MLE chooses the parameters that maximize the likelihood that the process described by the model would produce the observed data. After conducting a DCE, the likelihood function is constructed based on the probabilities of the observed choices across all individuals and choice tasks (Roh & Khan, 2013). It is the product of the individual probabilities for the choices that were actually made (Roh & Khan, 2013). Because the likelihood can be a very small number due to the product of probabilities, it is common to take the natural logarithm of the likelihood function, turning the product into a sum (Berbeglia et al., 2018). This log likelihood is what will be maximized.

Maximizing log likelihood is performed by testing different coefficients in vector  $\beta$  and examining how they impact the log likelihood. This trial-and-error is conducted algorithmically using gradient descent until optimal  $\beta$  coefficients are found (Helveston, 2023b). After the optimization process and once the model is estimated, the reports often include the log likelihood (calculated using the optimal  $\beta$  coefficients) as well as the null log likelihood. Null log likelihood is calculated by assuming that all coefficients are zero, which is to say that none of the attributes have any impact on the outcome (Kjaer, 2005). Comparing the null log likelihood and log likelihood helps evaluate the model's goodness-of-fit measures, such as McFadden's  $R^2$  (T. J. Smith & McKenna, 2013).

There is a noteworthy distinction between the definition and interpretation of the conventional  $R^2$  and McFadden  $R^2$  (also known as Pseudo  $R^2$ ). In linear regression,  $R^2$  represents the proportion of variance in the dependent variable that is predictable from the independent variables (Zhang, 2017). McFadden  $R^2$ , on the other hand, compares the likelihood of the model under consideration to a baseline null model, which is the model with no predictors beyond an intercept (Hemmert et al., 2018). McFadden  $R^2$  is calculated as 1 minus the ratio of the log likelihood of the model to the log likelihood of the null model (Kjaer, 2005). In other words, McFadden  $R^2$  shows how well the estimated model is predicting the outcomes compared to a model that assumes there is no relationship between the predictors and outcomes. According to McFadden, values between 0.2 and 0.4 "represent an excellent fit" (McFadden, 1977). This range for McFadden  $R^2$  is shown in simulations to be similar to the linear regression  $R^2$  values between 0.7 and 0.9 (Domencich & McFadden, 1975; J. Louviere et

al., 2000; Ugba & Gertheiss, 2023). To penalize overfitting, a variation of McFadden  $R^2$  termed Adjusted McFadden  $R^2$  can be used (Walker & Smith, 2016).

Additionally, the Akaike Information Criterion (*AIC*) and Bayesian Information Criterion (*BIC*) are commonly used for MNL model selection, each with unique properties and implications. The *AIC* focuses on the tradeoff between the model goodness-of-fit and its complexity, and does not inherently assume that a true model exists within the estimated models (Vrieze, 2012). It is particularly efficient in selecting a model that minimizes the expected divergence between the estimated and true model, making it a robust choice when the true model is complex or unknown (Vrieze, 2012). The *BIC* is derived from Bayesian probability, and models with more parameters are penalized more heavily compared to the *AIC*, thus favoring simpler models (Acquah, 2010). It is consistent in selecting the true model as the sample size grows (Acquah, 2010; Kasali & Adeyemi, 2022). The *AIC* and *BIC* values are typically included as a part of MNL model estimation reports.

One important reason behind the popularity of MNL models in estimating DCEs is that the log likelihood function for MNL models is globally concave, meaning that a global solution will always be found by MLE (Helveston, 2023b). In other words, MNL models have a closed-form solution (Kjaer, 2005). According to Helveston (2023), utility models such as Mixed Logit lack a concave log likelihood function and are less likely to have a global solution. In addition, MNL models are very robust and simpler to implement and interpret (Haan, 2004). However, the tradeoff for this simplicity and robustness is a few restrictive assumptions (Soekhai et al., 2019; Szinay et al., 2021):

- Independence of Irrelevant Alternatives: IIA was discussed previously under Luce's Choice Axiom. This assumption implies that the relative odds of choosing between any two alternatives are unaffected by the presence or absence of other alternatives (Arrow, 2012). In MNL model estimation, this assumption can become problematic if the stochastic components of utility are correlated between different alternatives. Violations of this assumption lead to biased parameter estimation, and typically manifest as poor model fit (Y. Lin et al., 1988). One way to avoid violating this assumption is to use unlabeled alternatives in the DCE such that if alternative  $j$  and  $k$  have exactly the same attribute levels, participants can be reasonably assumed to have no reason to prefer one over the other (Lancsar et al., 2017). In other words, the sole differentiator between two alternatives must be the differences between their attribute levels. For example, instead of introducing interventions with their names (e.g., Occupational Therapy or Inpatient Surgery), they can be presented with generic labels (e.g., Treatment A or Treatment B), since labels may carry preconceived notions that can adversely impact the IIA assumption.
- Independent and Identical Distribution (IID) of Errors: This assumption implies that the stochastic component of utility functions (also referred to as error terms) are distributed independently and identically across all alternatives and participants (Soekhai et al., 2019). In practice, these unobserved factors for different alternatives might be correlated. Similar to IIA, violations of this

assumption can also manifest as bias in parameter estimates and poor model fit (Xie et al., 2012).

- Homogeneity of Preferences: Historically, DCEs have been used to provide a summary model of sample preferences, with the underlying assumption that all individuals respond homogeneously to changes in the independent variables (Soekhai et al., 2019). This assumption implies that the preferences of individuals are consistent across the population. However, preferences can be heterogeneous, and given the same choice task, individuals may choose different alternatives under different circumstances (Vass et al., 2022). MNL models assume homogeneity of preferences and provide a summary report of the study population's preferences. While this assumption is not inherently problematic, if the goodness-of-fit for an MNL model is poor, models that allow for preference heterogeneity can be tested as alternatives. As such, it is recommended that researchers collect personal characteristics information along with the DCE (Lancsar et al., 2017). In addition, if there are only a few parameters to estimate compared to the number of choice tasks, this assumption becomes justifiable (Kessels, 2016).

These assumptions and their implications must be considered in MNL model estimation to avoid bias and erroneous conclusions. However, as explained above, IIA and homogeneity of preferences violations are preventable to a significant degree with appropriate study design. Additionally, if the MNL model explains a considerable portion of the utility function (as determined by goodness-of-fit measures), violating the IID and IIA assumptions typically has

negligible consequences (Vojáček & Pecáková, 2010). One of the most distinguished scholars in decision analysis and the inventor of the DCE methodology (J. J. Louviere & Woodworth, 1983), the late Dr. Jordan Louviere, posits that MNL models are effective and robust even with these underlying assumptions (J. J. Louviere, 2001). With all the above considered, I chose MNL as the analytical model to design my experiment.

**Main and Interaction Effects.** From 628 reviewed DCEs in health economics between 1990 and 2017, about 49% exclusively estimated the main effects, while only 13% reported estimating main and two-way interaction effects, and the majority of the remaining studies did not report their design specifications (Soekhai et al., 2019). Due to the novel nature of this dissertation's central hypothesis, no empirical evidence yet exists to support the existence of interaction effects between the attributes of this study with PwoD participants. I chose to estimate only the main effects. However, various interaction effects were explored a posteriori if the performance metrics indicated poor model fit.

**Labeled or Unlabeled.** Labeled designs refer to choices where each alternative is given a specific name that impacts the inherent characteristics of the alternative, such as "Clinic" and "Hospital" visits, or "Laptop" and "Desktop" computers. These labels help respondents anchor choice tasks in real-world contexts (Thai et al., 2023). Unlabeled designs, on the other hand, present alternatives without such identifiers, focusing instead on the attributes and levels of the choice tasks. For example, an unlabeled DCE may ask participants to choose between different healthcare programs based solely on attributes like cost, waiting time, and treatment outcomes, without naming the programs. This approach is more abstract and focuses on the tradeoffs between attributes, which can be useful for understanding preferences in a more

general sense, without the influence of preconceived notions tied to specific labels (Bliemer & Rose, 2014; Lancsar & Louviere, 2008; Szinay et al., 2021). In this experiment, since the alternatives are only BPMs with different attribute levels, an unlabeled design is appropriate. In addition, the unlabeled design helps mitigate concerns about the violation of IIA assumption, as discussed above.

**Structure.** After establishing the parameters above, the choice matrix structure can be designed. The choice matrix structure determines the combinations of alternatives that create each choice task. This is a very important stage of the DCE design process as it determines what will be presented to the participants, how much cognitive burden will be imposed, and what information can ultimately be extracted from the choice data. The factors contributing to appropriate choice matrix design will be discussed in this section.

- **Number of Choice Tasks and Blocking:** Various guidelines and recommendations exist to help identify the appropriate number of choice tasks with the objective of maximizing choice information and minimizing task complexity (Lancsar & Louviere, 2008). Bliemer & Rose (2014) suggest that if the number of alternatives in each choice task is  $J$  and the objective is to estimate  $K$  parameters, the minimum required number of choice tasks can be identified by finding the smallest integer  $S$  that satisfies the equation below.

$$S \geq \frac{K}{J-1} \quad (4)$$

The degrees of freedom is calculated by subtracting  $S$  as identified in (4) from the actual number of choice tasks, and to ensure sufficient degrees of freedom,

the number of choice tasks must be at least equal to  $2S$  or  $3S$  (Bliemer & Rose, 2014).

Marshall et al. found that DCEs in healthcare typically include between 7 to 16 choice tasks (Marshall et al., 2010), which is in agreement with other recommendations (Hensher et al., 2001). Christofides et al. and Hanson et al. recommend not to exceed 18 choice tasks (Christofides et al., 2006; Hanson et al., 2005).

If the number of choice tasks exceed the recommended sizes, blocking can be used to break down the choice matrix into evenly sized partitions and reduce the number of choice tasks per participant (Bliemer & Rose, 2014; Soekhai et al., 2019). It must be noted that the number of choice tasks in each block should be evenly divisible by the number of attribute levels, and blocking increases the sample size requirements (Szinay et al., 2021).

Increasing the number of alternatives in each choice task allows for capturing more choice data from each choice task while also increasing the task complexity (Bliemer & Rose, 2014). Evidence suggests that the one of the most influential choice matrix characteristics in terms of the impact on unobserved variability, is the number of alternatives (Caussade et al., 2005; DeShazo & Fermo, 2002). For unlabeled DCEs, between two to four alternatives per choice task are deemed appropriate (Bliemer & Rose, 2014), with evidence suggesting that presenting four alternatives per task is more desirable than presenting three or five due to scale factors (Caussade et al., 2005).

Based on the attributes and levels established for this dissertation research, the number of factors to be estimated for main effects can be determined. Cost and User Rating are numerical variables and require one factor estimate each. Brand is a categorical attribute with two levels. By selecting one level as reference for coding, one factor estimate will be required for estimating the part-worth utility of the other. Similarly, Usability/Accessibility Score as a categorical attribute has four levels and requires three factor estimates. Thus,  $K$  in (4) for this study is calculated as  $1 + 1 + 1 + 3 = 6$ . Based on the evidence described above, I chose to include four alternatives per choice task, which means in (4),  $J = 4$ . Based on these values for  $J$  and  $K$ , the minimum required choice tasks  $S$  is calculated as shown below.

$$S \geq \frac{6}{4 - 1} = 2 \quad (5)$$

To accommodate for the degrees of freedom requirements:

$$S \geq 3 \times 2 = 6 \quad (6)$$

Considering the minimum choice task number found in (6), and to comply with the evidence-based recommendations above, I chose to include 16 choice tasks in this study. This ensures that blocking will be unnecessary, and the task complexity is controlled in accordance with the literature recommendations while capturing as much choice information as possible.

- Type of Choice Matrix Design: Another crucial facet of choice matrix design is identifying whether a full-factorial design will be used or a fractional-factorial design. A full-factorial design includes all possible permutations of attribute level

combinations (Bliemer & Rose, 2014). While full-factorial designs facilitate independent estimation of all main and interaction effects, they typically impose a high cognitive burden on participants due to their large number of choice tasks (Lancsar & Louviere, 2008). In contrast, fractional-factorial designs use a systematic approach to select a subset of the full-factorial design to reduce cognitive burden, while still facilitating the estimation of necessary main and interaction effects (Street et al., 2008).

Fractional-factorial structures can be created using random, orthogonal, or efficient design approaches (Bliemer & Rose, 2014). A randomized design creates different choice matrices for different participants by randomly selecting choice tasks from the full-factorial design. According to Bliemer & Rose (2014), the main advantage of this approach is that it requires minimal experimental design skills and if the study is presented to a large enough sample size, any model estimation from the choice information can be successful. However, Bliemer & Rose note that the main flaw of this approach lies in its need for a very large sample size which must include at least about 1,000 participants. This is due to the fact that randomized designs require no prior plans for the utility function design, and therefore use no systematic approach is employed to optimally select the choice sets in a way that facilitates the estimation of the desired main and interaction effects (Bliemer & Rose, 2014). In summary, while random designs are easy to generate, they are also highly inefficient in their creation, and require very large sample sizes for effectiveness.

An orthogonal design, in contrast to random designs, is a systematic approach to choice matrix generation based on orthogonal arrays (Hedayat et al., 1999). Orthogonal design features are (a) statistical independence of attributes (orthogonality); and (b) equal representation of all attribute levels across the choice matrix (level balance) (Szinay et al., 2021). Orthogonal arrays are predesigned and can be found in reference tables in textbooks (Hahn & Shapiro, 1967), online (Sloane, 2007), or by using statistical software packages to retrieve known designs based on the number of attributes and levels (Groemping et al., 2023). These predetermined orthogonal arrays make generating the choice matrix very simple and reduce the experimental design proficiency required, while also preventing underrepresentation of some attributes and levels (Bliemer & Rose, 2014). Some have alleged that orthogonal designs allow for independent parameter estimation, however, this is not applicable to choice experiment modeling and pertains only to linear regression models (Bliemer & Rose, 2014). The downsides of using orthogonal designs are (a) only certain combinations of attribute and level numbers can have orthogonal array representations (Szinay et al., 2021); and (b) by definition they do not accommodate for imposing restrictions and removing undesirable alternatives (Bliemer & Rose, 2014; Street et al., 2008). Orthogonal designs have been the de facto strategy in generating fractional-factorial choice experiments for many decades (Kuhfeld, 2006). However, while acknowledging design orthogonality as an important concern, scientists have posited that ensuring realism, minimizing

cognitive burden, and reducing parameter estimate variances should take precedence over adherence to rigid orthogonal design constraints (Bateman et al., 2002; J. Louviere et al., 2000).

Efficient designs prioritize reducing parameter estimate variances while slightly relaxing the strict requirements of orthogonal designs and allowing for minimal correlation between attributes (F. R. Johnson et al., 2013; Lancsar & Louviere, 2008; Szinay et al., 2021). If the objective is to estimate  $K$  parameters using the MNL model, Fisher information matrix  $\mathbf{F}$  will be a  $K \times K$  matrix constructed using choice probabilities and attribute levels, irrespective of actual choice observations (Bliemer & Rose, 2014; McFadden, 1974). The variance-covariance matrix of parameter estimates  $\mathbf{\Omega}$  then equals the inverse of the Fisher information matrix (Bliemer & Rose, 2014; Lancsar & Louviere, 2008).

$$\mathbf{\Omega} = \mathbf{F}^{-1} \quad (7)$$

The goal of efficient designs is to maximize the amount of Fisher information by maximizing the volume of  $\mathbf{F}$ , which is tantamount to minimizing the variance-covariance matrix volume and reducing parameter estimate variances (Bliemer & Rose, 2014). The volume of  $\mathbf{F}$  is equal to the determinant of  $\mathbf{F}$ , and the commonly used efficiency measure D-error is thus defined as (Bliemer & Rose, 2014):

$$\text{D-error} = \left( \frac{1}{\det(\mathbf{F})} \right)^{\frac{1}{K}} \quad (8)$$

Consequently, to maximize the volume of  $\mathbf{F}$ , D-error must be minimized.

As can be seen from (8), the value of D-error is specific to the design and cannot be compared across different studies with different analytical models and choice task characteristics (Bliemer & Rose, 2014). The “lowest D-error” is also illogically difficult to determine. For example, based on the established attributes and levels,  $3 \times 2 \times 3 \times 4 = 72$  unique alternatives can be created for this study. This means that if each choice task has four alternatives,  $72^4 = 270,935,376$  unique choice tasks can be created (permutations with replacement). To design the choice matrix for this dissertation by selecting 16 choice tasks (as determined previously), the required number of D-error calculations is:

$$P(270935376, 16) = 843,045,937,389,405,191,267,959,839,010,909,322,728,871,223,863,637,912,114,406,760,376,484,003,124,434,823,817,913,162,341,455,720,535,927,513,328,571,053,768,860,401,971,200,000 \text{ calculations}$$

When I designed the choice matrix for this study, each reiteration took roughly one minute to compute on a relatively powerful 10-core (20 threads) and overclocked 5 GHz processor with all threads fully engaged. To carry out the number of calculations required to find the lowest D-error on this processor, the number of years it would take is approximated as:

$$P(270935376, 16) \div (60 \times 24 \times 30 \times 12) = 162,624,602,119,869,817,481,800,629,967,589,606,850,009,819,270,951,835,361,487,050,439,863,263,185,055,876,194,5886,721,496,211,583,478,608,104,160,522,331,827,625,525,248 \text{ years}$$

Based on the subjectivity of D-error to each experimental design, and the difficulty of finding the lowest D-error for a design as demonstrated above, the only benchmark that can be offered for the evaluation of D-error is “the lower the better”, and values smaller than 1 are typically considered desirable (Bliemer & Rose, 2014).

Choice matrices that are created following the D-error minimization procedure are termed D-Efficient designs (F. R. Johnson et al., 2013; Lancsar & Louviere, 2008; Szinay et al., 2021). These designs are capable of (a) virtually maximizing the amount of choice information gathered from participants; (b) minimizing the parameter estimate errors; and (c) being used with smaller sample sizes compared to other designs (Bliemer & Rose, 2014). The downsides of D-Efficient designs are (a) the steps involved in their design are complex and require computational algorithms; and (b) their success hinges on appropriate estimation of the choice probabilities (Bliemer & Rose, 2014).

As mentioned earlier, the Fisher information matrix is generated using choice probabilities. These probabilities or their estimates, as well as the analytical model and utility function, must be introduced to the design algorithm to facilitate appropriate Fisher information matrix design (Szinay et al., 2021). To identify these choice probabilities, educated guesses about the magnitude and sign of the unknown study parameters must be made (Bliemer & Rose, 2014). These estimates of the sign and magnitude of parameters are called “priors”,

and three types of D-Efficient designs can be constructed based on how the priors were estimated (Szinay et al., 2021):

- $D_z$ -Efficient, where no priors are available, and the design algorithm assumes the parameters are zero.
- $D_p$ -Efficient, where the priors are deterministically estimated by a fixed value and sign.
- $D_b$ -Efficient, where the priors are not deterministically identified, however their probability distribution is available. Due to their use of prior probability distributions, these designs are known as Bayesian designs.

Given the appropriate priors, D-Efficient designs perform noticeably better than orthogonal designs (Lancsar & Louviere, 2008). However, novel experiments may not be able to establish reliable estimates for priors due to the scarcity of the literature around their topics. The best practice in such cases is to first conduct a small-scale study (or pilot) using a  $D_z$ -Efficient design to avoid making inaccurate assumptions about the magnitude and sign of the parameters (Szinay et al., 2021). Then, using the parameter estimates identified in the first phase, the large-scale study can be designed using a  $D_b$ -Efficient approach (Szinay et al., 2021). This is the approach I elected to follow for this study.

- **Attribute Level Balance:**

The last step in identifying the DCE structure is ensuring that all attribute levels appear an equal number of times in the experimental design, which is

paramount for maintaining statistical efficiency and reducing bias in the estimation of preferences. This principle, known as attribute level balance, is crucial to minimizing the risk of confounding effects, where the influence of one attribute on the choice is mistakenly attributed to another (J. Louviere et al., 2000). When attribute levels are not balanced, certain levels may be overrepresented, leading to an overestimation of their importance in the model (Hensher et al., 2005). Moreover, balanced designs facilitate the estimation of main and interaction effects with greater precision, as each level's contribution to the utility function is observed in different contexts (Kuhfeld, 2010). Design balance and proper overlapping also aids in reducing cognitive fatigue and choice task complexity, thereby enhancing the quality of the collected data (Bonoli Escobar et al., 2013; Jonker et al., 2019; Meyerhoff et al., 2015). Based on these factors, attribute level balance in DCE is a critical design feature that ensures the reliability and validity of the inferred preferences, ultimately enhancing the accuracy and generalizability of the conclusions.

The computational method I employed in choice task design creates an initial choice matrix by randomly sampling the full-factorial design and assigning alternatives to various choice tasks. Using the Fedorov algorithm (Fedorov, 1972), the program then iterates upon the initial design by randomly exchanging alternatives. This process is repeated until either the maximum number of permitted iterations are exhausted, or program arrives at the threshold beyond which no further beneficial exchanges are possible (Wheeler & Braun, 2022). In

D-Efficient designs, as discussed previously, the objective is to maximize the volume of Fisher information (or minimize D-error). The program uses this metric to iteratively identify the most performant design, given the imposed constraints (Wheeler, 2009).

The optima found in this iterative process are local, and their proximity to the global optimum is influenced by the starting design. The starting design depends on random number generation to sample the full-factorial design. As with any random number generation software, the numbers are not truly random and the procedure follows a pattern after starting with a “seed” (Rubin, 2011). By manually identifying this seed, the choice matrix design algorithm creates the same design consistently.

While attribute level balance is not the primary objective of D-Efficient design procedure, based on the emphasis placed on its significance in many different sources, I decided to create multiple D-Efficient designs and manually inspect their balance. My priority was minimizing D-error without sacrificing attribute level balance, within a set of designs that I can reasonably generate (finding “the lowest D-error” is impractically time consuming, as discussed earlier). This was accomplished by generating different designs in a loop, recording the seeds used for reproducibility, then ranking all generated designs based on their D-error and attribute level balance.

Another facet of attribute level balance is ensuring that tradeoffs will be made in each choice task. If a choice task includes an alternative with the most desirable

attribute levels compared to all other alternatives in the choice task, any reasonable participant would select the most desirable alternative without making any tradeoffs with other alternatives. These alternatives are referred to as “Dominant Alternatives” in the literature (Szinay et al., 2021). Including Dominant Alternatives in the design provides the least amount information on preferences. If the design does not account for such alternatives, they may ultimately harm the accuracy of conclusions. I excluded the Dominant Alternative in the choice matrix design process. However, if used methodically, Dominant Alternatives can help increase the validity of the findings. Along with others, this technique will be discussed in the following section.

***Internal Validity Assessment.*** As a SP methodology, the validity of DCE implications may be scrutinized. Evaluating and reporting the external validity of DCEs may be difficult or expensive due to the hypothetical nature shared by all SP methods. In contrast, the internal validity of the study can be evaluated very effectively, using various techniques. However, this remains one of the least examined or reported aspects of DCEs (Clark et al., 2014; Soekhai et al., 2019). Considering the importance of this aspect of DCE design and its implications for the findings, I elected to assess internal validity using two methods: Within-Set Dominated Pairs and Stability Validity tests.

Within-Set Dominated Pairs tests, also known as non-satiation, intentionally include a choice task with a Dominant Alternative and observe whether participants select the Dominant Alternative over other options (F. R. Johnson et al., 2019; Szinay et al., 2021). Stability Validity tests repeat a choice task later in the DCE and examine whether participants remain consistent

in their decision-making (F. R. Johnson et al., 2019; Szinay et al., 2021). While the usage of Within-Set Dominated Pairs tests is reported more frequently in DCE literature than most other tests (Soekhai et al., 2019), Stability Validity tests are also fairly common in this context (Pearce et al., 2021). Both of these tests allow for the examination of rational decision-making, a fundamental assumption in SP methods (Pearce et al., 2021). However, failing the internal validity tests may not be sufficient grounds for data exclusion, and no consensus exists in the literature on how to approach the data from seemingly “irrational” participants (Lancsar & Louviere, 2006; Szinay et al., 2021). Regardless, these tests are means to gain a better understanding of the data quality. It must be noted that the choice tasks that are explicitly designed for testing internal validity need to be excluded from choice modeling (Szinay et al., 2021). I incorporated these tests in addition to the 16 choice tasks required for this study, and the total choice tasks presented to the participants (including internal validity tests) increased to 18.

To ensure the highest data quality, in addition to the tests mentioned above, I embedded an expansive array of fraud and inconsistency checks throughout the body of the DCE survey. From using features such as bot detection through reCAPTCHA technology (Google, n.d.), to verifying and cross-referencing a variety of demographics information and using state-of-the-art technologies for language comprehension, I meticulously examined the truthfulness of the participants. I preferred false positives in the automated fraud detection procedures, and with the smallest of violations flagged the entry for further examination. Subsequently, I manually reviewed all flagged responses to evaluate the severity of violations and establish

whether they are justifiable. These checks are detailed further under Fraud Detection and Data Quality Assessment, and their findings are presented in Results.

### ***Study Phases & Variables***

Overall, this study consisted of three phases, each with their own unique objectives and outcomes of interest. The scope of each phase is generally compatible with the phases of clinical trials, as defined by the National Institutes of Health (NIH, 2022). Due to the novel hypothesis of this study and the complexities of the methodology, it was imperative that every aspect of the experiment design was tested and iterated upon between phases. Below, I describe these phases and provide overarching details for each. In the continuation of this dissertation, I will specify the context for various information accordingly.

**Feasibility Study.** This study focused on evaluating task comprehension and cognitive burden, measuring the time required to participate in the experiment, improving the presentation of information, and eliciting feedback for study improvement. It also allowed for the mechanical evaluation and improvement of the study pipeline, from designing the choice matrix to ultimately analyzing the data. This study did not attempt to create a preference model based on the data, and instead focused on ensuring the study can be carried out soundly. The main outcomes of this phase were validating the study design procedures and improving the task presentation.

Independent Variables:

- DCE attribute levels
- Sociodemographic information

- Rating perceived task complexity, time requirements, information presentation, and appropriateness of attributes and levels

Dependent Variables:

- Experiment design and analysis procedures evaluation
- Time requirement estimates
- Necessary design or presentation improvements

**Phase I.** This phase collected data from a small sample consisting of participants from the experiment target demographic, using a fully-featured and functional  $D_2$ -Efficient DCE design. Similar to the objectives of the Feasibility Study, I focused on ensuring that the task was comprehensible, imposed minimal burden, presented the information effectively, and elicited feedback for improvement. In addition, this phase helped determine the appropriateness of attribute levels and validated the study design procedures. The main outcomes of this phase were criteria for task or procedure improvement, internal validity report, and preferences information which was also used in identifying the priors for a  $D_b$ -Efficient design.

Independent Variables:

- DCE attribute levels
- Sociodemographic information
- Rating perceived task complexity, time requirements, information presentation, appropriateness of attributes and levels, familiarity with disability and accessibility, and importance of Usability/Accessibility Score
- Ranking subjective attribute priorities in decision-making
- Current use of medical devices

- Missing attributes identification

Dependent Variables:

- MNL model parameter estimates
- Experiment design and analysis procedures evaluation
- Internal validity
- Time requirement estimates and participant burden
- General sociodemographic characteristics of the sample
- Reports on sample medical device use, level of familiarity with accessibility and disability, importance of Usability/Accessibility Score, attribute priorities, and task comprehensibility
- Necessary design or presentation improvements

**Phase II.** This phase addressed the shortcomings identified in previous phases and collected preference data from a substantially larger sample of participants, using a Bayesian D-Efficient ( $D_b$ -Efficient) design based on the priors identified in Phase I. This phase of the study was preregistered with the Open Science Framework prior to data collection, as an indication of commitment to research transparency (Ardehali, 2023).

Independent Variables:

- DCE attribute levels
- Sociodemographic information
- Rating perceived task complexity, time requirements, information presentation, appropriateness of attributes and levels, familiarity with disability and accessibility, and importance of Usability/Accessibility Score

- Ranking subjective attribute priorities in decision-making
- Current use of medical devices
- Commitment and attention questions

Dependent Variables:

- MNL model parameter estimates
- Internal validity
- General sociodemographic characteristics of the sample
- Reports on sample medical device use, level of familiarity with accessibility and disability, importance of Usability/Accessibility Score, attribute priorities, and task comprehensibility

### **Data Collection**

Data collection took place in June, July, and November 2023. Participants answered an online survey that elicited their preferences for hypothetical Blood Pressure Monitors with various attribute levels. The recruitment and initial screening were facilitated by a reputable online research panel provider. Compliance with ethical research guidelines was approved and monitored by the University of Wisconsin-Milwaukee (UWM) Institutional Review Board (Study ID: 24.010, approved in July and November 2023). Below, various aspects of data collection for this study are discussed in detail.

### ***Sample***

Due to the reliance of SP methods on the evaluation of hypothetical alternatives, it is paramount to ensure the participants find the experiment realistic. In this study, part of the realism of choice scenarios was achieved by selecting a prevalent and important medical device

(BPM) to serve as a proxy for medical devices and facilitate examination of preferences for the study attribute levels. Subsequently, appropriate attributes and levels were determined based on extensive market research. Another facet of improving plausibility involved designing pertinent and appropriate choice tasks. The next step in maximizing the believability of choice scenarios and enhancing the validity and reliability of the study conclusions was the adequate recruitment of participants, identified appropriately in accordance with the experiment's hypothetical scenario.

**Sample Size.** Unlike power analysis for classical quantitative research methods, determining a sample size for DCEs is not a very straightforward task. There is no consensus among scholars on a statistical procedure to determine an appropriate sample size for DCEs (Bridges et al., 2011; Marshall et al., 2010). On page 670 of their DCE design guidelines for healthcare applications, Lancsar & Louviere state “our empirical experience is that one rarely requires more than 20 respondents per version to estimate reliable models, but undertaking significant post hoc analysis to identify and estimate co-variate effects invariably requires larger sample sizes” (Lancsar & Louviere, 2008). The late Dr. Richard Johnson (the founder of Sawtooth Software and author of Conjoint-Based Choice System) & Bryan Orme (the president of Sawtooth Software), proposed a rule-of-thumb to determine minimum sample size required for full-profile DCEs (R. Johnson & Orme, 2003; Orme, 2019). This rule-of-thumb, which has been used frequently in the literature (E. W. de Bekker-Grob et al., 2015), suggests that the minimum sample size  $N$  required for a DCE with the maximum number of levels  $c$  (if only main effects are going to be considered), number of choice tasks  $t$ , and number of alternatives per choice task  $a$ , can be calculated based on the following equation (Orme, 2019):

$$N > \frac{500c}{ta} \quad (9)$$

In this study, the maximum number of levels is for Usability/Accessibility Information with four levels. Also, the choice matrix includes 16 choice tasks, each with four alternatives. According to Johnson & Orme's rule-of-thumb, and based on (9), the minimum sample size required for this study is calculated as:

$$N > \frac{500 \times 4}{16 \times 4} = 31.25$$

This means that based on the Johnson & Orme's rule-of-thumb, this study will need at least 32 participants to estimate the main effects with statistical significance.

Another method of finding a proper sample size is the simulation approach. In this approach when a choice matrix is identified, by using computational methods and statistics software, the choice data can be simulated either randomly, or according to a prior model (Helveston, 2023a, 2023b). Subsequently, a simulated power analysis can be carried out by increasing the sample size incrementally, estimating the model for each simulated sample, and calculating the parameter standard errors (Helveston, 2023c). This approach allows for sample size estimation based on the study's requirements for standard error.

In homogeneous designs such as DCEs, Fisher information has a linear relationship with sample size  $N$  (Bliemer & Rose, 2014). Based on (7), this means the variance-covariance matrix of parameter estimates is inversely proportional to  $N$ , and as the sample size increases, standard errors decrease at a rate of  $\sqrt{N}$  (Bliemer & Rose, 2014). Johnson et al. demonstrated an example of using a simulation technique to examine the relationship between DCE parameter estimate precision and simulated sample size (F. R. Johnson et al., 2013). However,

instead of simulating choices for a design, Johnson et al. used the data from existing studies with very large sample sizes and simulated smaller sample sizes by randomly selecting participants for each target size (with replacement). Based on their findings, the precision of parameter estimates increases sharply as the sample size approaches 150, and beyond 300 increasing the sample size returns only minimal improvements to precision (F. R. Johnson et al., 2013).

Based on a systematic review of the literature by de Bekker-Grob et al., out of 69 publications in 2012 on the application of DCE in healthcare roughly 73% used a sample size of up to 300, with the majority between 100 to 300 (E. W. de Bekker-Grob et al., 2015). In another systematic review of 301 publications in health economics between 2013 and 2017, Soekhai et al. found the median sample size was about 400. A similar literature review by Trapero-Bertran et al. examined 72 research publications between 2008 and 2015, related to the application of DCE in healthcare, and reported the average sample size as 299 (Trapero-Bertran et al., 2019).

Due to the variety of recommendations from different sources, and the diversity of sample sizes used in the literature on the applications of DCE in healthcare, arriving at a definitive sample size proved challenging. I determined the sample size for this research by attempting to satisfy as many criteria as possible. For the Feasibility study, I estimated that about 10 participants were enough to examine the identified dependent variables, especially considering that this phase was not focused on constructing a preference model based on the DCE data. I conducted Phase I with 20 participants, which satisfied Lancsar & Louviere's recommendation for an adequate sample size. Next, I simulated the standard error for parameter estimates based on the model identified in Phase I and noticed that the precision

curve starts flattening for sample sizes between 150 and 200. Thus, I conducted Phase II by using a sample size of 180. The simulation results are presented in the Results chapter.

**Inclusion Criteria.** As discussed earlier in this chapter (p. 46), considering the prevalence of hypertension in adults aged 50 and over, and the importance of blood pressure monitoring in its management, I defined the inclusion criteria as American PwOD aged 50 or older and fluent in English. These criteria ensured the choice scenarios are realistic.

**Recruitment for Study Phases.** I recruited for the Feasibility Study by convenience sampling. I invited colleagues at the Rehabilitation Research Design & Disability (R<sub>2</sub>D<sub>2</sub>) Center at UWM, friends, and other personal and professional connections to participate in the online survey and share their feedback and experiences. For Phase I & II, participants were recruited through Prolific (prolific.com, London, UK).

Prolific is highly reputable and facilitates research for many renowned organizations such as The World Bank, Google, Meta, Yale University, and University of Oxford. This platform has a strict verification procedure for participants that includes photo identification and name and age verification using driver's licenses or national passports (Prolific, 2023c). In addition to providing more than 300 filters to set the inclusion criteria, Prolific also allows for including participants based on their approval rate from previous studies (Prolific, 2023a). For Phase I, aside from the inclusion criteria mentioned above, I only allowed participants with more than 90% approval rate to participate. After the conclusion of Phase I, the participants were excluded from the pool, and I recruited for Phase II with the same 90% approval rate criterion in addition to the study inclusion criteria. Eligible participants who fully participated in the data collection and provided quality data were compensated at a rate of \$12/hour.

I requested a sex-balanced recruitment for both phases, and Prolific attempted to recruit an equal number of male and female participants. The option for a population representative recruitment was also available but cost prohibitive. Instead, I elected to study aspects of the sociodemographic characteristics of the available participant pool by configuring various sociodemographic filters individually and recording the number of eligible and recently active participants reported by Prolific. While this would not guarantee data collection from a representative sample, at the minimum it would help demonstrate the potential for reaching it. Figure 8 shows the values reported by Prolific compared to the data gathered from the American Community Survey 2022 (U.S. Census Bureau, 2022) and Gallup Poll 2022 (Jones, 2023) where the census data were unavailable. This figure shows how the eligible participants on Prolific with more than 90% approval rates and active within the 90 days prior to July 6, 2023 ( $N = 2786$ ), compared to the US population of PwOD above the age of 50 in 2022 ( $N = 88,401,084$ ). The only exception is the comparison between sexual orientations where the American Community Survey (2022) did not have any reports available, and I used the Gallup Poll (2022) for comparison. Due to the discrepancy in labeling I attempted to match the labels to the extent possible to allow for a relatively accurate comparison. The exclusion of any category from the individual comparisons only reflects data unavailability either on Prolific or from the sources used for cross-examination. The results showed that while discrepancies existed, they were mostly minimal, and the experiment would be made available to a relatively representative pool of participants.

**Figure 8:**

*Comparison Between Total Eligible Respondents' sociodemographic Information from Prolific and the US Population Characteristics*



**Instrumentation**

Each phase of this study was developed in two stages: DCE Design, and Survey Design. While the attributes and levels remained the same, each phase required its own design parameters. Three software were predominantly used in the design stage: R (R Core Team,

2023) in RStudio (Posit team, 2023), Python (Python Software Foundation, 2023) in Visual Studio Code (Microsoft Corporation, 2023b), and Excel (Microsoft Corporation, 2023a).

**DCE Design.** I used the `cbcTools` (Helveston, 2023a) and `logitr` (Helveston, 2023b) packages in R to design the  $D_z$ -Efficient and  $D_b$ -Efficient choice matrices, examine their D-error and balance, and simulate choices for power analysis to estimate sample sizes. After identifying the appropriate design, I exported the data as a spreadsheet and used the `pandas` package (The pandas development team, 2023) in Python to transform choice matrices into tables of alternatives. The tables created in this stage were compatible with direct placement in the survey and maintained appropriate aesthetics. The syntax used in this stage is included in Appendix A and B.

The objective of the Feasibility Study was mainly the mechanical evaluation of survey flow and data processing pipeline, handling ineligibility for participation, and estimating the time and cognitive burden on the participants. The participants I recruited for this phase were mainly from backgrounds adjacent or relevant to disability sciences and were extensively familiar with the need for accessible design. Thus, the choice information obtained in this phase was not meaningful due to bias. Nonetheless, to examine the experiment design and data analysis procedures comprehensively, I designed a choice matrix assuming zero priors for this phase ( $D_z$ -Efficient). The full design is presented in Appendix C.

I conducted Phase I with participants who met the inclusion criteria, and their stated preferences were valuable for a small-scale evaluation of the hypotheses. Since no appropriate priors were available yet, I generated a new  $D_z$ -Efficient design. In contrast to the Feasibility Study, for this phase I examined 24 random seeds to identify a design with the lowest D-error

that did not substantially compromise attribute level balance. After identifying the desired design, I performed simulated power analysis based on the parameter standard errors curve relative to sample size. Since I was assuming zero priors, the simulated choices mimicked derivation from an MNL model with zero-adjacent parameters. In other words, the choices were simulated such that the estimated MNL model would have near-zero part-worth utility for all parameters. Consequently, I was not expecting the simulation results to accurately predict my observations after data collection. Next, I generated the corresponding tables for alternatives and finalized the survey by incorporating the DCE questions. The full design is available in Appendix D.

Pursuant to the conclusion of data analysis in Phase I and estimating the MNL model parameters, I obtained the appropriate priors to create a  $D_b$ -Efficient design for Phase II. I updated the syntax to use these priors and proceeded to generate 229 unique designs in a loop. I documented the randomization seeds as well as  $D$ -error and balance information and used these data points to rank the generated designs. Through this process, I identified the design with the lowest  $D$ -error that did not overtly disregard attribute level balance. After identifying the desired  $D_b$ -Efficient design, I simulated choices for power analysis and estimated an appropriate sample size. Finally, I transformed the choice matrix into tables and incorporated them into the Phase II survey version. This  $D_b$ -Efficient design is presented in Appendix E.

**Survey Design.** I used Qualtrics (Qualtrics, 2023) to create online surveys that contained the DCE design specific to each study phase, as well as many other segments to capture various information from the participants. A link to this study was embedded into Prolific for

recruitment. After conducting each phase of this study, I revised the survey based on the findings to ensure improved data quality.

To help improve the participants' experience, and consequently the quality of the data, I chose to present survey pages individually and use a swiping animation when moving from one question to the next. This was especially important for the DCE questions as without the animation, the page and information appeared to remain static while only some levels changed in the choice tasks. This animation helped reinforce and confirm the sense of accomplishment from answering each question, while providing prominent visual feedback about the change in the page contents. I also did not include a "back" button to prevent participants from going back and changing their responses based on the information they may have obtained later in the survey. However, I made all questions mandatory such that the participants could not proceed without answering all the questions on each page. The only exceptions were the demographics questions where the participants were allowed not to respond, for research ethics purposes. This ensured that all participants who finished the survey and were eligible for payment had provided a response to almost all questions. At the bottom of each page, I included a progress bar to help inform participants about their advancement in the survey.

The segments of the survey are listed and explained below, and their inclusion in different phases is identified:

1. Consent Form (Feasibility, Phase I, Phase II): The consent form approved by the UWM Institutional Review Board was presented to the participants to inform them of the study procedures, data collection, safety, and privacy. Participants provided their consent by selecting the corresponding option. If participants

decided not to provide consent, they were redirected to Prolific with an “incomplete” code to indicate they did not finish the study and a replacement participant is required. An unapproved version of this consent form was also made available to the Feasibility Study participants since approval for this stage was unnecessary.

2. Commitment Check (Phase II): After noticing the challenges in ensuring data quality in Phase I, I decided to implement various tools to improve the probability of collecting high quality data. A study conducted by Qualtrics showed that commitment checks are very effective. This study found that by asking the participants to commit to providing truthful and thoughtful responses, the rate of data quality problems decreased substantially (Geisen, 2022). As a side, I believe this type of question may be exploiting an interesting manifestation of Dr. Festinger’s Cognitive Dissonance theory (Festinger, 1957), or perhaps Dr. Bem’s Self-Perception theory (Bem, 1967). In this section I briefly explained the purpose of the study and its potential for positive impact on society, then asked the participants whether they will commit to reading and following instructions and providing honest and thoughtful responses.
3. Screener Validation (Feasibility, Phase I, Phase II): Prolific makes the survey available only to eligible participants, however, sometimes participants who may not be eligible for the study may find their way into the survey. To comply with the Prolific guidelines (Prolific, 2023b), and ensure recruitment of eligible participants, I asked participants to verify their age and disability status. If

participants failed the screener validation questions, they were redirected to Prolific with an “incomplete” code.

4. Task Comprehensibility Potential (Feasibility, Phase I, Phase II): Two questions followed the screening questions. The first question inquired about BPM use by participants or by individuals they know. The second question asked whether participants think they will be able to imagine needing to purchase a BPM for themselves or someone they know. The purpose of these questions was to assess the likelihood of collecting realistic choice data from participants, however, they were not used to exclude individuals from the study.
5. Attention Check 1 (Phase II): To assess and improve data quality, two attention checks were added to the survey in Phase II. These questions were generic, and with obvious logical answers, to evaluate whether participants are paying attention to the questions. These attention checks were implemented in accordance with Prolific guidelines for payment rejection (Prolific, 2023b), and their inclusion is supported by literature (Abbey & Meloy, 2017; Gummer et al., 2021). The first attention check appeared at this point in the survey and asked participants to rate their agreement with the statement “I live with one or more unicorns” on a four-point Likert-type scale (strongly disagree, disagree, agree, strongly agree). Only “strongly disagree” and “disagree” options are considered acceptable.
6. DCE Instructions and Mindset (Feasibility, Phase I, Phase II): This section of the survey provided information about the DCE questions, different attributes and

the range of their levels, and how the questions must be answered. This section asked participants to confirm their understanding of the instructions and followed up with a question on one of the attributes for evaluation.

7. Practice Question/Dominant Alternative Check (Feasibility, Phase I, Phase II):

Following the DCE instructions, participants were presented with a Within-Set Dominant Alternative test disguised as a practice question. The instructions prompted participants to attend to the difference between attribute levels in the hypothetical BPMs presented in the question and select the one they are most likely to purchase. I disguised the Dominant Alternative question as a practice question immediately after the instructions to directly evaluate how much the participants comprehended the task. Additionally, in selecting the dominant alternative, I had a strong sense of which attribute levels are most desirable, except for brand. For example, it can be reasonably assumed that lower Cost and higher User Rating and Usability/Accessibility Score are likely more desirable. However, I was unsure whether participants would prefer one brand over the other, despite Omron being introduced as a trusted brand, and the imaginary brand “HomeHealth” as a newly established one. Therefore, I included both alternatives in this test and accepted their selection as the correct identification of the Dominant Alternative.

8. DCE Block (Feasibility, Phase I, Phase II): The 16 choice tasks were individually presented next. Each survey version presented the choice matrix tailored

specifically for the corresponding phase. The order of the choice tasks and the alternatives in each task were randomized to prevent bias.

9. Stability Validity Test (Feasibility, Phase I, Phase II): This question repeated a question from the DCE block to assess preference consistency and help establish internal validity.
10. Follow-Up Questions (Feasibility, Phase I, Phase II): On all survey versions, I asked four follow-up questions in this section. The first question presented all four attributes and asked participants to rank them from most to least important in their purchasing decisions. The purpose of this question was to observe how participants consciously and explicitly rank these attributes overall, regardless of what they implicitly prioritized in their DCE choices. The second question asked participants to identify whether they have a medical device at home and if so, to name the device. Next, participants were prompted to rate their familiarity with disability and accessibility concepts, on a five-point Likert-type scale, from “not familiar at all” to “extremely familiar.” The last question asked participants to explicitly rate the importance of Accessibility/Usability Score in their purchasing decisions on a five-point Likert-type scale, from “not important at all” to “extremely important.” This question provided an opportunity to compare descriptive statistics of the responses to the findings of the preference models.
11. Attention Check 2 (Phase II): The second attention check was presented at this point in the Phase II survey, and asked participants to rate their agreement with the statement “all living humans need to breathe”, on a four-point Likert-type

scale (strongly disagree, disagree, agree, strongly agree). Only “strongly agree” and “agree” options are considered acceptable. I intentionally selected a statement that required a positive response, in contrast to the previous attention check.

12. General Sociodemographic Questions (Feasibility, Phase I, Phase II): Five questions in this section collected data on the gender, race, age group, state of residence, and annual household income. These questions were not mandatory. Participants could select multiple races or select “Other” and identify their race. The questions on age and income provided ranges instead of specific values. These questions were presented at this point in the survey flow to prevent dropouts before the DCE if sharing personal information is upsetting for some participants.
13. Effort Check (Phase II): I realized after Phase I that most questions on the survey require participants to select an option, and there were no mandatory long-answer questions. Some long-answer questions were present but not mandatory, and asked questions that could simply be responded with one or two words at best, and with “NA” or no response at worst. This shortcoming denied me access to a potentially important measure of honest and attentive participation. I decided to include a mandatory long-answer question in the Phase II survey that asked participants to describe their perception of the study purpose in two or more sentences. This question accomplished two objectives at once. Firstly, it created an opportunity for me to evaluate the participants’

perceptions, examine how they approached the questions, and determine how accurately they approximated the true purpose of the study (which was never explicitly disclosed). Secondly, it was an opportunity to observe the participants' level of adherence to instructions and their willingness to invest effort into their responses. Incoherent, short, or irrelevant responses to this question are also considered grounds for payment rejection on Prolific, as the required effort level was clearly identified (Prolific, 2023b).

14. Feedback Questionnaire (Feasibility, Phase I, Phase II): The purpose of this questionnaire was to evaluate the experiences of participants. In different phases of the study, this questionnaire included different questions based on the phase objectives. The common questions between all versions mainly asked participants to rate their experiences with the length of the survey, clarity of instructions, comprehensibility of the study purpose, mindset assumption difficulty, and choice task difficulty. An open-ended question in this section asked participants to identify attributes absent from this study that they deem important while purchasing products online. This question was not made mandatory to avoid pressuring participants into fabricating untruthful responses. The purpose of this open-ended question was to help establish the validity of this study's attributes. The last question on all versions asked participants to share their thoughts or feedback.

Following the successful completion of the survey, participants were redirected to Prolific with a completion code via a link and awaited manual approval of their submissions. In

the survey flow described above, it may be evident that different versions were improved and iterated upon based on the findings of previous phases. I compiled survey versions in Appendix F and G.

### **Data Analysis**

Phase I and Phase II data first underwent a rigorous quality check to identify and exclude entries from fraudulent participants. Although Prolific appears to diligently vet their online panels, I found that the probability of encountering a few fraudulent respondents is not zero. After examining data quality with a reasonably high standard, the data from each phase went through a transformation step to attain a structure recognizable by the MNL analysis module. In this step the internal validity tests were also checked, with the option to produce a secondary dataset and exclude participants who failed validity tests. As detailed previously, failing the internal validity tests does not offer sufficient motivation to exclude entries from data analysis entirely. However, I decided to estimate models with and without the questionable entries and compare the findings. Finally, after transformation into the appropriate structure, the data were made available to the MNL analysis module to estimate the utility model, its performance metrics, and the parameter estimate variance-covariance matrix described by (7). I programmed the fraud detection and analysis preparation modules in Python, and the MNL analysis module in R.

### ***Fraud Detection and Data Quality Assessment***

The validity of SP methods such as DCE relies heavily on participants and data quality. Due to the crucial importance of this factor in making reliable and valid conclusions, I prioritized developing a robust hybrid approach to evaluate data quality and detect participants who are,

at best, dishonest or inattentive. This hybrid approach hinged on four key resources: Qualtrics platform's built-in fraud detection, Qualtrics survey metadata, Prolific demographics report for the participants, and participants' own responses to unique or redundant questions.

Under the survey settings, Qualtrics provides four fraud detection features (Qualtrics, n.d.-a), two of which appeared relevant to my application: multiple submissions prevention, and bot detection using reCAPTCHA technology. The first option did not behave as advertised and I still found duplicate submissions. The second option, however, provided a score between 0 and 1 to indicate the likelihood that the participant is a human and not a bot. Unfortunately, these features are not enabled by default, and I was only able to enable them for Phase II after the conclusion of Phase I. This information was used as one of the vectors for fraud detection.

Qualtrics also records survey metadata such as public IP address and geographical latitude and longitude. IP-based geolocation can be accurate only to an extent. At the country level the accuracy is usually about 99%, while at the state level the accuracy drops to about 80% to 90% (ipify.org, 2019; IPinfo Team, 2020; IPLocation, 2018). Qualtrics clarifies that their latitude and longitude reports are based on IP information, and are accurate mostly to the city level inside the United States (Qualtrics, n.d.-b). Qualtrics also facilitates recording each participant's unique Prolific ID once they are directed to the survey from Prolific. In addition, if timing questions are embedded in the survey, Qualtrics records page submission times and click counts (timers were not shown to the participants). However, the timing option is not enabled by default, and I embedded timing into the survey only for Phase II. These metadata provided additional fraud detection vectors.

Prolific has very strict policies on VPN use, and bans participants who use these services (Croissant, 2021). After participants are deemed eligible and the survey is made available to them, Prolific shares their unique Prolific ID with the survey system to help distinguish participants across platforms. It also shares the responses participants provided in their profiles relating to the inclusion criteria identified by the researcher for recruitment. These data were made available as the demographics report after the conclusion of each phase and served as additional fraud detection vectors to cross-validate responses.

Finally, participants' responses to various survey questions created other opportunities for fraud detection. I strategically designed some of the questions to play a second role as data quality checkpoints, in addition to their primary role of gathering data pertinent to the research questions. For example, Prolific demographics report includes the actual age of the participants. One of the screener validation questions on the survey asks participants to identify their age based on wide ranges (e.g., 18 to 34, 35 to 49, or 50 and above). After responding to more than 30 questions, participants are asked to select their age range again under the General Sociodemographic Questions, this time in finer detail (e.g., below 49, 50 to 54, 55 to 59, 60 to 64, and so on). For an inattentive or fraudulent participant, the data from two independent sources verified three times in different forms may be inconsistent. By embedding more such checkpoints into the survey I further facilitated the detection of dishonest behavior.

I examined six potential fraud detection vectors in Phase I and added another 14 in Phase II. Some checks were as simple as identifying whether a participant scored lower than 0.5 on reCAPTCHA, or comparing the time spent on various questions to find the outliers. Others were more elaborate and, for example, cross-validated responses using independent sources

(such as the age example above). Regardless of their individual elaborateness, these vectors synergized in conjunction with one another to identify suspicious entries in need of more granular inspection. These vectors are individually identified and briefly described below to provide a better overview of the depth and breadth of data quality examination procedure in this study:

1. Multiple Entry Check (Phase I, Phase II): Identified participants who had more than one entry associated with their Prolific ID. Due to the severity of this violation, a report was created from the IDs of these participants, and all their entries were removed from the data. The rest of the functions below analyzed only the unique participants.
2. Age Cross-Validation (Phase I, Phase II): For the participants who identified as being above the age of 50 on the screener validation questions, the age range they identified under sociodemographic questions was compared to their true age as reported by Prolific demographics report. The sociodemographic questions were not mandatory, and I treated missing values as discrepancy due to preference for false positives.
3. Disability Cross-Validation (Phase I, Phase II): Compared the response to the screener validation question on disability status, with the information from Prolific demographics report.
4. Ethnicity Cross-Validation (Phase I, Phase II): Compared the response to the survey's sociodemographic question on race, with the information from Prolific demographics report on ethnicity. It is evident that there is a terminology

difference between “race” and “ethnicity.” Also, there was a discrepancy in labeling different races and ethnicities. I attempted to match the definitions to the degree possible. However, I maintained my preference for false positives in this stage, even if the discrepancy was something entirely justifiable. For example, individuals who identified as White on Prolific may have selected the “other” option on the survey and typed in Hispanic. I ignored justifiable discrepancies in later stages.

5. IP-Based US Location (Phase I, Phase II): Used IP-based geolocation to confirm the participants were in the United States. IPinfo API was used for IP geolocation (IPinfo, 2023). The alternative to use latitude and longitude data from Qualtrics (also IP-based) was available, handled through the GeoPy library (GeoPy, 2023).
6. IP-Based State of Residence Validation (Phase I, Phase II): Validated the state of residence identified by each participant in sociodemographic questions, based on the IP geolocation or latitude-longitude data from Qualtrics. IPinfo API or GeoPy were used respectively.
7. Total Survey Duration Outliers (Phase II): Prolific identifies a specific criteria for payment rejection if the participant finished the survey “exceptionally fast” (Prolific, 2023b). Outliers in this case are defined as those who completed the survey in less time than three standard deviations below the mean. This criterion was used to identify participants who finished the survey unreasonably quickly.

8. Bot Detection (Phase II): Checked reCAPTCHA score provided by Qualtrics metadata and flagged the entry if this score was below 0.5, the threshold identified by Qualtrics (Qualtrics, n.d.-a).
9. Commitment Check (Phase II): Flagged participants who responded to the question asking for their commitment to attentive, honest, and thoughtful participation, with anything other than “Yes, I will.”
10. DCE Mindset Check (Phase II): Participants were asked if they can imagine needing to purchase a BPM for themselves or someone they know. If the response was anything other than “yes,” the participant was flagged.
11. Attention Checks (Phase II): If the responses to either or both attention checks were not acceptable, the participant was flagged.
12. DCE Instructions Comprehension Test (Phase II): After providing information on the attributes and levels of the survey, participants’ comprehension was examined by asking a question on the instructions still visible on the page. If a response was incorrect, the participant was flagged.
13. Gender Cross-Validation (Phase II): This was another point of terminology discrepancy between the survey and Prolific. Prolific reported “sex,” while we asked participants about their “gender.” Nevertheless, discrepancies were flagged to uphold preference for false positives.
14. Survey Page Times Outliers (Phase II): The time spent on each page by each participant was compared to the time spent by all other participants on the same page. Using the Interquartile Range (*IQR*), values below  $Q_1 - 1.5 \times IQR$  or

above  $Q_3 + 1.5 \times IQR$  were flagged ( $Q_1$  and  $Q_3$  are the first and third quartiles respectively).

15. Survey Page Click Counts Outliers (Phase II): Using the *IQR* method, participants who clicked on a page substantially more or less than others were flagged.
16. Time on DCE Instructions Outliers (Phase II): Using the *IQR* method, participants who spent substantially more or less time than others on the DCE instructions were flagged.
17. Time on Dominant Alternative Outliers (Phase II): Using the *IQR* method, participants were flagged if they spent substantially more or less time than others on the Dominant Alternative choice task.
18. Time on Choice Consistency Outliers (Phase II): Like 17, participants who were outliers based on the time they spent on the Choice Consistency test were flagged.
19. Approximate Time on DCE Outliers (Phase II): Qualtrics survey platform issued an incompatibility warning about using timing questions in blocks with custom functions for randomization. Due to the use of randomization in ordering DCE choice tasks, as well as the order of alternatives within each choice task, I was unable to time choice tasks individually. However, I estimated the time spent on the DCE by deducting the time spent on all other pages from the total duration of the survey for each participant. The outliers were flagged using the *IQR* approach.

20. Study Purpose Response (Phase II): This function used a state-of-the-art technology for language comprehension and compared perceptions of the study purpose as reported by participants to the “canonic” study purpose which was a sentence I wrote to summarize the purpose of the study. My summary of the study’s purpose was in fact provided in the brief description of the study on Prolific, which participants saw before taking the survey. However, I did not overtly disclose the study’s intent, which was to evaluate the utility of Usability/Accessibility Score. Instead, to avoid bias, I described what the participants were going to see on the survey by stating “The purpose of this study is to understand how people purchase and use blood pressure monitors, and what factors impact their decision-making in the purchasing process.” I used the GPT-3.5 Turbo model, an Artificial Intelligence model designed by OpenAI, via the OpenAI API for Python (Brockman et al., 2020). I instructed the model to determine whether the study purpose stated by each participant was relevant to my “canonic” purpose. If the model determined that the purpose was misidentified by a participant, they were flagged.

The functions above carried out the automatic portion of fraud detection and a new spreadsheet was created by the program, comprised of entries from unique participants who met the inclusion criteria. The output spreadsheet included two new columns in addition: one for listing the violation categories for all participants, and one to provide the details of their violations. I manually reviewed this spreadsheet to determine the severity of violations, evaluate other responses from the flagged participants, and determine what discrepancies

were justifiable. Even those who were not flagged by the code were still manually reviewed to ensure response consistency and appropriateness.

After manually examining the data, I removed and reported entries from participants who met any of Prolific's rejection criteria, as well as those with multiple submissions that had already been excluded from the output spreadsheet. These criteria included failing both attention tests, exceptionally low total submission times, and objectively demonstrable low effort when the effort requirements were explicitly identified (Prolific, 2023b). Next, I evaluated entries that were not rejected on Prolific to determine if they were questionable enough to exclude from data analysis. During the automated phase, I prioritized identifying false positives to extract as many discrepancies as possible. In contrast, the manual review focused on identifying redeeming factors and understandable violations. For example, a misclick might be reasonably inferred if a participant misidentified their age range, but their true age fell in the category directly above or below their selection, particularly if the rest of their entry showed no concerning signs. Similarly, a mismatch between a participant's state and IP geolocation data might be explained by travel. Time spent on questions was also considered; outliers who consistently took longer were not a cause for concern. In fact, I often viewed this as a redeeming factor against minor violations. Discrepancies in labeling, as described earlier in the Race and Sex Cross-Validation section, were also overlooked if they seemed reasonable.

Figure 9 shows an example of the outcome of this module after manual review. The row highlighted in red denotes that the participant had severe violations and was excluded and reported (due to failing both attention checks). Rows highlighted in orange indicate that the participants had two important violations, but I was indecisive about removing them entirely

from the analysis, based on my review of the rest of their entries. For example, a participant may have said they cannot imagine needing to purchase a BPM, and incorrectly responded to the DCE instructions comprehension question, but did not have any other severe violations, appeared truthful in their responses, and proceeded to identify the purpose of the study accurately. In such cases, my concerns regarding the validity of answers were somewhat alleviated. Nonetheless, I marked these entries as the first ones to be removed from the analysis, if necessary. Similarly, rows highlighted in yellow had only one important violation, or a few less significant ones, but the rest of their entries did not raise any major concerns. I was

**Figure 9:**

*Example of Manual Review of Data Quality After Automatic Fraud Detection*

ProlificID	ViolationCategories	Violations
	State - AttentionChecks	['Ohio', 'Arizona'] - ['Attn. 1 Strongly Agree', 'Attn. 2 Strongly Agree']
	MindsetAssumption - InstructionCo	No - Omron - ['Q3_Click Count 1.0 >>> 0.0', 'Q43_Click Count 8.0 >>> 2.0']
	MindsetAssumption - PageSubmiss	No - ['Q7_Page Submit 22.05 >>> 12.94'] - ['Q16_Click Count 3.0 >>> 1.0']
	MindsetAssumption - InstructionCo	No - Omron - ['Q7_Click Count 5.0 >>> 2.0', 'Q16_Click Count 2.0 >>> 1.0']
	MindsetAssumption - InstructionCo	No - Omron - ['Q3_Page Submit 331.42 >>> 2.95'] - ['Q7_Click Count 8.0 >>> 2.0']
	InstructionComprehension - PageC	Omron - ['Q18_Click Count 5.0 >>> 2.0'] - I think this study was to
	State - AttentionChecks - Instructio	['California', 'Texas'] - ['Attn. 1 Agree'] - Omron - ['Q1_Page Submit 8.0 >>> 2.0']
	MindsetAssumption - InstructionCo	No - Omron - ['Q3_Page Submit 7.48 >>> 2.95', 'Q41_Page Submit 299.4 >>> 2.0']
	InstructionComprehension - PageS	Omron - ['Q18_Page Submit 89.16 >>> 49.51'] - ['Q39_Click Count 2.0 >>> 1.0']
	InstructionComprehension - PageC	Omron - ['Q12_Click Count 2.0 >>> 1.0', 'Q21_Click Count 2.0 >>> 1.0']
	InstructionComprehension - PageC	Omron - ['Q3_Click Count 1.0 >>> 0.0', 'Q18_Click Count 5.0 >>> 2.0']
	PageClickCounts	['Q1_Click Count 8.0 >>> 2.0', 'Q18_Click Count 20.0 >>> 2.0']
	InstructionComprehension - PageC	Omron - ['Q12_Click Count 2.0 >>> 1.0'] - It may have to do with
	InstructionComprehension - PageS	Omron - ['Q14_Page Submit 60.1 >>> 9.2', 'Q41_Page Submit 58.6 >>> 2.0']
	InstructionComprehension - PageC	Omron - ['Q16_Click Count 2.0 >>> 1.0']
	InstructionComprehension - PageS	Omron - ['Q49_Page Submit 48.3 >>> 13.5'] - ['Q21_Click Count 2.0 >>> 1.0']
	MindsetAssumption - PageSubmiss	No - ['Q41_Page Submit 47.35 >>> 27.04', 'Q55_Page Submit 299.4 >>> 2.0']
	State - InstructionComprehension -	['South Carolina', 'Texas'] - Omron - ['Q1_Page Submit 105.87 >>> 2.0']
	State - MindsetAssumption - PageS	['North Carolina', 'New York'] - No - ['Q3_Page Submit 8.33 >>> 2.0']
	PageSubmissionTime - PageClickCo	['Q1_Page Submit 67.6 >>> 28.15', 'Q18_Page Submit 475.07 >>> 2.0']
	PageSubmissionTime - PageClickCo	['Q1_Page Submit 273.14 >>> 28.15'] - ['Q3_Click Count 1.0 >>> 0.0']
	State - MindsetAssumption - PageC	['Washington', 'West Virginia'] - No - ['Q3_Click Count 1.0 >>> 0.0']
	PageClickCounts - StudyPurpose	['Q41_Click Count 1.0 >>> 0.0'] - Its purpose was to judge how pe
	PageSubmissionTime - PageClickCo	['Q67_Page Submit 17.46 >>> 4.01'] - ['Q16_Click Count 2.0 >>> 1.0']
	PageSubmissionTime - PageClickCo	['Q39_Page Submit 162.92 >>> 12.93'] - ['Q39_Click Count 2.0 >>> 1.0']
	PageSubmissionTime - TimeInstruc	['Q12_Page Submit 38.48 >>> 7.71', 'Q14_Page Submit 15.34 >>> 9.2']
	StudyPurpose	It was probably to see how desirable a new name is. It was to ex
	PageClickCounts	['Q39_Click Count 2.0 >>> 1.0']
	PageSubmissionTime - PageClickCo	['Q3_Page Submit 33.23 >>> 2.95', 'Q7_Page Submit 21.93 >>> 12.94']
	PageSubmissionTime - PageClickCo	['Q41_Page Submit 47.35 >>> 27.04', 'Q55_Page Submit 299.4 >>> 2.0']

Note: While Prolific IDs are not personally identifiable information, they were blurred out of respect for the participants. For brevity, only part of the spreadsheet is shown.

not confident about fully removing these entries. Finally, rows highlighted in green either had negligible or justifiable flags, or no flags at all. To avoid introducing bias into the results by removing the “undesirable” entries (Lancsar & Louviere, 2006), I refrained from excluding any of the orange or yellow rows. The script I wrote for this module is available in Appendix H.

### ***Analysis Preparation and Internal Validity Evaluation***

The penultimate stage before estimating the model parameters involved (a) assessing the internal validity of DCE choices using the tests dedicated to this purpose; and (b) restructuring the data to a format appropriate for the MNL analysis module. To accomplish (a), the syntax read the participants’ responses to the Dominant Alternative question and compared them to the alternatives designated as dominant. The report consisted of the participant IDs for those who failed the test, the selection frequencies for each alternative designated as dominant, and the overall success rate. Additionally, the syntax identified the alternative selected by each participant in the Choice Consistency question and compared it with the alternative they had selected when they saw the same choice task earlier. The report consisted of the participant IDs for those who were not consistent in their choices, an overview of the discrepancy between their choices, and the overall success rate. As discussed earlier, there is no consensus on whether these participants should be excluded from the analysis. However, this module allowed for excluding the participants who failed internal consistency tests and creating a new spreadsheet for a secondary analysis, if necessary.

To accomplish (b) as described above, DCE choice data extracted from the survey needed to be transformed from the wide format into the long format. Wide format, or “unstacked” data, is a structure where each observed unit (each participant) is represented

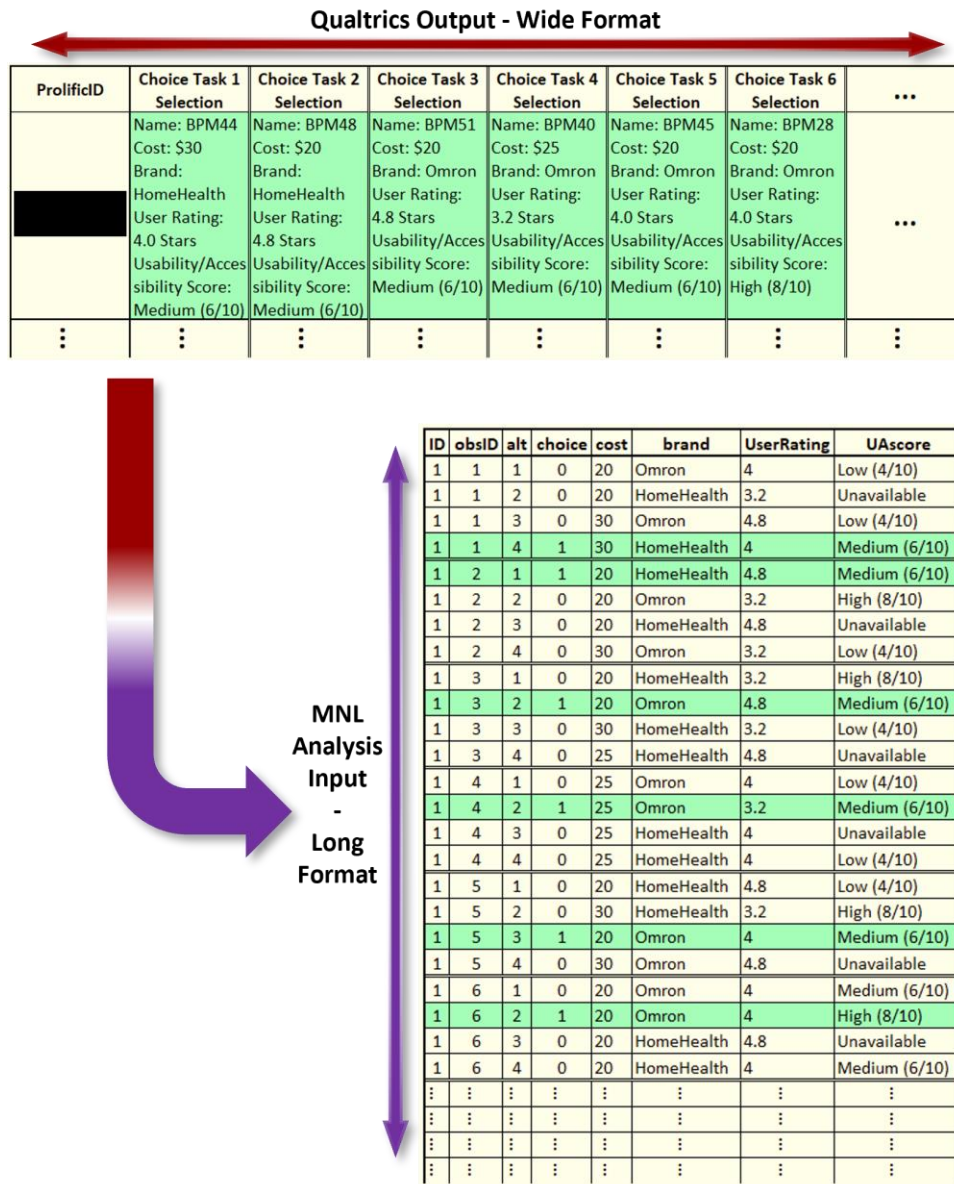
with a row, and the observed variables (selected alternatives) are spread across various columns. Conversely, in long format or “stacked” data, each row corresponds to a single observation. In this structure, more than one row may be dedicated to each observed unit, depending on the number of observations made from that unit.

The exported survey from Qualtrics presents participants’ responses in a wide format. Each participant has a dedicated row, and their selection in each choice task is presented as separate columns. The MNL analysis module requires this data to be transformed into the long format where each choice is represented on a single row. In addition, these choices need to be put in context, which means identifying the alternatives that were rejected from the same choice task. This context is crucial for identifying what tradeoffs were made. Figure 10 shows an example of data transformation from the wide to long format with the addition of context. It should be noted that only a portion of the data is shown, and only for one participant, due to size and legibility considerations.

I used the pandas and NumPy (C. R. Harris et al., 2020) packages in Python to automate the processes described in this module. An additional report on the survey’s total duration was also created at this stage. The script I used for this module is available in Appendix I.

Figure 10:

Example of Data Transformation from Wide to Long Format with the Addition of Choice Context



**Multinomial Logit Analysis**

Once the reports on the internal validity tests were generated and the DCE data were transformed into the appropriate format, I estimated the MNL model using the logitr package in R. Subsequently, I extracted the estimated parameters and statistics, as well as model

performance parameters. I also obtained the variance-covariance matrix of the parameter estimates at this stage. If the findings required additional a posteriori analyses, these procedures were conducted either after or concurrently with this module. The R syntax for this module is available in Appendix J.

### **Methods Summary Review**

The need to evaluate the hypotheses of this study, as well as its significance and relevance, were extensively discussed in prior chapters. This chapter meticulously detailed the process of establishing an experimental design that facilitates proper examination of the hypotheses. More specifically, DCEs were discussed as a prominent approach in the examination of stated preferences. I documented the prevalence of DCEs in healthcare choice experiments and their solid theoretical foundation and elected to use this methodology to evaluate my hypotheses.

To make the research questions more tangible, I identified blood pressure monitors as a prevalent medical device, to use as a proxy for medical devices. The prevalence and importance of using BPMs also aided in further grounding the research questions, and to make the conclusions relatively more generalizable. Next, I identified the demographics of individuals for whom purchasing a BPM may seem more realistic and appropriate. Through conducting comprehensive market research, I identified the most common attributes and levels of blood pressure monitors and made an informed decision on which attributes and levels can be evaluated more objectively. Due to the novelty of this study's central hypothesis, I deemed the inclusion of attributes with more likelihood for preference subjectivity inappropriate. Additionally, research guidelines and best practices caution against the inclusion of too many

attributes, and especially ones that may not be objectively evaluated. Based on the market research, and general guidelines on designing choice experiments, I determined that Cost, Brand, and User Rating may be the most prominent and reasonably quantifiable attributes of BPMs. Additionally, I introduced Usability/Accessibility Score as the fourth attribute, to serve as a representation of Usability/Accessibility Information. Realistic and evidence-based levels for each of these attributes were identified and summarized in Table 3.

Following the establishment of attributes and levels, I decided to use full-profile choice tasks. This meant that all attributes were presented for each alternative. Since only four attributes were involved in this study, full-profile choice tasks were not expected to impose an undue cognitive burden. I also decided against including an opt-out option. While opt-out options add realism to a choice scenario, they also detract from collecting enough information about the tradeoffs between attribute levels. In addition, there is no consensus on modelling approaches for choice experiments with opt-out options.

Next, I thoroughly examined and discussed the options for experimental design in DCEs. Generally, MNL is a popular analytical model in economics and econometrics, and specifically, it is very prevalent in health econometrics literature. I discussed the merits of this analytical model, as well as its assumptions and potential limitations. With all the considerations and implications, this analytical model still seemed appropriate for this study. To the best of my knowledge, no studies to date have examined hypotheses adjacent to this study's, within the scope of its target population. As such, no empirical evidence is available to support the existence of interaction effects between the parameters of this study. Systematic reviews of the literature on DCE in healthcare have also reported that most studies examine only the main

effects. Thus, I chose to examine only the main effects via MNL analysis. Since this study includes only BPMs with different permutations of attribute levels, an unlabeled design seemed appropriate. After meticulous examination of various DCE guidelines and best practices, I determined that a 16x4 choice matrix design (16 choice tasks with four alternatives each) without blocking will be able to capture the required choice data for hypothesis evaluation.

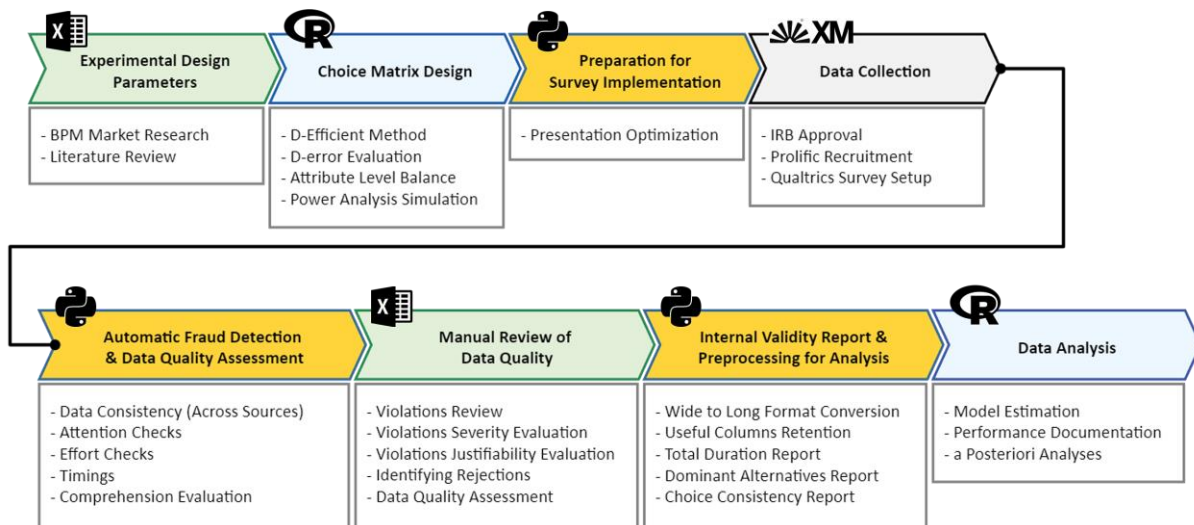
Different approaches to choice task design were examined and D-Efficient designs appeared to hold promise for this study's requirements. No relevant priors could be established for the study parameters, due to the unavailability of relevant literature. As such, I decided to use a  $D_2$ -Efficient design for Phase I of the study and use the established informative priors to design a  $D_b$ -Efficient choice matrix for Phase II. The use of D-Efficient design approaches also allowed for exclusion of the Dominant Alternate. While D-Efficient designs do not prioritize attribute level balance primarily, research suggests that it is an important consideration. I decided to generate many different D-Efficient designs for each phase and rank them based on D-error and attribute level balance characteristics, to identify designs with comparatively minimal D-error that avoid blatantly sacrificing balance.

Approaches to internal validity assessment were discussed next, and the reasons behind this paramount consideration were identified. I elected to use the Within-Set Dominant Alternative and Stability Validity tests to evaluate internal validity. Since failing these tests is not enough evidence for rejecting an entry, I decided to rely on the model performance characteristics to inform me whether conducting a separate analysis is warranted for the data segment with higher internal validity.

The different phases of data collection were Feasibility Study, Phase I with a small sample size, and Phase II with a large sample size. The justification, dependent and independent variables, sample size, inclusion criteria, and recruitment procedures were detailed for each phase. The software and online platforms used in data collection were Microsoft Excel, Python, R, Qualtrics, and Prolific. The survey design was described next with an overview of its sections, followed by a thorough description of data analysis modules. The data analysis modules aided with fraud detection, data quality assessment, internal validity assessment, data preprocessing, and choice modeling. Figure 11 illustrates the flow of data collection and analysis activities, provides an overview of all the modules and subcomponents, and presents a task-oriented summary of this chapter. The objectives and general steps involved in each module are listed below, and the software or online platform predominantly used in each stage is also identified in this figure.

**Figure 11:**

*Summary of Data Collection and Analysis Pipeline*



*Note:* The logos of Microsoft Excel, R, Python, Prolific, and Qualtrics XM are trademarks of their respective owners, used in this figure solely for non-commercial, academic, and decorative purposes with no implication of endorsement or affiliation.

## IV. Results

In this chapter, the empirical findings from the Discrete Choice Experiments are presented, with a specific focus on evaluating the influence of usability and accessibility information on the medical device purchasing decisions of People without Disabilities. The central objective of this research was elucidating the preferential inclinations and priorities of PwoD when presented with conventional medical device attributes such as Cost, Brand, and User Rating, augmented by a Usability/Accessibility Score, as a manifest of accessibility and usability information. These results have the potential to catalyze a paradigm shift via encouraging greater emphasis on the reporting of usability/accessibility information by responsible stakeholders. This shift not only will promote informed decision-making for all, fostering a more effective healthcare ecosystem, but will also ameliorate the disparities in healthcare outcomes experienced by People with Disabilities.

To bolster the validity, reliability, and generalizability of these findings, I studiously adhered to the available guidelines and best practices for deploying DCEs in the health econometrics context. Furthermore, I aimed to ensure proper sampling and data collection, and to verify data quality while mitigating the introduction of any unwarranted bias. The presentation of results in this chapter is organized according to the various phases of the study. Each section details the characteristics and objectives pertinent to the respective phase, followed by a sociodemographic synopsis of the sample, an analysis of survey findings extraneous to the DCE, a report on data quality indicators and internal validity, and ultimately, a detailed presentation of the DCE preference model and performance information.

The findings presented here directly address the research questions outlined in the introductory chapters and offer empirical insights into the decision-making patterns of PwOD in the context of this study. This chapter sets the foundation for the following Discussion chapter, where these results will be interpreted against the backdrop of existing literature, theoretical frameworks, and their pragmatic implications will be detailed.

## **Feasibility Study**

### ***Objectives***

- Evaluation of the data collection and analysis procedures
- Time requirement estimates
- Identifying potential improvements based on feedback

Overall, this phase aimed to evaluate the mechanics of data collection and analysis procedures (Fig. 11), comprehensibility of the instructions, burden on the participants, and eliciting feedback. Convenience sampling of colleagues and personal connections yielded 18 total submissions, consisting of nine complete and nine incomplete entries. I asked the participants to test the screening questions and investigate the enforcement of inclusion criteria, and the incomplete entries were the byproduct of this testing process. Data collection took place in June 2023.

### ***Recruitment Characteristics***

Prior to data collection, I was aware of the bias this sample may exhibit towards more usable and accessible products. In agreement with my expectations, six out of nine participants stated that they were very or extremely familiar with accessibility and disability, on a five-point

Likert-type scale. Due to the bias of this sample, I mainly focused on the analysis of data relevant to the objectives of this phase.

**Design Report**

I confirmed that the data collection and analysis modules were performing as expected, from choice matrix design to MNL analysis of the choice data. To evaluate the stages prior to data collection, I generated a  $D_z$ -Efficient choice matrix with  $D$ -error = 0.37. Table 4 shows the attribute level balance of this design. This matrix was subsequently converted to individual tables for each alternative and transferred to the Qualtrics Survey.

**Survey Results**

On average, the survey took roughly 10 minutes to complete ( $SD = 3.9$ ). In the Feedback Questionnaire, two out of nine participants indicated that they found it somewhat difficult to understand the purpose of the study, on a five-point Likert-type scale. The rest were either neutral ( $N = 1$ ), or found it somewhat or extremely easy ( $N = 5$ ). One participant found it difficult to assume the mindset of a person looking to buy a BPM, and the rest were either neutral or

**Table 4:**

*Attribute Level Balance of Feasibility Study Choice Matrix*

Cost		Brand		User Rating		Usability/Accessibility Score	
Levels	Counts	Levels	Counts	Levels	Counts	Levels	Counts
\$20	27	Omron	32	3.2 Stars	24	Unavailable	15
\$25	20	HomeHealth	32	4.0 Stars	22	Low (4/10)	16
\$30	17			4.8 Stars	18	Medium (6/10)	17
						High (8/10)	16

found it easy. Similarly, when asked about the difficulty of comparing the alternatives in each choice task, one participant indicated finding the task somewhat difficult, with the rest of the participants reporting it easy. For the time requirement of the survey, all participants indicated either neutrality or satisfaction, apart from one participant who was somewhat dissatisfied.

### ***Feedback Results and Potential Adjustment Needs***

The last question asked participants for their general feedback on the survey. One participant provided a helpful insight and mentioned that the tables for choice tasks were difficult to see. I followed up on this comment and learned that the uniform styling of the tables made them difficult to tell apart. For the subsequent phases, I adjusted the font colors slightly, so the attribute names were dark grey instead of black, which was uniform with the rest of the content of the choice tasks. I also adjusted the borders and padding of the tables to help each choice task stand out further. I confirmed with the participant that the aesthetic improvements were effective in addressing their comment.

After evaluating the experiences of participants and implementing the necessary adjustments, I tested the performance of the data analysis modules. Fraud analysis was unnecessary for this stage. I transformed the survey results into the long format, and manually confirmed the resulting spreadsheet. As mentioned earlier in this segment, I did not intend to utilize the results of MNL analysis due to the small sample size and high potential for bias. Nevertheless, I executed the MNL analysis script and studied the results to confirm the required output data were generated appropriately.

By the conclusion of MNL model estimation, I had successfully achieved the desired outcomes of this stage. I had confirmed that the entire data collection and analysis pipeline is

operating robustly, aside from fraud detection which was not tested. I had also established reasonable estimates for the burden on participants and time requirements. After addressing the feedback and making minute adjustments to the procedure overall, I concluded this phase.

**Phase I**

***Objectives***

- Data collection and analysis procedures validation
- Identification of the sample sociodemographic characteristics
- Data quality evaluation internal validity assessment
- Examination of task comprehensibility potential
- MNL model parameter estimation
- Participant burden evaluation based on self-reports

***Design Report and Recruitment***

I generated a  $D_z$ -Efficient choice matrix with  $D\text{-error} = 0.340$  and implemented the choice tasks into the Qualtrics survey. In Table 5, the attribute level balance of this matrix is presented. Despite the non-informative priors, I performed power analysis simulation and the

**Table 5:**

*Attribute Level Balance of Phase I Choice Matrix*

Cost		Brand		User Rating		Usability/Accessibility Score	
Levels	Counts	Levels	Counts	Levels	Counts	Levels	Counts
\$20	27	Omron	32	3.2 Stars	25	Unavailable	15
\$25	21	HomeHealth	32	4.0 Stars	23	Low (4/10)	16
\$30	16			4.8 Stars	16	Medium (6/10)	17
						High (8/10)	16

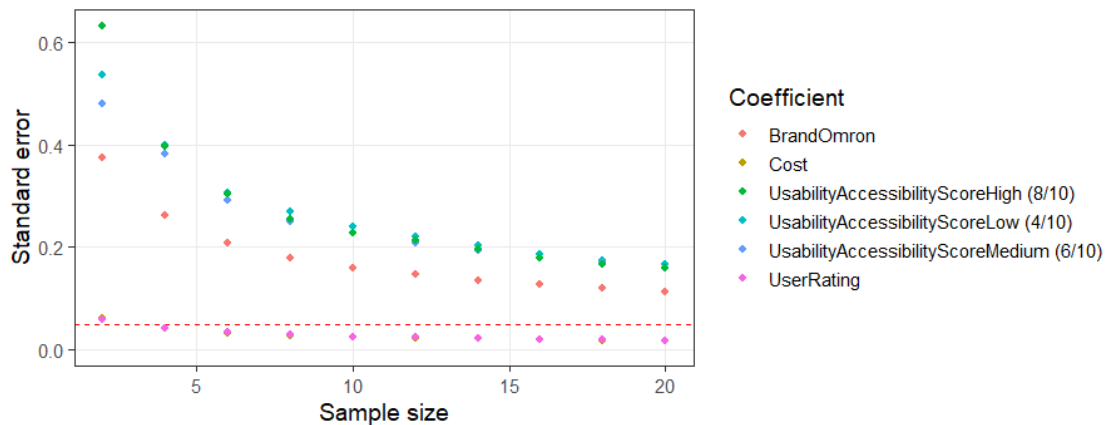
standard error estimates are depicted in Figure 12. A sex-balanced sample of 20 participants with approval ratings of above 90% received the survey through Prolific. These participants had stated that they were (a) above 50 years of age; (b) with no long-term health conditions or disabilities; (c) fluent in English, and (d) residing in the United States. Data collection from this sample took place in July 2023.

**Fraud Detection and Data Quality Assessment**

The fraud detection module identified two duplicate submissions, three discrepancies in ethnicity, and one discrepancy in state of residence. Through manual examination, I was able to verify this report. The participants with multiple entries had failed the screener validations previously, retaken the survey, and changed their responses to qualify. These entries were excluded from the subsequent analyses. Ethnicity discrepancies had arisen from the labeling mismatch between the survey and Prolific categories and were deemed inconsequential. The participant with state discrepancy also did not exhibit any other violations, and their entry was retained. Figure 13 shows a summary of fraud detection and data quality assessment results for all 20 participants, after automatic and manual inspection.

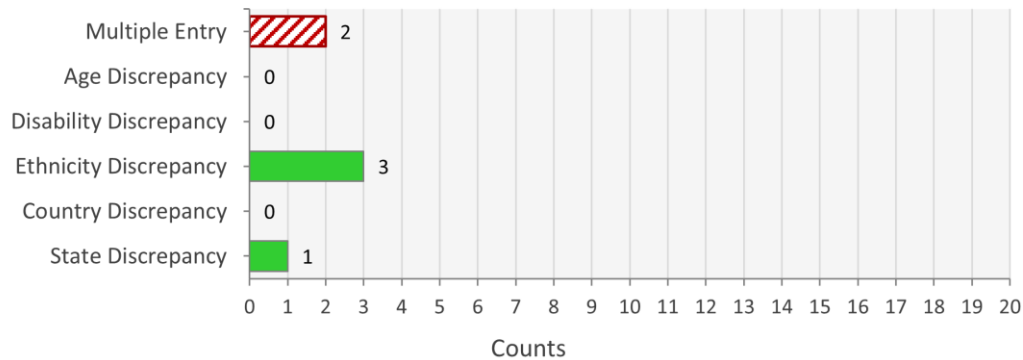
**Figure 12:**

*Parameter SE Simulation for Increments of Sample Size in Phase I*



**Figure 13:**

*Phase I Violations Summary After Automatic and Manual Inspection*



### ***Sociodemographic Characteristics***

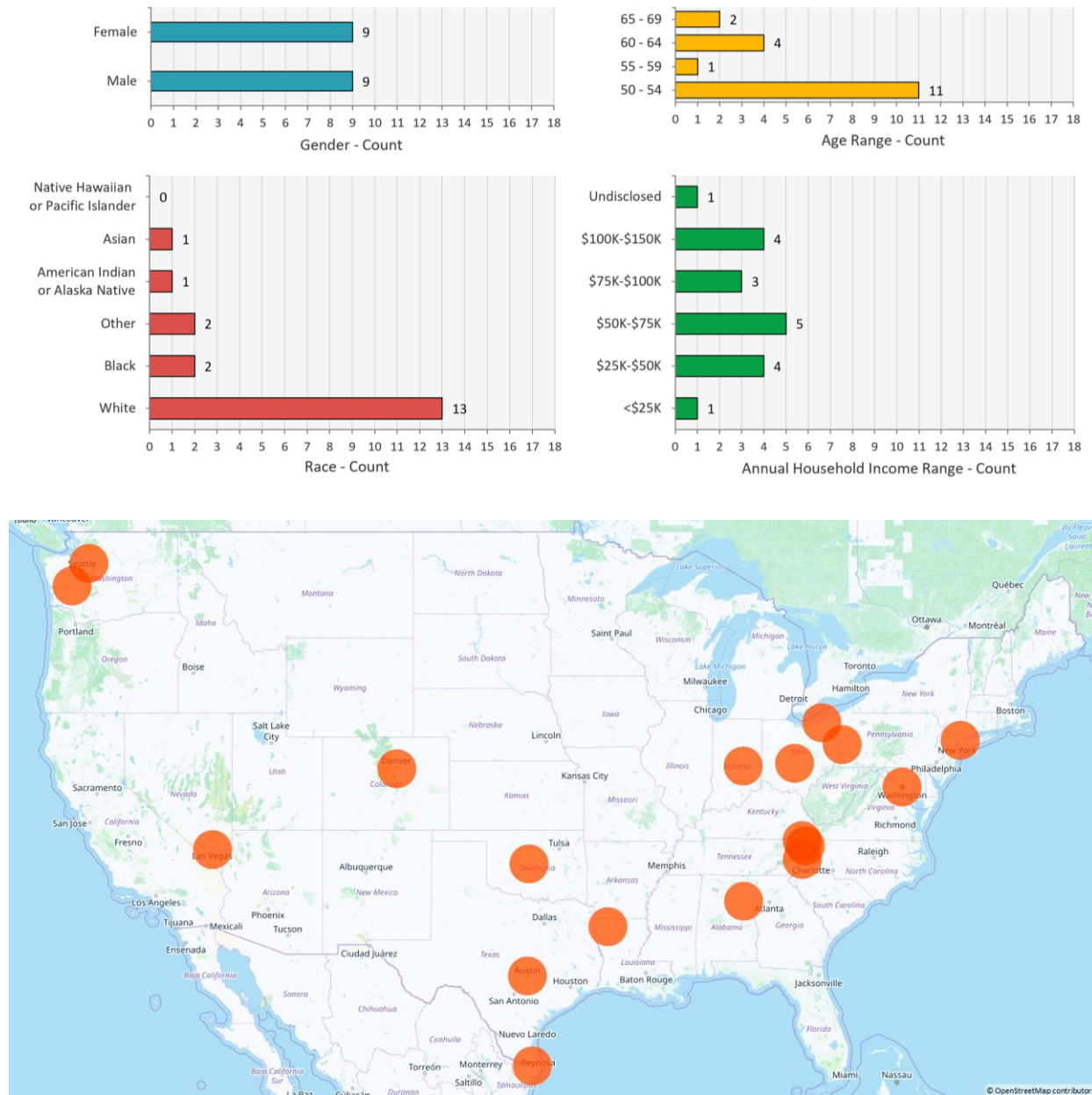
The remaining eligible participants ( $N = 18$ ) comprised an equal distribution of 50% male and 50% female, according to their responses to the sociodemographic questions. Their racial background was as follows: 76.5% White, 11.8% Black, 5.9% American Indian or Alaska Native, 5.9% Asian, and the remaining 11.8% selected “Other.” It is important to note that participants had the option to select more than 2 one race. The age categories placed 61.1% between 50-54, 5.6% between 55-59, 22.2% between 60-64, and 11.1% between 65-69. Regarding annual household income, 5.6% reported earning less than \$25K, 22.2% in the \$25K-\$50K range, 27.8% in the \$50K-\$75K range, 16.7% in the \$75K-\$100K range, and 22.2% in the \$100K-\$150K range. One participant chose not to disclose their annual household income. The sociodemographic data were complete with no missing information. Figure 14 shows the sociodemographic data as well as the geographical distribution of this phase’s participants.

### ***Potential for Context Resonance***

Due to the vital significance of task realism for the validity and reliability of the subsequent conclusions, the likelihood that a participant may find the hypothetical choice tasks

**Figure 14:**

*Sociodemographic Data and Geographical Distribution of Participants in Phase I*



pertinent and relatable was estimated based on the responses to two questions before, and four questions after engagement with the DCE. When asked whether they use a BPM or know someone in their friends or family who does, 61.1% reported using one themselves, or having friends or family members who use BPMs. The other questions asked if the participants were

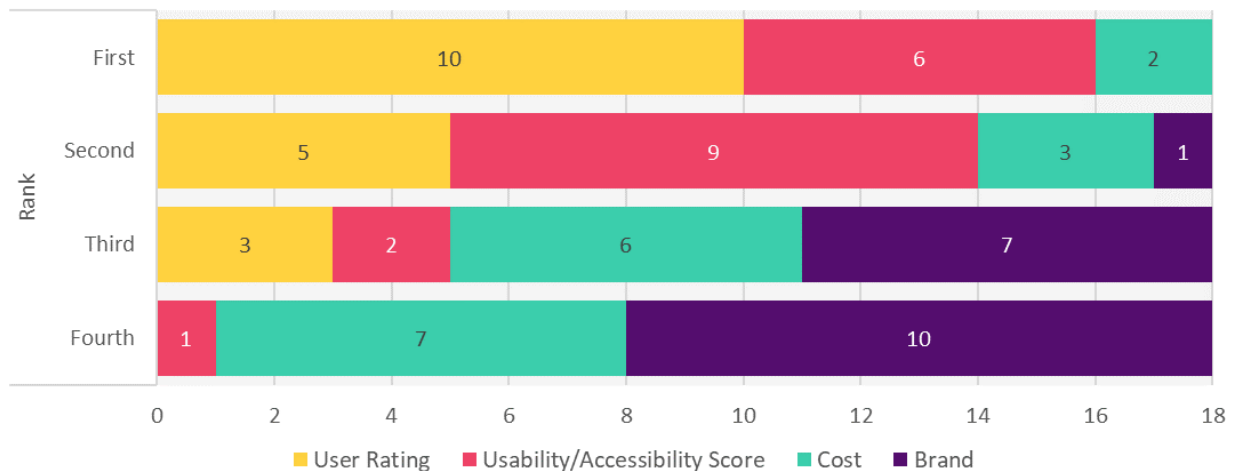
able to imagine needing to purchase a BPM for themselves or a friend or family member. All participants responded positively to this question with no exceptions.

Immediately after engaging with the DCE (approximately 20 questions later) the participants were asked four additional questions. Two questions aimed to further examine potential topic pertinency, and the other two were directly observing the attitudes towards the role of Usability/Accessibility Score. Participants were asked if they currently own any home medical devices and 77.8% responded positively. They also rated their familiarity with disability and accessibility concepts on a five-point Likert-type scale, and the responses were 16.7% very familiar, 33.3% moderately familiar, and 50.0% were slightly familiar and not familiar at all.

When prompted to explicitly rank the importance of the four study attributes in their choices, the participants collectively ranked User Rating and Usability/Accessibility Information as the first and second most important factors, respectively. Cost and Brand were ranked lower, claiming the third and fourth place in decision priorities. The average ranks based on 18 responses were 1.6 for User Rating, 1.9 for Usability/Accessibility Score, 3.0 for Cost, and 3.5 for

**Figure 15:**

*Distribution of Participants' Priority Ranking of the Study Attributes in Phase I*



Brand (see Figure 15). Lastly, using a five-point Likert-type scale, 70.0% of participants rated the availability of Usability/Accessibility Score as very or extremely important in their decisions.

All participants responded to the questions discussed in this segment and the data were complete. The placement of these questions was discussed previously in the Instrumentation section, under Task Comprehensibility Potential and Follow-Up Questions. While these responses alone neither guarantee nor directly measure task realism and pertinence, they may evidence the inclination of participants to resonate with the study question.

**Internal Validity Assessment**

All participants had successfully passed the Dominant Alternatives (Within-Set Dominant Alternative) and Choice Consistency (Stability Validity) tests.

**Multinomial Logit Analysis of Preferences**

Table 6 shows the estimated part-worth utilities of attributes. Model performance and the variance-covariance matrix are shown in Tables 7 and 8, respectively.

**Table 6:**

*Phase I Preference Model Parameter Estimates*

Parameter	Estimate	SE	95% CI <sup>b</sup>	z	p
Cost	-0.131	0.048	-0.23, -0.04	-2.720	0.007
Brand					
HomeHealth <sup>a</sup>	-	-	-	-	-
Omron	0.416	0.261	-0.09, 0.93	1.595	0.111
User Rating	3.242	0.358	2.6, 4.0	9.060	<.001
Usability/Accessibility Score					
Unavailable <sup>a</sup>	-	-	-	-	-
Low (4/10)	-1.795	0.629	-3.0, -0.58	-2.852	0.004
Medium (6/10)	1.893	0.403	1.1, 2.7	4.696	<.001
High (8/10)	5.237	0.574	4.1, 6.4	9.119	<.001

<sup>a</sup> Reference Category, <sup>b</sup> Confidence Interval

**Table 7:***Phase I MNL Model Performance Indicators*

Indicator	Value
Log-Likelihood	-112.502
Null Log-Likelihood	-399.253
AIC	237.003
BIC	258.981
McFadden $R^2$	0.718 <sup>a</sup>
Adjusted McFadden $R^2$	0.703
Prediction Success Rate <sup>b</sup>	77.43%

*Note.* Number of Observations = 288 (Sample Size × Number of Choice Tasks)

<sup>a</sup> Values between 0.2 and 0.4 represent an “excellent fit” (McFadden, 1974)

<sup>b</sup> Using the model to predict the outcomes of each choice task within the same dataset

**Table 8:***Phase I Parameter Estimates Variance-Covariance Matrix*

Parameter	1	2	3	4	5	6
	Cost	Brand - Omron	User Rating	UAS <sup>a</sup> - Low (4/10)	UAS <sup>a</sup> - Medium (6/10)	UAS <sup>a</sup> - High (8/10)
Cost	0.002					
Brand - Omron	0.003	0.068				
User Rating	-0.004	0.035	0.128			
UAS <sup>a</sup> - Low (4/10)	-0.004	0.010	0.041	0.396		
UAS <sup>a</sup> - Medium (6/10)	-0.011	-0.005	0.070	0.079	0.163	
UAS <sup>a</sup> - High (8/10)	-0.016	0.025	0.166	0.108	0.175	0.330

<sup>a</sup> UAS: Usability/Accessibility Score

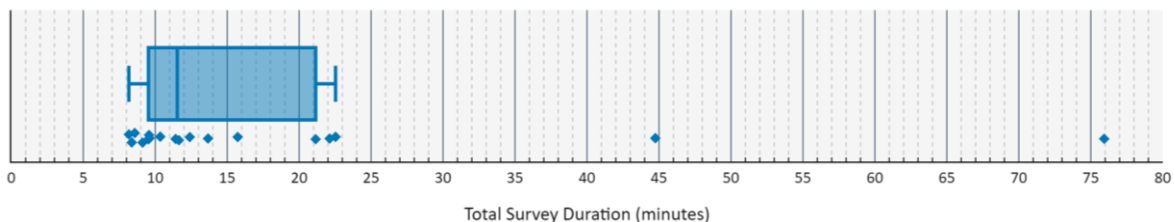
### ***Participant Burden and Feedback***

On average, the survey took about 18 minutes to complete, however, some participants spent significantly more time than others ( $SD = 16.48$ ). Figure 16 shows that most participants spent between about 8 to 14 minutes, while contextualizing the outliers. On a five-point Likert-type scale (a) 88.9% indicated their satisfaction with the length of the survey; (b) 94.5% agreed that the instructions were simple and clear; (c) 100% found the study's purpose easy to understand; (d) 100% found it easy to assume that they were looking to purchase a BPM, and (e) 94.4% expressed having no difficulties when comparing hypothetical products to make a purchasing decision. These ratings were performed by all participants, with no missing data.

Lastly, to ensure the appropriateness of study attribute levels, an open-ended question asked participants whether they needed more information about any other factors to make their decisions. Of 17 responses, 52.9% said they had all the information they needed, 11.8% mentioned product durability, 11.8% mentioned battery or electricity requirements, and 5.9% mentioned accuracy. The remaining participants mentioned they would have liked to know the manufacturer's country and history, and the product's warranty.

**Figure 16:**

*Time Requirement of Phase I Survey*



### ***Potential Adjustment Needs***

Based on the data collected in this phase, and the explicit or implicit feedback from the participants, the following areas were identified as potential improvements for the next phase:

1. Bot detection capabilities are not enabled by default in Qualtrics, and they needed to be enabled.
2. Multiple submission rejection is also not enabled by default and needed to be enabled.
3. Some participants spent significantly more time than others on the survey. Timing needed to be added to as many questions as possible to identify where the additional time is being spent. This also had the added benefit of identifying if some participants were not spending enough time on the instructions, or other questions.
4. The consent form overtly disclosed the eligibility criteria, which may have led to multiple entries with changed responses to screeners, to qualify for the study. Also, the screener validation was using leading questions. These issues needed to be addressed to avoid fraudulent entries.
5. Only crucial questions, such as consent, screener validation, and DCE choice tasks were mandatory. This allowed some participants to skip some questions. All questions needed to be made mandatory, except for sociodemographic questions. This would prevent participants from finishing the survey without having answered important questions.

6. Attention checks needed to be included to improve data quality assessment procedures. An explicit commitment request, as well as an open-ended question with explicit effort requirements (e.g., in two sentences describe A or B) would also aid in that process.
7. Instruction comprehension needed to be examined immediately after providing the instructions.

With all the potential improvements identified, I concluded Phase I of this study.

## **Phase II**

### ***Objectives***

- Sample sociodemographic analysis
- Data quality and internal validity assessment
- Task comprehensibility examination
- Preference model estimation
- Participant burden evaluation

### ***Design Report and Recruitment***

I used the parameter estimates from Phase I that had reached statistical significance as priors to design a  $D_b$ -Efficient choice matrix for this phase. Methodical evaluation of 229 designs identified one with  $D\text{-error} = 0.543$  which maintained an acceptable attribute level balance (see Table 9). Power analysis simulation with informative priors from the previous phase, yielded parameter standard error estimates for increments in sample size, as depicted in Figure 17. A sex-balanced sample of 179 participants who satisfied the inclusion criteria, with approval

**Table 9:**

*Attribute Level Balance of Phase II Choice Matrix*

Cost		Brand		User Rating		Usability/Accessibility Score	
Levels	Counts	Levels	Counts	Levels	Counts	Levels	Counts
\$20	25	Omron	31	3.2 Stars	20	Unavailable	17
\$25	19	HomeHealth	33	4.0 Stars	19	Low (4/10)	19
\$30	20			4.8 Stars	25	Medium (6/10)	17
						High (8/10)	11

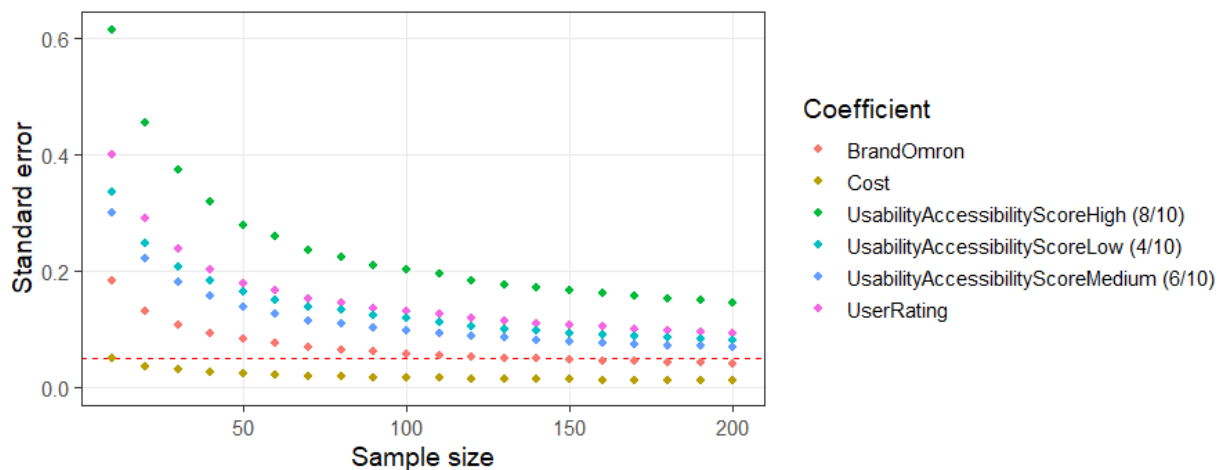
ratings higher than 90%, were recruited via Prolific. In addition, participants from Phase I were excluded from recruitment. Data collection took place in November 2023.

***Fraud Detection and Data Quality Assessment***

The multiple submission prevention feature on Qualtrics proved ineffective, and four participants were identified as having duplicate submissions. These participants had changed their responses to the screener validation questions, after being disqualified the first time.

**Figure 17:**

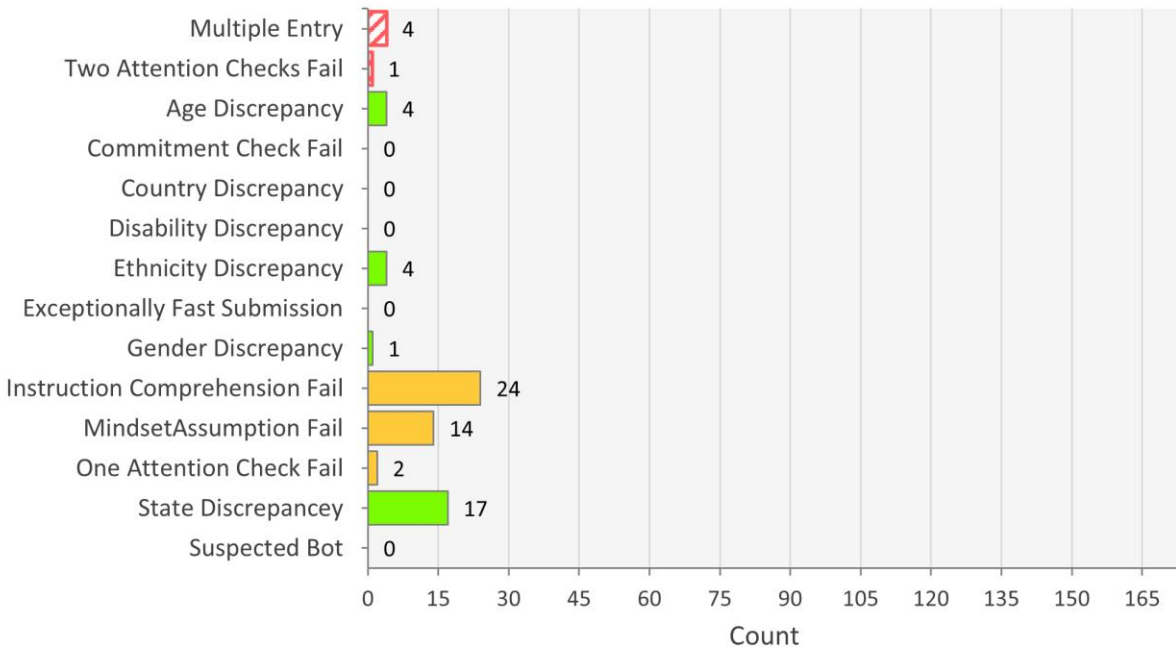
*Parameter SE Simulation for Increments of Sample Size in Phase II*



Additionally, one participant had failed both attention checks, and manual review showed incoherent and irrelevant responses. These five entries were discarded, and the subsequent analyses commenced with 174 participants. Similar to Phase I, many discrepancies were justifiable due to the differences in labeling the categories and the phrasing of questions. In some sociodemographic categories, such as age, some discrepancies arose since the questions were not mandatory and a few participants elected not to disclose the answer. I found that the IQR criteria for page submission times and click counts was perhaps too strict, as it raised many flags while the manual review of the data showed the entries were within reason. Figure 18 shows a summary of violations identified for the qualifying participants, after automatic and manual inspection.

**Figure 18:**

*Phase II Violations Summary After Automatic and Manual Inspection*

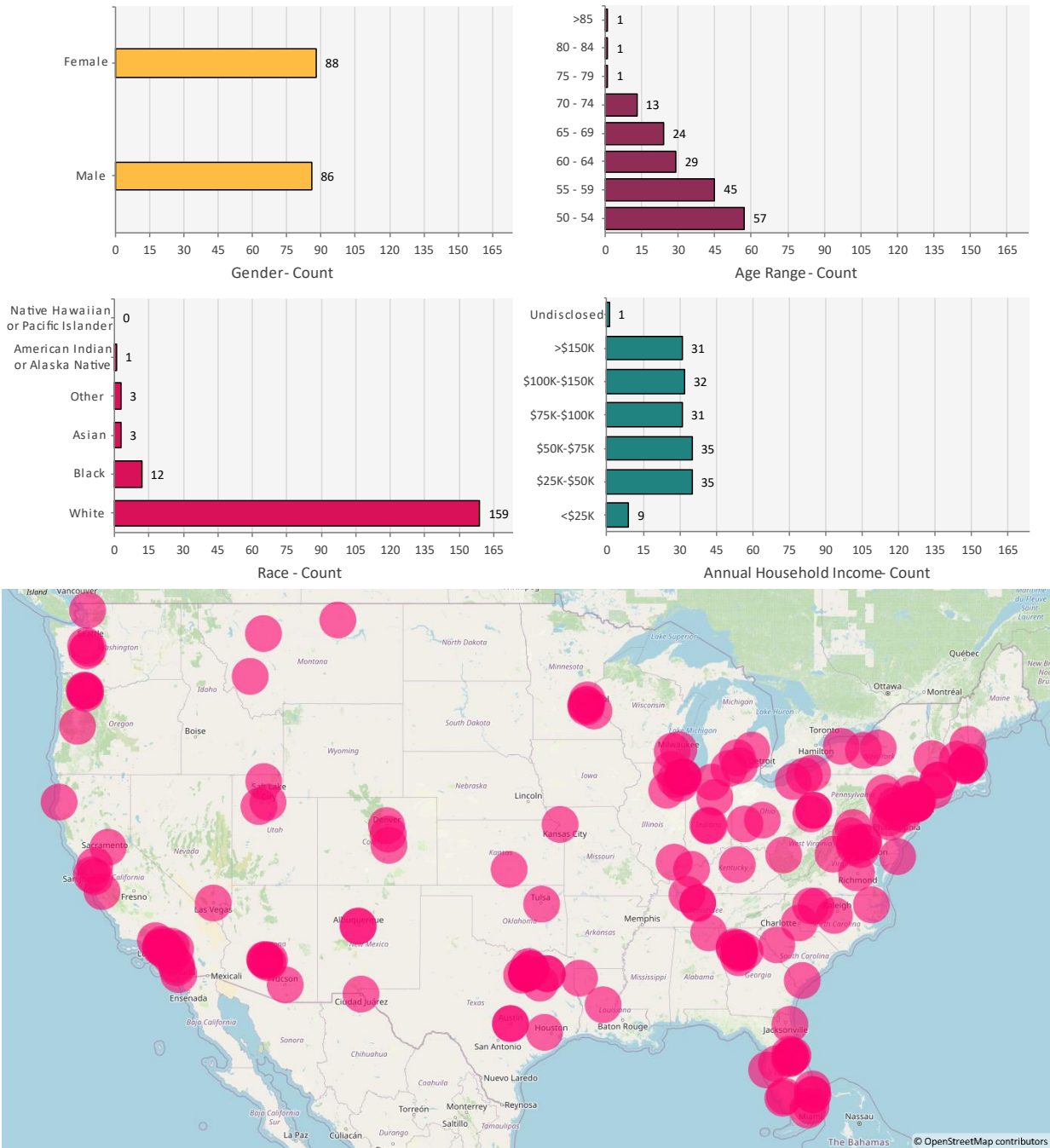


### ***Sociodemographic Characteristics***

This phase's sociodemographic profile is reported similarly to Phase I, based on gender, race, age groups, and annual household income characteristics. The gender composition of the study cohort was nearly balanced, with females representing 50.6% and males accounting for 49.4%. The racial composition showed most participants were White, constituting 91.4% of the total participants. Black participants made up 6.9%, Asians 1.7%. American Indians or Alaska Natives 0.6%, and Native Hawaiian or Pacific Islanders were not represented in the participants. The age distribution of the participants primarily fell within the 50 to 54 age group, which comprised 33.3% of the total. This was followed by the 55-59 age group at 26.3%, the 60-64 age group at 17.0%, and the 65-69 age group at 14.0%. Participants aged 70-74 accounted for 7.6%, and those in each of the 75-79, 80-84, and 85 or above age groups constituted 0.6% respectively. The annual household income of the participants varied, with 20.1% each in the \$25K-\$50K and \$50K-\$75K income brackets. Those earning between \$75K-\$100K and \$100K-\$150K were similarly represented at 17.8% and 18.4%, respectively. Participants with a household income of \$150K or more also constituted 17.8% of the sample. The lowest income group, earning less than \$25K, accounted for 5.2% of the participants, and a marginal 0.6% preferred not to disclose their income. This overview accounts for all participants, except for the age category, where three participants opted not to respond. Figure 19 provides an overview of the sociodemographic composition of the participants and depicts their geographical distribution across the United States.

**Figure 19:**

*Sociodemographic Data and Geographical Distribution of Participants in Phase II*



**Potential for Context Resonance**

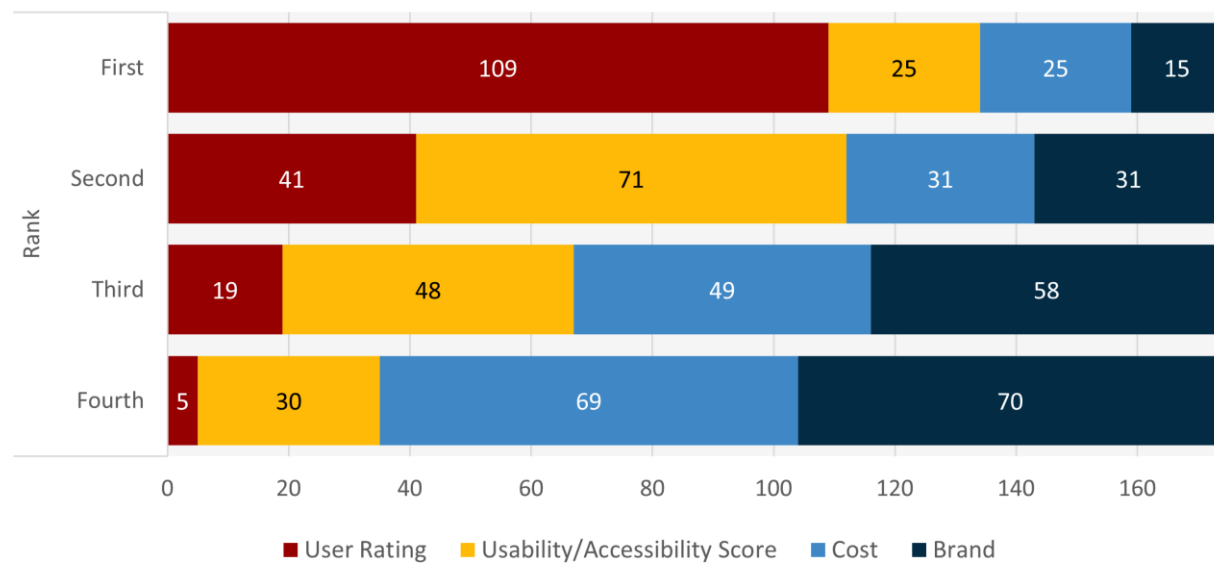
Similar to Phase I, two questions before and four questions after the DCE aimed to evaluate the pertinence potential of the hypothetical choice tasks to the participants. When

asked about using BPMs, or having friends or family members who do, 75.3% responded positively. The question on the participants' ability to assume they need to purchase a BPM for themselves, or a friend or family member, elicited 92.0% positive responses. After engaging with the DCE, 77.0% of the participants stated they own medical devices at home. In the next question, using a five-point Likert-like scale, 17.8% rated themselves as familiar with disability and accessibility concepts, 38.5% as moderately familiar, and 35.5% as unfamiliar or slightly familiar.

Participants' rankings of the study attributes placed User Rating first, with an average rank of 1.54. Usability/Accessibility Score claimed the second rank with an average of 2.48. Cost and Brand placed third and fourth very closely, with 2.93 and 3.05 average ranks, respectively. Figure 20 shows how the priority rankings were distributed among the four attributes in this phase. Lastly, rating on a five-point Likert-type scale, 31.6% stated that the presence of

**Figure 20:**

*Distribution of Participant's Priority Ranking for the Study Attributes in Phase II*



Usability/Accessibility Score did little to impact their selections, while 36.8% claimed the impact moderate, and 31.6% indicated it played an important role. All participants responded to all six questions with no data missing.

**Internal Validity Assessment**

All participants successfully selected the Dominant Alternative, and 73.6% remained consistent in their decision-making and selected the same alternative when presented with the same choice task.

**Preference Model Estimation**

Part-worth utilities of attributes were estimated using a MNL model, as tabulated in Table 10. Tables 11 and 12 show the model performance and the variance-covariance matrix of parameter estimates respectively.

**Table 10:**

*Phase II Preference Model Parameter Estimates*

Parameter	Estimate	SE	95% CI <sup>b</sup>	z	p
Cost	-0.060	0.010	-0.08, -0.04	-6.199	<.001
Brand					
HomeHealth <sup>a</sup>	-	-	-	-	-
Omron	0.410	0.042	0.33, 0.49	9.743	<.001
User Rating	2.308	0.073	2.2, 2.4	31.550	<.001
Usability/Accessibility Score					
Unavailable <sup>a</sup>	-	-	-	-	-
Low (4/10)	-0.855	0.069	-0.99, -0.72	-12.382	<.001
Medium (6/10)	1.219	0.063	1.1, 1.3	19.383	<.001
High (8/10)	2.740	0.107	2.5, 2.9	25.556	<.001

<sup>a</sup> Reference Category, <sup>b</sup> Confidence Interval

**Table 11:***Phase II MNL Model Performance Indicators*

Indicator	Value
Log-Likelihood	-2903.285
Null Log-Likelihood	-3859.444
AIC	5818.569
BIC	5854.159
McFadden $R^2$	0.248 <sup>a</sup>
Adjusted McFadden $R^2$	0.246
Prediction Success Rate <sup>b</sup>	42.89%

*Note.* Number of Observations = 2784 (Sample Size × Number of Choice Tasks)

<sup>a</sup> Values between 0.2 and 0.4 represent an “excellent fit” (McFadden, 1974)

<sup>b</sup> Using the model to predict the outcomes of each choice task within the same dataset

**Table 12:***Phase I Parameter Estimates Variance-Covariance Matrix*

Parameter	1	2	3	4	5	6
	Cost	Brand - Omron	User Rating	UAS <sup>a</sup> - Low (4/10)	UAS <sup>a</sup> - Medium (6/10)	UAS <sup>a</sup> - High (8/10)
Cost	0.000					
Brand - Omron	0.000	0.002				
User Rating	0.000	0.000	0.005			
UAS <sup>a</sup> - Low (4/10)	0.000	0.000	-0.002	0.005		
UAS <sup>a</sup> - Medium (6/10)	0.000	-0.000	0.002	0.001	0.004	
UAS <sup>a</sup> - High (8/10)	0.000	0.001	0.006	-0.001	0.004	0.011

<sup>a</sup> UAS: Usability/Accessibility Score

While the MNL model for Phase II appears to have performed very well in differentiating the part-worth utility for all attributes based on the significance of all parameter estimates,

small  $CIs$ , and McFadden  $R^2$ , the prediction success rate is 42.89%. The implication is that the MNL model, while still robust in identifying the part-worth utilities, falls short of accurately estimating all choice outcomes. It appeared that at least one of the core assumptions of the MNL model had been violated (IIA, IID, or preference homogeneity, discussed previously on page 60). The findings presented in Figure 20 hinted that the preferences may be heterogenous, which triggered an a posteriori investigation.

Traditionally, two extensions of MNL analysis are utilized in accommodating such violations: the Mixed Logit model (MXL), and the Latent Class MNL model (LC-MNL) (Greene & Hensher, 2003). The MXL model relaxes the IIA and preference homogeneity assumptions of the MNL, however, it requires making deliberate assumptions about the shape of parameter distributions across participants (Kjaer, 2005; McFadden & Train, 2000). In contrast, the LC-MNL model does not demand these assumptions and instead, accounts for preference heterogeneity by dividing participants into unobserved and discrete partitions on their preference parameters (Greene & Hensher, 2003; Hoedemakers et al., 2022). The LC-MNL approach, after segmenting the participants, estimates the MNL model for each class, thereby assuming that the restrictive MNL assumptions hold for each class, instead of the total observations pool (Sarrias & Daziano, 2017).

The LC-MNL model seemed congruent with the a priori assumptions of this study. I used the `gmnl` package (Sarrias & Daziano, 2017) to estimate the LC-MNL model for two to six classes. Based on the  $AIC$ ,  $BIC$ , and Log-Likelihood, increasing the number of classes consistently improved the model fit. Attempts to construct beyond seven classes were unsuccessful. Table 13 shows the gradual improvement of LC-MNL model over the MNL model as the number of

classes increases. I estimated the LC-MNL model with the assumption that the participants comprise of at least six distinct segments. Table 14 presents the model parameter estimates, and Figure 21 provides a representation of how the six classes differ in their parameter estimates, to aid in the visualization of class differences in part-worth utility.

The individual-specific probabilities were extracted from the model. These parameters determine the likelihood that each participant belongs to one of the identified classes. Utilizing these probabilities, along with the Kruskal-Wallis test and Chi-square test of independence, I was able to determine the summary profiles of the six classes based on the survey data. This approach not only facilitated a nuanced understanding of each class, but also allowed for a comparison of the distinctive features and underlying tendencies of different classes. Characteristics such as sociodemographic factors, preferences, and inferred experiment participation behaviors were examined. This comprehensive analysis of class profiles has important implications for a more nuanced and precise understanding of the findings. Figure 22 highlights the summary characteristics of various classes if the test statistic for the data indicated significant differences between classes.

**Table 13:**

*Gradual Improvement of the LC-MNL Model Performance as the Number of Classes Increases*

Performance Indicator	MNL Model	LC-MNL Models				
		2 Classes	3 Classes	4 Classes	5 Classes	6 Classes
AIC	5818.569	5099.135	4846.729	4859.937	4460.541	4373.173
BIC	5854.159	5176.172	4965.247	5019.935	4662.021	4616.134
Log-Likelihood	-2903.285	-2536.568	-2403.365	-2402.968	-2196.270	-2145.587

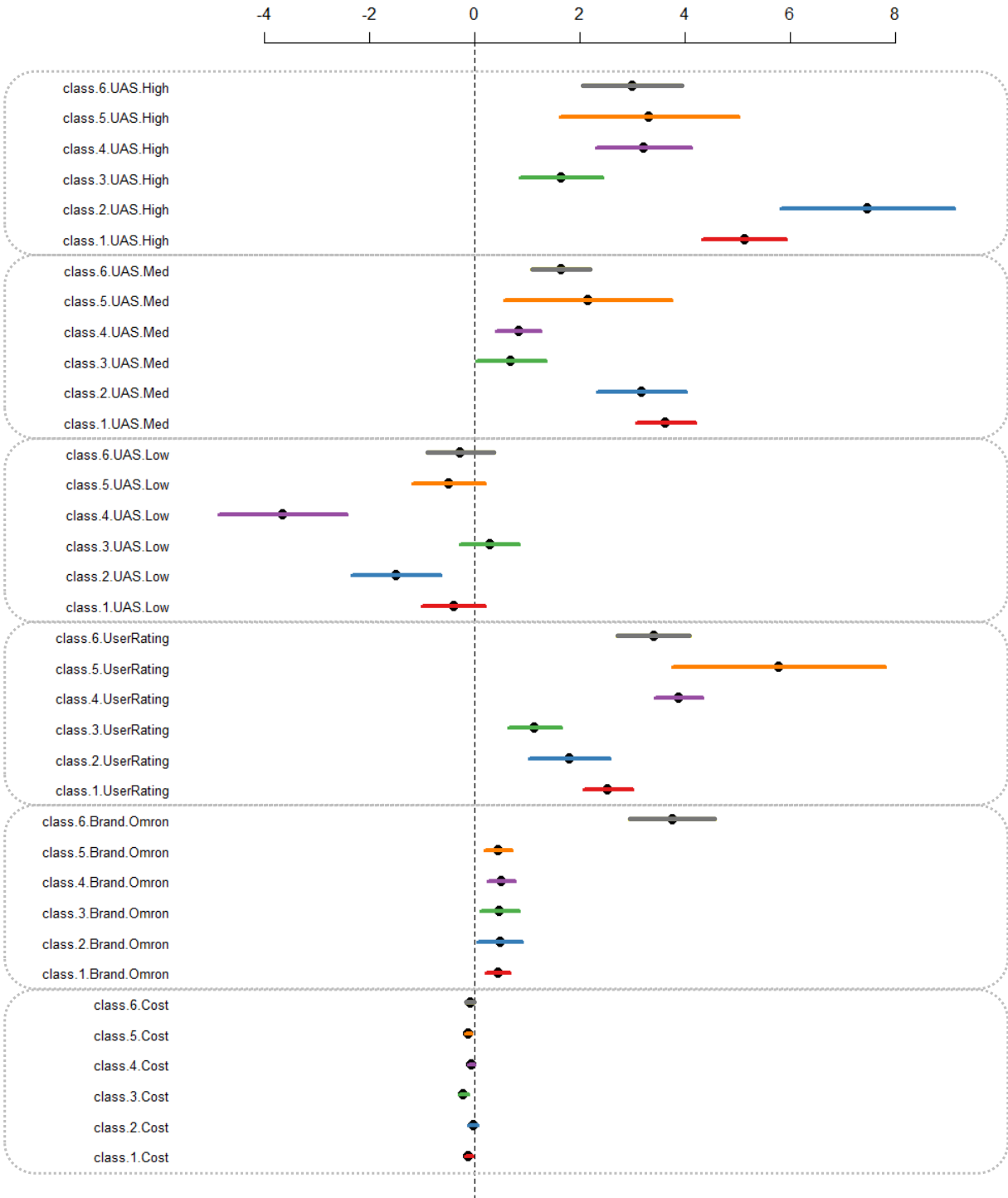
**Table 14:***Six-Class LC-MNL Model Parameter Estimates*

Parameter	Estimate	SE	95% CI <sup>a</sup>	z	p
class.1.Cost	-0.115	0.033	-0.18, -0.05	-3.462	<.001
class.1.Brand.Omron	0.442	0.113	0.22, 0.66	3.898	<.001
class.1.UserRating	2.532	0.232	2.08, 2.99	10.915	<.001
class.1.UAS.Low	-0.404	0.306	-1, 0.2	-1.322	0.186
class.1.UAS.Med	3.625	0.286	3.07, 4.19	12.686	<.001
class.1.UAS.High	5.124	0.401	4.34, 5.91	12.763	<.001
class.2.Cost	-0.030	0.047	-0.12, 0.06	-0.635	0.525
class.2.Brand.Omron	0.477	0.215	0.06, 0.9	2.225	0.026
class.2.UserRating	1.797	0.387	1.04, 2.56	4.644	<.001
class.2.UAS.Low	-1.485	0.430	-2.33, -0.64	-3.453	<.001
class.2.UAS.Med	3.170	0.433	2.32, 4.02	7.313	<.001
class.2.UAS.High	7.465	0.836	5.83, 9.1	8.927	<.001
class.3.Cost	-0.215	0.046	-0.3, -0.13	-4.727	<.001
class.3.Brand.Omron	0.474	0.184	0.11, 0.84	2.568	0.010
class.3.UserRating	1.140	0.258	0.63, 1.65	4.418	<.001
class.3.UAS.Low	0.285	0.283	-0.27, 0.84	1.005	0.315
class.3.UAS.Med	0.688	0.338	0.03, 1.35	2.037	0.042
class.3.UAS.High	1.649	0.400	0.86, 2.43	4.120	<.001
class.4.Cost	-0.067	0.032	-0.13, 0	-2.111	0.035
class.4.Brand.Omron	0.501	0.129	0.25, 0.75	3.891	<.001
class.4.UserRating	3.875	0.231	3.42, 4.33	16.778	<.001
class.4.UAS.Low	-3.652	0.619	-4.87, -2.44	-5.900	<.001
class.4.UAS.Med	0.832	0.214	0.41, 1.25	3.892	<.001
class.4.UAS.High	3.206	0.458	2.31, 4.1	6.997	<.001
class.5.Cost	-0.126	0.029	-0.18, -0.07	-4.272	<.001
class.5.Brand.Omron	0.451	0.131	0.19, 0.71	3.430	<.001
class.5.UserRating	5.784	1.029	3.77, 7.8	5.620	<.001
class.5.UAS.Low	-0.495	0.347	-1.18, 0.19	-1.427	0.154
class.5.UAS.Med	2.152	0.811	0.56, 3.74	2.652	0.008
class.5.UAS.High	3.312	0.864	1.62, 5.01	3.834	<.001
class.6.Cost	-0.084	0.033	-0.15, -0.02	-2.560	0.010
class.6.Brand.Omron	3.757	0.416	2.94, 4.57	9.040	<.001
class.6.UserRating	3.402	0.356	2.7, 4.1	9.553	<.001
class.6.UAS.Low	-0.272	0.324	-0.91, 0.36	-0.840	0.401
class.6.UAS.Med	1.634	0.285	1.07, 2.19	5.728	<.001
class.6.UAS.High	2.989	0.477	2.05, 3.92	6.269	<.001
(class)2	-0.683	0.077	-0.83, -0.53	-8.822	<.001
(class)3	-1.092	0.107	-1.3, -0.88	-10.223	<.001
(class)4	0.197	0.085	0.03, 0.36	2.327	0.020
(class)5	0.385	0.088	0.21, 0.56	4.393	<.001
(class)6	-0.526	0.076	-0.67, -0.38	-6.954	<.001

<sup>a</sup> Confidence Interval

**Figure 21:**

*Six-Class LC-MNL Parameter Estimate Variations Per Class*



Note. UAS = Usability/Accessibility Score

**Figure 22:**

*Survey Data that Differed Between Classes with Statistical Significance*



*Note.*  $N_{Class1} = 34$ ,  $N_{Class2} = 17$ ,  $N_{Class3} = 12$ ,  $N_{Class4} = 43$ ,  $N_{Class5} = 47$ , and  $N_{Class6} = 20$ . For the significance of between-class differences, in order from the top left to bottom right,  $p = 0.041, 0.008, 0.054, <0.001, <0.001, \text{ and } 0.009$ .

The categories depicted in Figure 22 only encompass the responses that were found to have statistically significant differences between classes. These six categories were identified after comparing the class responses to the following 17 questions:

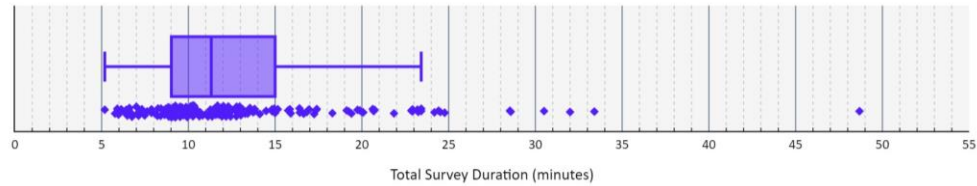
- Sociodemographic characteristics: Age, gender, and annual household income
- Propensity to find the study personally relevant: BPM user, ability to assume the mindset of needing to purchase a BPM, owning other medical devices, familiarity with disability and accessibility, rankings of attributes, and ratings of the importance of Usability/Accessibility Score in decision-making
- Survey participation characteristics: Total survey duration, instruction comprehension, and choice consistency (Stability Validity)
- Self-reported attitudes towards the study: Satisfaction with the survey length, clarity of instructions, difficulty of understanding the study purpose, difficulty of assuming the appropriate mindset, and difficulty of making decisions in choice tasks.

### ***Participant Burden Evaluation***

Participants on average spent 12.97 minutes on the survey ( $SD = 6.21$ ). There were relatively fewer outliers compared to the portion of outliers in Phase I (see Figure 23). Overall, 90.2% of participants expressed on a five-point Likert-type scale that they were satisfied with the time requirement of the experiment. On the other hand, 6.9% expressed neutrality, and 2.8% were dissatisfied. The ratings on the clarity of instructions indicated that 89.1% found the instructions clear, while 10.9% strongly disagreed. Additionally, 79.9% expressed they found the study easy to understand, 17.8% selected the neutral option, and 2.3% found it somewhat

**Figure 23:**

*Time Requirement of Phase II Survey*



difficult. On assuming the mindset of needing to purchase a BPM, 95.4% found this task easy, 3.4% were neutral, and 1.1% found it somewhat difficult. Most participants found it easy to compare the alternatives in choice tasks and make a final decision, comprising 89.7% of the ratings, while 6.3% rated the task as neither easy nor difficult, and 4.0% found it somewhat difficult.

The last question asked the participants to leave any comments they may have or share their general feedback. While the question was voluntary, 159 participants responded, expressing a variety of positive responses. Many found the study engaging, describing it as a “fascinating study”, “fun survey”, and “interesting quiz”. Some expressed gratitude, with remarks like “good luck with this important research!” and “thank you for the opportunity to participate”. The survey's clarity and relevance were also noted by comments such as “I like the clearness of this survey and the focus for what you needed”, “excellent project - no errors or issues, whatsoever”, or “the study was relevant to my life and I had no trouble putting myself in the place of someone who would need to choose between these products”. Some participants expressed their desire for more information about product features, for example “In real life purchase decisions, the more info, the better.” Personal connections to the product were also mentioned, including plans to purchase related items. A few participants simply acknowledged

their participation. Overall, the feedback was positively skewed, with participants appearing actively engaged and appreciative, alongside offering suggestions for enhancements, and sharing their personal connections with the topic.

## **Key Findings Summary**

### ***Feasibility Study***

The Feasibility Study established the mechanical validity and viability of the data collection and analysis procedures and aided in identifying necessary improvements to the survey presentation. It also helped set the expectations for time and cognitive burden on the participants. A  $D_z$ -Efficient design with  $D$ -error = 0.37 was used. However, due to the known bias of the sample, and their differentiation with the target sample size, the results were deemed inconsequential to the design of the next phase.

### ***Phase I***

The examination of 10 different designs for Phase I helped identify a  $D_z$ -Efficient design with  $D$ -error = 0.340. The balance of this design was the least compromised compared to the other designs examined. The numerical variables (Cost and User Rating) had higher variations in attribute level frequencies, however, the categorical variables (Brand and Usability/Accessibility Score) remained balanced. The  $SE$  of study parameters were simulated for the increments of sample size, and the results showed a flattening curve for  $SE$  with between 15 to 20 participants.

After data collection from 20 eligible participants, facilitated by Prolific, the data were thoroughly inspected for evidence of fraudulent participation. Two participants had failed the screener validation questions and made a second attempt with different responses. After

discarding the entries from these participants, other violations and discrepancies found in the data appeared justifiable. No concerning irregularities were observed in the sociodemographic composition of the sample or their geographical distribution.

Internal validity assessment showed that all participants had successfully selected the dominant alternative products, and remained consistent in their decision-making when the same choice task was presented twice with some time in between. After assessing the quality of the data, the MNL model was estimated. The results showed that Usability/Accessibility Score had a major impact on the purchasing decisions. User Rating followed as the second most influential factor. Cost had a negative impact which was comparatively small, but nonetheless significant. Brand did not achieve statistical significance and participants appeared indifferent toward this attribute. Examination of the model performance indicators revealed a remarkable model fit, as evidenced by the McFadden  $R^2$  and choice prediction success rate. In addition, the parameter estimates did not show any concerning behavior in the variance-covariance matrix. Participant burden appeared very low, based on responses to the feedback questionnaire. The time investment requirements also did not appear concerning; however, some participants had spent significantly more time on the survey than others. Potential adjustments to the survey and data collection and analysis procedures were identified before the conclusion of this phase.

## ***Phase II***

Parameter estimates that had achieved statistical significance in the previous phase were used as priors for this phase's  $D_b$ -Efficient design. Evaluation of 229 designs identified a design that offended attribute level balance the least, with a low D-error = 0.543. Parameter estimate *SEs* were subsequently simulated against increments of sample size and showed a

flattening curve for  $N$  between 150 and 200. Data collection commenced by recruiting 179 participants through Prolific. After automatic and manual inspection of the data, five participants were rejected due to duplicate submissions or failing the embedded attention checks. Examination of the sociodemographic information revealed that White participants were more represented in the sample than the national estimates of the population (PwOD aged 50 and over). However, it remained mostly in line with the proportion of different ethnicities on Prolific. There were no major discrepancies in other characteristics, and the participants appeared to represent most geographical locations in the United States.

Internal validity assessment revealed that all participants had passed the Dominant Alternative test, and 73.6% had passed the choice consistency test. The MNL model was constructed, and all parameters achieved notable statistical significance. The analysis revealed that Usability/Accessibility Score and User Ratings were the most influential factors in decision-making, with Brand preference also playing a modest yet significant role. Cost, while statistically significant, had a comparatively low impact. Based on the performance indicators, the model performed very well in fitting the data based on the McFadden  $R^2$ . However, it did not successfully predict more than 42.89% of the choices. This hinted at the violation of some MNL assumptions.

Exploratory a posteriori analysis using the LC-MNL model confirmed that multiple classes can be identified within the data, and preferences were in fact heterogenous. Comparisons of the AIC, BIC, and log-likelihood showed that a LC-MNL model with six classes performs better than models with smaller class numbers. Estimating the model yielded parameter estimates for all attributes in all six classes. In this model some classes showed

indifference towards various attributes, however, other parameters maintained a very high statistical significance. By estimating the individual-specific probabilities, each participant's classification was determined. Classes one through six had 34, 17, 12, 43, 47, and 20 members, respectively. Parameter estimates for each class showed different sensitivities to various attributes and prioritization approaches. Using the survey data, six out of the 17 categories examined showed statistically significant differences between classes: average total survey duration, familiarity level with accessibility and disability, satisfaction with the length of survey, average attribute rankings, importance of Usability/Accessibility score in decision-making, and choice consistency.

## V. Discussion

The primary objective of this study was to empirically evaluate the utility of usability and accessibility information in the purchasing decisions of people without disabilities. To my knowledge, this inquiry has never been made in the past with such breadth, depth, and scope. In this chapter, I will thoroughly interpret the results obtained from this novel study and examine their significance in light of the research questions posed. Given the pioneering nature of this research, the findings stand as initial forays into a previously unexplored domain, without the benefit of direct comparison to existing studies. However, I will place these discoveries against a backdrop of other synergetic investigations, regulations, and pressing extant market need. This unique position permits these discussions to revolve around the potential theoretical implications of the results and present a significant contribution to the nascent theoretical framework focused on the utility of usability and accessibility information for all individuals. Additionally, the practical implications of these findings will be explored, highlighting their potential applications and impact in real-world scenarios. It is also crucial to acknowledge the limitations of this study, which impose restrictions on the interpretation of results. Within the context of these constraints, this study opens numerous avenues for future research, setting a foundational precedent for subsequent future inquiries. As such, this chapter aims to provide a rigorous analysis of the study's outcomes, discuss practical and theoretical implications, contextualize the study constraints and limitations, and establish a baseline for future exploration and discourse.

## **Outcomes Interpretation**

In this section, I will examine the research outcomes in direct relation to the initial research questions, with the aim of clarifying how the findings align with, or diverge from, the anticipated outcomes. Emphasis will be placed on succinctly interpreting these results to underscore their significance and implications.

### ***Utilities and Tradeoffs***

The MNL analysis results of Phases I and II clearly demonstrate not only that the availability of Usability/Accessibility Score significantly impacts participants' decisions, but also given a high enough score, this attribute patently overshadows the impact of any other singular attribute. In fact, in Phase I, all other parameters combined cannot match the part-worth utility of Usability/Accessibility Score when it is "High (8/10)". This holds true despite assuming the statistical significance of Brand's impact. Even in Phase II all other parameter estimates combined outmatch this effect only by a small margin (2.778 against 2.740).

In both phases of the study, having a low Usability/Accessibility Score was significantly detrimental. When all other variables remained constant, participants demonstrated a stronger preference for a product with no Usability/Accessibility Score, over one with a low score. The implication is that participants are more inclined to take a chance on a product lacking a Usability/Accessibility Score, rather than choosing one known to have poor accessibility and usability. This general attitude in the choice scenarios underscored the significant utility of usable and accessible products for PwOD, and their responsiveness to the availability of Usability/Accessibility Information.

The performance of the medium level of Usability/Accessibility Score aligned with logical expectations. This level significantly improved the likelihood of a product being chosen; however, its influence was less pronounced compared to that of a high score. Considering solely the magnitude (and omitting the direction) of effect estimates, the medium Usability/Accessibility Score provided greater utility than both Cost and Brand attributes. Nonetheless, in these scenarios, User Rating assumed a more significant role in predicting choice outcomes.

The Marginal Rate of Substitution in a linearly additive main effects model (such as the one used in this study) is calculated as the ratio of two parameter estimates (Lancsar et al., 2007). It denotes the rate at which a consumer is willing to trade off one attribute in exchange for another while maintaining the same level of utility. The implication of the findings above is that based on MRS as calculated for the MNL models, participants were likely to trade off more of other attribute levels in exchange for a high Usability/Accessibility Score. This remained true for medium and low levels of this attribute compared to Cost and Brand. However, at lower levels for this attribute, User Rating was more likely to be the dominant deciding factor.

A noteworthy facet of the discussion above on the impact sign and magnitude of various Usability/Accessibility Score levels, is the congruence of the elicited reactions with logical expectations of accessibility and usability indicators. If participants had not perceived this attribute as reasonable, realistic, and reliable, variations in its levels might not have significantly influenced their decision outcomes to such degree. This suggests that the implementation of usability and accessibility indicators can profoundly influence the purchasing decisions of PwOD, and the extent and direction of this influence is directly proportional to the product's

accessibility and usability. Considering this implication, and since accessibility and usability can be assessed prior to a product's introduction to the market, the evaluation results may potentially serve as a predictor of successful reception by the market.

### ***Sociodemographic Characteristics and Familiarity with Disability***

The MNL model in Phase II presented a very good fit, based on traditional performance indicators such as  $p$ ,  $CI$ , and McFadden  $R^2$ . However, the prediction success rate (42.89%) suggested that in many cases, the model's estimations were not reflective of the observed choices. With the chance level at 25% (four alternatives in each choice task), the MNL model was outperforming this threshold by a noticeable margin. However, the violation of "preference homogeneity" assumption, as one of the cornerstones of the MNL model, was found responsible for the subpar prediction performance. The ensuing LC-MNL analysis identified the utility model for different classes of participants.

Based on the individual-specific probabilities, in conjunction with the survey data, I profiled these classes based on their distinguishing features. While age, gender, and income were not statistically significant in class differentiation, the role of familiarity with disability and accessibility concepts emerged as a crucial factor. The profiles below will help put the utility models of different classes against a backdrop of their characteristics. For brevity, I will use "Rating" for User Rating and UAS for Usability/Accessibility Score in the profile labels below.

**Class 1 Profile:** *Primarily UAS-Responsive and Cost-Averse, Less Brand-Selective.* Class 1 was predominantly characterized by its substantial responsiveness to Usability/Accessibility Score, particularly at medium and high levels. This class also exhibited a notable aversion to higher costs, demonstrating significant sensitivity in this regard. This class also placed lesser

emphasis on Brand compared to others. Comprising about 20% of the participants, they generally expressed the least familiarity with disability and accessibility concepts but ranked Usability/Accessibility Score as the second most important attribute by a small margin. Despite spending the most time on average on the survey, the participants in this class rated their satisfaction with the survey length as one of the highest. The average age in this class was 57.5 years ( $SD = 7.13$ ), with a nearly equal distribution of male and female members, and their income brackets were predominantly normally distributed.

**Class 2 Profile:** *UAS-Driven, Cost-Indifferent*. The defining characteristic of Class 2 was its exceptional responsiveness to the Usability/Accessibility Score. Both high and medium scores for this attribute outranked the effects of other class-specific parameter estimates. Additionally, this class exhibited minimal sensitivity towards Cost, with price fluctuations having an insignificant impact on its preferences. Comprising approximately 10% of participants, Class 2 ranked the Usability/Accessibility Score as the most important attribute above all others. They also often rated this factor as “Extremely Important” in their decisions. Additionally, this class performed well in the Stability Validity test, marking them as one of the most consistent classes in decision-making. The average age of Class 2 members was 60.8 years ( $SD = 6.39$ ), with a near-equal distribution of male and female participants. Their incomes were primarily concentrated in the \$25K to \$50K and \$100K to \$150K ranges.

**Class 3 Profile:** *Cost-Driven, Least UAS-Responsive*. Class 3 was notably the most sensitive to Cost among its counterparts. Another distinguishing attribute of this class was its minimal valuation of the Usability/Accessibility Score, assigning the least utility to this parameter comparatively. This class constituted about 7% of the participants, who spent the

least amount of time on the survey and had the highest failure rate in the Stability Validity test. They consistently ranked Cost as their top priority, more than any other class. The average age in this class was 59.0 years ( $SD = 6.95$ ), with an equal distribution of male and female members. Income brackets for this class were predominantly \$75K and above, with a smaller proportion ranging from \$25K to \$50K.

**Class 4 Profile:** *Primarily Brand-Selective and Rating-Sensitive, Less Cost-Averse.* Class 4 was more motivated by Brand and User Rating compared to other classes. Additionally, the impact of Cost on the decision-making process of this class was considerably low compared to others. With approximately 25% of participants, this class reported a moderate familiarity with disability and accessibility concepts, having the lowest proportion of individuals who were “not familiar at all” with these topics. Additionally, it also had the highest number of members extremely dissatisfied with the survey length. User Rating was frequently ranked first in this class, more so than in most other classes. Their valuation of the Usability/Accessibility Score was typically moderate or higher, with the fewest “not important at all” ratings among all classes. The average age of participants in Class 4 was 59.0 years ( $SD = 7.22$ ), with a majority being female. The income bracket of \$150K and above was most represented in this class, while the distribution in other brackets was almost normal.

**Class 5 Profile:** *Rating-Driven, Less Brand-Selective or UAS-Responsive.* Class 5 primarily made choices influenced by User Rating, exhibiting the most significant impact from this factor among all classes. The influence of User Rating superseded all other factors in this class comfortably. Additionally, this class assigned noticeably less utility to Brand and was less influenced by the Usability/Accessibility Score compared to others. Comprising the largest

group at approximately 27% of participants, they most frequently prioritized User Rating over other attributes. Echoing Class 1, Class 5 spent the second-highest amount of time on the survey and expressed the highest satisfaction with its length. The average age in this class was 57.9 years ( $SD = 6.38$ ), with a nearly equal number of male and female participants. Income distribution was mostly normal across all ranges, with a higher representation in the above \$100K categories.

**Class 6 Profile:** *Brand-Driven, Less UAS-Responsive.* The predominant characteristic of Class 6 was its strong influence from Brand, assigning the highest utility to this attribute compared to other classes. Additionally, this class assigned lower utility to the Usability/Accessibility Score compared to its counterparts. Making up about 11% of the total sample, members of this class were the most consistent in their choices, as indicated by the Stability Validity test. Brand was most frequently ranked as this class's top priority. Participants in this class also reported the highest frequency of being "not familiar at all" with disability and accessibility concepts and were most likely to rate the Usability/Accessibility Score as slightly important or less in their decisions. The average age of Class 6 members was 60.3 years ( $SD = 7.00$ ), with a nearly equal distribution of male and female participants. Their income brackets were mostly skewed towards the lower ranges, with fewer members above \$100K and more in the below \$25K range than other classes.

It must be emphasized that while in the profile labels above I use verbiage such as "*Less or Least UAS-Responsive,*" these characterizations are being made in comparison with other classes and their utility assignments. They are not intended to describe the magnitude and significance of parameters within classes. The LC-MNL model parameters that achieved

statistical significance impact decisions in the same direction explained by the MNL model.

However, the magnitude and consequently the MRS are different for various classes. The low level for Usability/Accessibility Score still negatively impacts the choice outcomes in all classes where it is statistically significant, and the high and medium levels still have a positive impact, with the magnitude of impact being larger for the high level.

### ***Usability/Accessibility Score - Perceptions vs. Observation***

One standout discovery of studying the unique characteristics of subgroups, was the propensity of many classes to place User Ratings in the highest rank when explicitly asked, while their class-specific utility model suggests they valued certain levels of Usability/Accessibility Score either just as much, or higher than User Rating. For example, Class 4 ranked User Rating the second highest on average, yet the MRS for both high and low levels of Usability/Accessibility Score compared to User Rating remain close to one in this class (Class4.MRS<sub>UAS-Low, UR</sub> = 0.942 and Class4.MRS<sub>UAS-High, UR</sub> = 0.827). Class 1 is another example of this scenario (Class1.MRS<sub>UAS-Med, UR</sub> = 1.432 and Class1.MRS<sub>UAS-High, UR</sub> = 2.024). Class 3 also showed the same behavior only for the high level of Usability/Accessibility Score (Class3.MRS<sub>UAS-High, UR</sub> = 1.446). The MNL model for both phases, in comparison to the distribution of their rankings for different variables overall (see Figs. 15 and 20), confirms this overall tendency to rank User Rating very high, while Usability/Accessibility Score empirically played a more notable role than perceived.

In reviewing the data collected during Phase II, one particular response stood out, offering a unique perspective on the earlier discussion about the contrast between User Ratings and Usability/Accessibility Scores. The participant commented:

Not sure if this helps, but if the user rating was high, I had a hard time accepting that the usability/accessibility rating could be really low, so when a user rating was at 4.8 and the usability/accessibility rating was 4/10, I discounted the latter completely.

While it is clear why individuals may be inclined to place significant value on User Rating, positive reviews are not necessarily always reflective of what a product cannot do and are likely left by individuals who anticipated being able to use the product before the purchase. The common perception may remain that the products with high user reviews, indicative of their utility for many others, must fit the needs of all individuals. However, it appears that the premise of this statement is founded on hasty generalization. I believe no population may reject this sentiment quite as emphatically as PwD. Even highly rated products can pose significant usability barriers for many of these individuals. Participants may have consciously perceived their utility assignment as skewed towards User Rating, however, the empirical observations evidence prioritization of Usability/Accessibility Score in many scenarios.

Another participant noted, “Negative reviews are just as crucial as positive ones, often revealing significant insights.” This comment underscores the value of considering feedback from users who encounter problems, whether with the device itself or the purchasing process. The benefit of an objective evaluation of accessibility and usability lies in highlighting potential barriers (with potentially life-threatening consequences in the case of medical devices) before the users face them. For example, a task analysis of three widely-used BPMs, aimed at assessing adherence to universal design principles, revealed that all violated principles of equitable use, simplicity, perceptibility, and error tolerance (Cifter, 2017). Among the seven universal design

principles (Connell et al., 1997), the only unviolated one was size and space for approach and use, which is possibly irrelevant to BPMs. Another study in 2019 reported that a popular blood pressure monitor had challenging instructions, display data, and application procedures, potentially leading to user errors and dissatisfaction (Cote et al., 2019).

Cifter (2017) detailed key features of different BPMs aimed at enhancing user accessibility and ease of use. For BPM 1, features included multilingual voice output, button backlighting, and color-coded blood pressure scales. BPM 2 offered an intuitive positioning system with visual cues, audio signals, and color-coded pressure indicators. For BPM 3, a cuff-wrapping guide and blood pressure level indicator were highlighted. These features were specifically designed to accommodate a wide range of user needs and abilities (Cifter, 2017). Based on the findings of my research, and the conclusions of the above studies, it seems evident that consumers will be able to make better informed decisions, with lower possibility for dissatisfaction or health complications, if objective information on the positive and negative aspects of device accessibility and usability are made available to them prior to purchase.

Contrasting the common attitudes towards user ratings with an objective report of a product's capabilities and accommodations, presented a unique opportunity for discovery in this study. The comments mentioned previously were the first indications that there may be distinct segments within the study cohort, each with unique and distinctive patterns of utility assignment. This hypothesis was substantiated by the LC-MNL analysis. Notably, the two individuals whose comments were referenced earlier were identified as belonging to Classes 6 and 5, respectively. These classes demonstrated a relative insensitivity to Usability/Accessibility Scores, showing a stronger preference for factors like Brand and Cost.

## Contextualization of Discoveries

The field of usability and user experience evaluation with medical devices has seen an exponential growth over the past two decades (Bitkina et al., 2020). With the proliferation of home medical devices worldwide, enhancing the usability of these devices has grown increasingly important due to the potentially life-threatening consequences. However, studies have uncovered skepticism from industry towards the implementation of practices that ensure usability of home medical devices, citing reasons mostly related to cost of implementation (Grant, 2014; Privitera et al., 2017).

People with Disabilities are disproportionately impacted by the lack of accessibility and usability in medical devices, facilities, and services (Coyle & Santiago, 2002; Grabois et al., 1999; Kailes, 2007). Their health outcomes show great disparities in comparison to PwOD, primarily to inaccessible medical equipment, and exacerbated by the unavailability of any means to anticipate and prepare for usability barriers (Mendonca, 2010; Sade, 2012; Schopp et al., 2002; R. O. Smith et al., 2007; Story et al., 2010). Consequently, hundreds of millions of dollars are spent on medical devices and equipment that fail to realize their purpose due to the access barriers (Kailes et al., 2006; Nosek, 2000; Nosek & Howland, 1997). In 2021, I led a survey to examine the perspectives of approximately 200 PwD, on the accessibility of medical devices and equipment they use, either at home or healthcare facilities. The findings overwhelmingly supported the need for medical device usability and accessibility information to facilitate informed decision-making when purchasing medical devices or visiting healthcare providers (Ardehali et al., 2023).

Valid and reliable instruments that evaluate usability and accessibility have the potential to (a) provide design and evaluation guidelines to the medical device industry stakeholders, and (b) facilitate informed decision-making by all consumers, with or without disability. This study empirically demonstrated that PwD assign significant utility to accessibility and usability information when purchasing medical devices. In fact, with some differences, the part-worth utilities for various levels of Usability/Accessibility Score generally followed the utility PwD assigned to these information (Mendonca, 2010). A comparison of these estimates is presented in Table 15, to demonstrate the extent to which PwD and PwoD similarly value usability and accessibility informing their medical device purchasing decisions.

## Implications

The adoption of accessible and universal design models in the medical device industry hinges on two pivotal factors: regulatory frameworks and profitability considerations (Vanderheiden & Tobias, 1998, 2000). Even the implementation of procedures to enhance user

**Table 15:**

*Usability and Accessibility Score Parameter Estimate Comparison Between PwD and PwoD*

Parameter	PwD – BPM <sup>1</sup>		PwD – Exam Table <sup>2</sup>		Parameter	PwoD – BPM	
	Estimate	<i>p</i>	Estimate	<i>p</i>		Estimate	<i>p</i>
Accessibility Score					Usability/Accessibility Score		
Unavailable	-	-	-	-	Unavailable	-	-
Low	0.75	<0.0001	0.06	0.63	Low	-0.855	<.001
Medium	2.29	<0.0001	1.72	<0.0001	Medium	1.219	<.001
High	3.46	<0.0001	2.98	<0.0001	High	2.740	<.001

<sup>1</sup> Mendonca (2010) – *N* = 98, Table 3

<sup>2</sup> Mendonca (2010) – *N* = 63, Table 8

experience for PwOD is hindered by cost, internal policies, and access to usable and reliable user feedback (Grant, 2014; Privitera et al., 2017). Especially with regards to accessible design, industries view accessible design as catering to a niche audience, primarily PwD, which could lead to higher design and manufacturing costs without reasonable returns (Shinohara et al., 2018). As such, medical device designers and manufacturers often rely on their experience, while minimally involving users in the final stages (Grant, 2014). Since PwOD constitute a larger market segment, manufacturers are driven towards profit maximization by targeting this audience, overlooking the significant needs of PwD in the process. Consequently, medical devices are contributing to health outcomes disparities between PwOD and PwD.

Many aspects of design for PwD are directly relevant to general user experience outside the context of disability. Evaluating the accessibility of medical devices, using a reliable and valid instrument, can yield results that (a) inform design accessibility improvements for PwD; (b) inform design usability improvements for PwOD, and (c) provide accessibility and usability indicators that can be used by PwD and PwOD to facilitate informed purchasing decisions. Additionally, these evaluations are easier to adopt and less costly to implement, and their results are directly translatable into design guidelines.

As discussed previously in the first chapter, MED-AUDIT is a notable example of an instrument for the evaluation of medical device usability and accessibility, specifically tailored to the needs of PwD. I conducted an experiment to demonstrate the pertinence and translatability of accessibility evaluations to general user experience, regardless of the level of ability. This experiment was limited in scope, and further studies may be needed to validate the findings. Nevertheless, I asked a friend without disability to review the items on the MED-AUDIT

taxonomy and determine whether information on those items would benefit PwD or PwoD, and exclusively or jointly. This individual was mostly unfamiliar with disability and concepts relating to accessibility, and I only provided basic instructions for how to review the items and determine to whom they may be pertinent. I also reviewed the taxonomy myself to determine the relevance of items.

In addition, I deployed the same state-of-the-art technique used previously in the fraud detection module and introduced an Artificial Intelligence rater to the cohort. Via the OpenAI API implementation in Python, this round I used the “gpt-4-1106-preview” language model which is the most advanced engine currently available. The model was tuned to behave as a layperson with no particular expertise, in the context of evaluating medical devices, and report whether information on each taxonomy item would be beneficial to PwD, PwoD, or both.

This taxonomy consists of 1,149 items and is designed specifically to evaluate the accessibility of medical products for PwD. The results indicated that at the minimum, 88.91% of the items provide information pertinent to both PwD and PwoD, and at the maximum 11.09% are exclusively relevant to PwD. The inter-rater reliability was measured using Gwet’s AC1 (Gwet, 2008; Wongpakaran et al., 2013), and indicated a very high level of agreement between raters ( $AC1 = 0.90, p < .001$ ). This experiment, though limited in scope, showed medical device evaluation results for PwD are also highly translatable to usability results for PwoD.

In this dissertation, the findings empirically demonstrate that large segments of PwoD substantially value accessibility and usability information. The implications of these findings may facilitate a paradigm shift in the medical device industry to evaluate medical devices for usability and accessibility. Additionally, the results suggest that the presentation of usability

and accessibility indicators informs purchasing decisions, and most PwD find this additional significantly impacting their decisions, influencing marketing strategies. By alleviating the industry's financial reservations and providing evidence to redefine market motivations, the consequences of this paradigm shift may aid in the reduction of healthcare disparities between PwD and PwoD through promotion of informed medical device purchasing decisions.

## **Limitations**

### ***Experiment Design***

Due to the novelty of the central hypothesis in this study, some limitations were imposed intentionally. The intent was to perform the most pragmatic observation possible, without sacrificing realism and the validity of conclusions, and lay the foundation for future efforts and less reductionistic approaches. In the DCE design stage, all facets were established with objectivity in mind. This resulted in the omission of attributes that are commonly present but highly subjective. The inherent subjectivity of such attributes would deter from reliably estimating the parameters of interest and determining their significance. For example, including product images, though realistic, might have resulted in the rejection of some products solely based on the color, shape, or other aesthetic features, regardless of the interplay between other important attributes. Regardless of these limitations, I attempted to design the study with the highest level of realism possible, by including the most significant and omnipresent attributes and identifying their levels directly informed by market research.

Another limitation of the experiment design in this study is its reliance on a stated preference method. While DCEs offer significant advantages, especially in scenarios where true observations are not practical, they also raise external validity concerns due to their

hypothetical nature. Participants might indicate a preference for a certain choice in the experiment, but real-life decisions could be swayed by other factors. This discrepancy between intention and behavior is present in all stated preference methodologies (Ajzen, 1991; Szinay et al., 2021). Verifying the external validity of these methodologies remains a difficult task, as it involves assessing preferences by realizing the hypothetical choice scenarios.

### ***On Blue Buses, Red Buses, and Homogeneity of Preferences***

The analytical model appears to be another facet that in retrospect, could have been substituted for other models with less restrictive assumptions. The justification for using the MNL model was discussed extensively under Analytical Model. These restrictions of the MNL model only manifested after the conclusion of data collection in Phase II, and this limitation is only identified retrospectively. Despite these restrictions the model still performed very robustly, based on the significance of attributes and McFadden  $R^2$ .

The title of this segment is acknowledging a famous paradox that challenges the IIA criterion in choice theory, which is another foundational assumption of the MNL model. The "red bus/blue bus" paradox illustrates a scenario where an agent who chooses to travel by car instead of red bus, may be irrationally influenced by the introduction of an otherwise identical blue bus to the choice task. According to IIA, the probability of choosing between the first two options should remain unchanged regardless of the addition of the "irrelevant" alternative. However, this paradox highlights that in real-world scenarios, people's preferences can be inconsistently altered due to seemingly irrelevant factors.

The a posteriori analysis using LC-MNL allowed for the identification of different study cohort segments. The LC-MNL model also utilizes the MNL model to estimate the parameters.

However, it does so within the confines of each class where the assumption of homogeneity and IIA is, by design, more likely to hold true.

### ***Scope***

The results of this study pertain to the population characteristics, specific product, and attribute levels used in hypothesis evaluation. In other words, the findings are explaining the impact of Usability/Accessibility Score for English-speaking PwD above 50 who live in the United States, when purchasing BPMs in the presence of Cost, Brand, User Rating, and Usability/Accessibility Score attributes. Generalization beyond these boundaries requires establishing further evidence. Nevertheless, these results may denote the likelihood of success, if future attempts are made to reasonably expand the scope of investigation beyond what was examined in this study.

### **Future Directions**

#### ***Multifaceted Accessibility and Usability Valuation Effect***

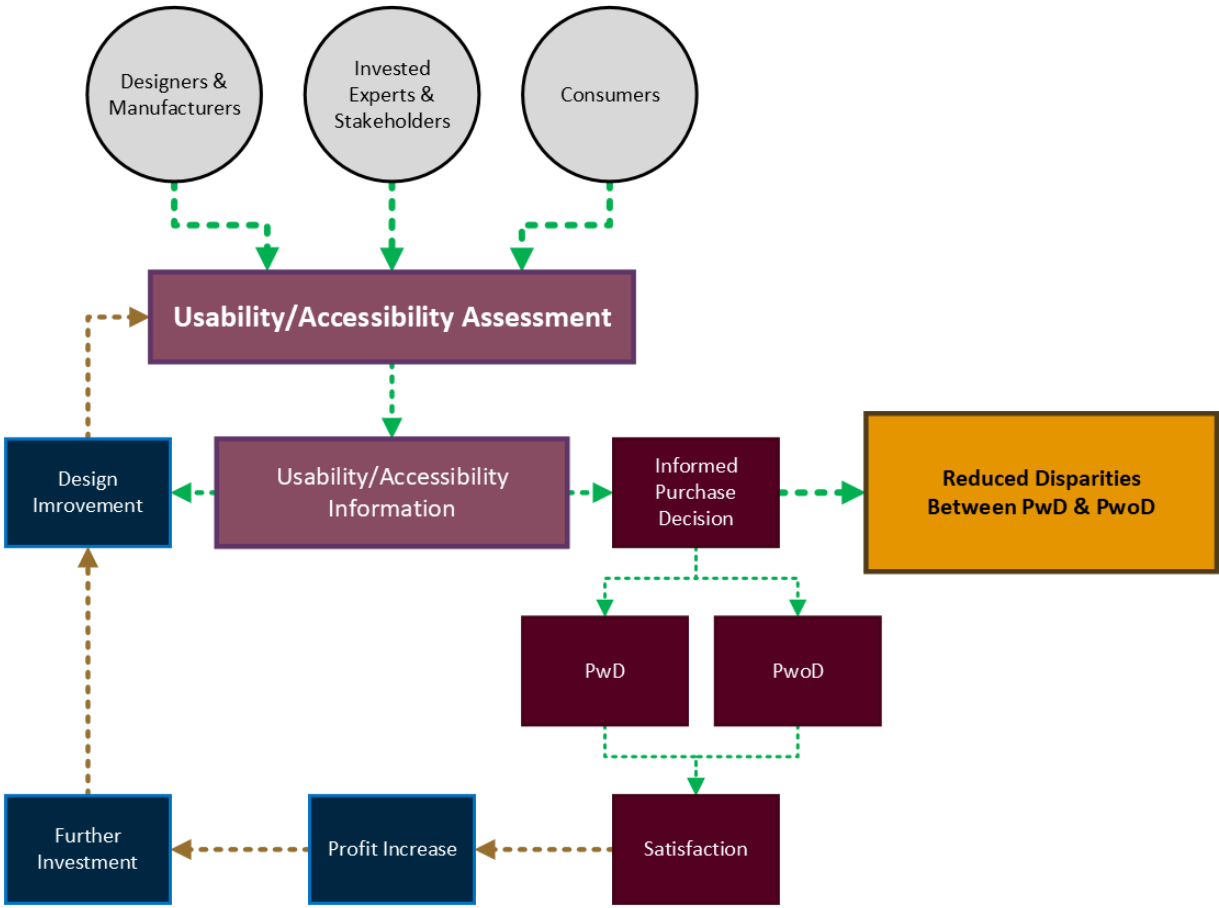
Previous scholarly efforts have evidenced the significant value PwD place on accessibility information when making purchasing decisions. This finding has been complemented in this study by a parallel discovery that PwD also exhibit a notable appreciation for accessibility and usability information in their decision-making processes. Based on the significant evidence uncovered in this study, and my observations over the years, I formulated the "Multifaceted Accessibility and Usability Valuation Effect" (MAUVE) hypothesis. The MAUVE hypothesis posits that the evaluation of accessibility and usability transcends the realm of disability, becoming a universal criterion in consumer decision-making. This perspective not only emphasizes the convergence of the preferences among PwD and PwOD with regards to Usability/Accessibility

Information, but also underscores the significance of this factor in the broader context of market dynamics.

The implications of the MAUVE hypothesis are threefold. Firstly, accessibility and usability assessments provide designers and manufacturers with invaluable insights into how their products, services, or environments can be enhanced for all. This perspective encourages a universal design approach, which iteratively and progressively elevates the quality of designs to benefit a wider range of users. Secondly, the generation of detailed usability and accessibility information serves as a vital resource for both PwD and PwOD, aiding them in making more informed decisions. Thirdly, usability and accessibility assessments need not be exclusively led by designers and manufacturers. Other invested stakeholders, including consumers, can participate in populating rating and evaluation databases, thereby enriching the landscape of insights. This increased information symmetry empowers consumers in general, and particularly ameliorates the disparities between PwD and PwOD. The resultant effect is a positive feedback loop where accessibility and usability assessment data navigate consumer purchasing decisions, which in turn motivates designers and manufacturers to further invest in evaluating and enhancing these aspects and achieve a larger market share, ultimately benefiting all stakeholders involved (Figure 24).

While the initial findings supporting the MAUVE hypothesis are promising, more research is necessary to evidence the missing connections and ascertain external validity. Future studies should aim to expand the body of evidence on the impact of accessibility and usability information on the decision-making processes of both PwD and PwOD across various

**Figure 24:**  
*The MAUVE Hypothesis*



contexts, including different types of products, services, and environments. Such future endeavors not only would solidify the foundations of the MAUVE hypothesis, but also provide actionable insights for stakeholders aiming to reduce disparities and foster a user-friendly market landscape. Further exploration of this hypothesis in various settings is crucial for understanding the extent of its applicability and for developing ability-agnostic strategies that cater to the needs of all consumers.

## VI. Conclusion

Significant disparities exist in health outcomes, between People with and without Disabilities, due mainly to the inaccessibility of medical devices and equipment. Despite federal regulatory guidelines and a substantial body of evidence emphasizing the necessity for improved medical device accessibility and usability, there appears to be a marked reluctance in designers and manufacturers to invest in improving the accessibility of their medical products. This resistance extends not only to the actual enhancement of product accessibility, even at a marginal level, but also to the evaluation and marketing of these products with adequate accessibility information. At the minimum, this information could aid consumers, especially PwD, in making informed purchasing decisions, thereby decreasing the health outcomes disparities.

The factors predominantly driving this hesitation are cost of implementation and perceptions of the market size that may potentially benefit from these enhancements. Since PwD comprise only a fraction of the general population, medical device manufacturers pursue profit maximization by prioritizing PwoD. Reinforcing market motivation for medical device manufacturers appears key to facilitate the provision of accessibility and usability information and increase the accessibility of medical devices.

I hypothesized that this boost to market motivation can be delivered by demonstrating that PwoD purchasing decisions may also depend on the usability and accessibility of medical devices. Thus, this study investigated the impact of usability and accessibility information on the medical device purchasing decisions of People without Disabilities. It examined the magnitude and direction of impact for Usability/Accessibility Score, as a representation of

usability and accessibility information, in conjunction with Cost, Brand, and User Rating attributes. Due to their widespread use, which increased the likelihood of task realism, blood pressure monitors were used as hypothetical medical devices in the experiment design.

In three key stages, this study discovered that PwoD value usability and accessibility information highly. The first stage established the significant impact of Usability/Accessibility Score with a smaller sample size. In the second stage, the sample size was increased by a factor of ten, and the impact of this factor was found significant and followed the same trend. The third stage systematically segmented the study cohort into smaller partitions in which participants were more likely to share the same values when assigning utility to different attributes. This a posteriori investigation demonstrated that in most classes, higher levels of Usability/Accessibility Score significantly increased the likelihood of purchase, lower levels significantly harmed this likelihood, and medium levels had a significant positive influence, though not as pronounced as the impact of higher levels. Subsequently the profiles of different classes were determined based on the information they provided in the study, including sociodemographic data. The findings strongly confirmed the central hypothesis of this dissertation by underscoring the value PwoD derive from usability and accessibility information.

The findings of this study hold the potential to instigate a significant paradigm shift by advocating for increased reporting of usability and accessibility information by key stakeholders. Such a shift is poised to enhance informed decision-making across the board, thereby contributing to a more efficient and effective healthcare ecosystem. Moreover, this emphasis on accessibility and usability stands to substantially mitigate the existing disparities in healthcare outcomes faced by People with Disabilities. Highlighting the importance of usability

and accessibility information in the healthcare sector by implication, this change could ultimately lead to more equitable health services and improved quality of life for individuals with disabilities.

The inherent and imposed limitations of this study may restrict the generalizability of some findings. The inclusion of factors that were least likely to elicit unobservable subjective variations was prioritized. While the attributes and levels included in this study were thoroughly researched to ensure realism in choice scenarios and maximize generalizability, the aforementioned prioritization may have excluded some factors that are otherwise commonly present on e-tailers. Another unforeseen limitation arose from the choice of analytical model. While the MNL model proved effective in estimating the parameters, demonstrating significance and a very good fit, it did not excel at outcome predictions due to the violation of its inherently restrictive assumptions. However, this impact was mitigated through the a posteriori analysis. Due to the ongoing uncertainty about observing any effect, throughout the entire course of this study, the scope was also restricted prior to experiment design to enhance the probability of detecting smaller effect sizes while maintaining the realism of choice scenarios.

### **Personal Reflection**

I believe the findings of this study can advocate for a significant improvement in the health outcomes disparities between PwD and PwoD by incentivizing medical device manufacturers to account for usability and accessibility of their products. The evidence for this optimism is the significant and sizable impact of different levels of Usability/Accessibility Score

on the choices. Even in segments of the study cohort that predominantly prioritized other attributes, this impact was still noticeable, significant, and in the expected direction.

Tangential to the primary objective, based on the evidence unveiled by this study, I posit that an entirely absent facet of product representation may have been discovered. In fact, I hypothesize that the utility derived from user ratings is only the subjective manifestation and a function of usability and accessibility. Of course, this experiment was but a small step in the direction of identifying this relationship. However, circumstantially and anecdotally, it appears to me that the reason for the high utility of user ratings is precisely that they provide the accounts of individuals, presumably similar to the consumer, on the experiences they have had with the usability and accessibility of products. Future investigations may help substantiate this intuition.

## VII. References

- Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53–56, 63–70. <https://doi.org/10.1016/j.jom.2017.06.001>
- Acquah, H. de-Graft. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2(1), 001–006.
- ADA.gov. (2020, June 26). *Access To Medical Care For Individuals With Mobility Disabilities*. [https://www.ada.gov/medcare\\_mobility\\_ta/medcare\\_ta.htm](https://www.ada.gov/medcare_mobility_ta/medcare_ta.htm)
- Adamowicz, W., Bunch, D., Cameron, T. A., Dellaert, B. G. C., Hanneman, M., Keane, M., Louviere, J., Meyer, R., Steenburgh, T., & Swait, J. (2008). Behavioral frontiers in choice modeling. *Marketing Letters*, 19(3), 215–228. <https://doi.org/10.1007/s11002-008-9038-1>
- Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. *JACCP: JOURNAL OF THE AMERICAN COLLEGE OF CLINICAL PHARMACY*, 4(10), 1358–1367. <https://doi.org/10.1002/jac5.1441>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ardehali, M. M. (2023). *Examining Stated Choice Preferences in Purchasing Blood Pressure Monitors*. <https://doi.org/10.17605/OSF.IO/YT4GP>
- Ardehali, M. M., Mendonca, R. J., Smith, R. O., Mendonca, R., & Smith, R. O. (2023). People With Disabilities & Medical Device Accessibility: Perspectives, Challenges, and the Role

- of Occupational Therapy Practitioners. *The American Journal of Occupational Therapy*, 77(Supplement\_2), 7711500014p1. <https://doi.org/10.5014/ajot.2023.77S2-PO14>
- Arrow, K. J. (2012). *Social Choice and Individual Values*. Yale University Press.  
<https://www.jstor.org/stable/j.ctt1nqb90>
- Baji, P., Pavlova, M., Gulácsi, L., & Groot, W. (2012). Preferences of Hungarian consumers for quality, access and price attributes of health care services—Result of a discrete choice experiment. *Society and Economy*, 34(2), 293–311.  
<https://doi.org/10.1556/socec.34.2012.2.7>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, 24(6), 574–594.  
<https://doi.org/10.1080/10447310802205776>
- Barnett, S., & du Toit, N. (2018, December 8). Disability is a spectrum, not a binary. 24 *Accessibility*. <https://www.24a11y.com/2018/disability-is-a-spectrum-not-a-binary/>
- Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., & Loomes, G. (2002). *Economic Valuation with Stated Preference Techniques*. Edward Elgar Publishing. <https://econpapers.repec.org/bookchap/elgebook/2639.htm>
- Becker, H., Stuijberger, A., & Tinkle, M. (1997). Reproductive health care experiences of women with physical disabilities: A qualitative study. *Archives of Physical Medicine and Rehabilitation*, 78(12 Suppl 5), e26–e33. [https://doi.org/10.1016/s0003-9993\(97\)90218-5](https://doi.org/10.1016/s0003-9993(97)90218-5)
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 183–200. <https://doi.org/10.1037/h0024835>

- Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.
- Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2014). Can Marginal Rates of Substitution Be Inferred From Happiness Data? Evidence from Residency Choices. *The American Economic Review*, *104*(11), 3498–3528.  
<https://doi.org/10.1257/aer.104.11.3498>
- Berbeglia, G., Garassino, A., & Vulcano, G. (2018). *A Comparative Empirical Study of Discrete Choice Models in Retail Operations* (SSRN Scholarly Paper 3136816).  
<https://doi.org/10.2139/ssrn.3136816>
- Billot, A., & Thisse, J.-F. (1999). A Discrete Choice Model When Context Matters. *Journal of Mathematical Psychology*, *43*(4), 518–538. <https://doi.org/10.1006/jmps.1998.1237>
- Bitkina, O. V., Kim, H. K., & Park, J. (2020). Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. *International Journal of Industrial Ergonomics*, *76*, 102932.  
<https://doi.org/10.1016/j.ergon.2020.102932>
- Bliemer, M. C. J., & Rose, J. M. (2014). Designing and Conducting Stated Choice Experiments. In *Handbook of Choice Modelling* (2nd ed.). Edward Elgar Publishing.  
<https://surveyengine.com/wp-content/uploads/2021/11/Designing-and-Conducting-Stated-Choice-Experiments.pdf>
- Böckenholt, U. (2006). Thurstonian-Based Analyses: Past, Present, and Future Utilities. *Psychometrika*, *71*(4), 615–629. <https://doi.org/10.1007/s11336-006-1598-5>

- Bonoli Escobar, M., Volpe, J. I., Mosca, J., & Picasso, E. (2013). Respondent engagement and task complexity effects in stated choice experiments. *Coloquio Argentino de Estadística (41° : 2013 : Mendoza, Argentina)*.  
<https://repositorio.uca.edu.ar/handle/123456789/5612>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Breckenridge, C. A., & Vogler, C. (2001). The Critical Limits of Embodiment: Disability's Criticism. *Public Culture, 13*(3), 349–358. <https://doi.org/10.1215/08992363-13-3-349>
- 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>
- Brockman, G., Murati, M., Welinder, P., & OpenAI. (2020, June 11). *OpenAI API*.  
<https://openai.com/blog/openai-api#OpenAI>
- Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind., 189*.
- Campbell, D., & Erdem, S. (2019). Including Opt-Out Options in Discrete Choice Experiments: Issues to Consider. *The Patient - Patient-Centered Outcomes Research, 12*(1), 1–14.  
<https://doi.org/10.1007/s40271-018-0324-6>
- Carlsson, F. (2011). Non-Market Valuation: Stated Preference Methods. In *The Oxford Handbook of the Economics of Food Consumption and Policy*. Oxford University Press.

- Caussade, S., Ortúzar, J. de D., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological*, 39(7), 621–640. <https://doi.org/10.1016/j.trb.2004.07.006>
- CDC. (2023, July 6). *Facts About Hypertension*. Centers for Disease Control and Prevention. <https://www.cdc.gov/bloodpressure/facts.htm>
- Center for Universal Design. (2002). *Universal Design: Product Evaluation Countdown*.
- Cheng, E., Myers, L., Wolf, S., Shatin, D., Cui, X., Ellison, G., Belin, T., & Vickrey, B. (2001). Mobility impairments and use of preventive services in women with multiple sclerosis: Observational studies. *British Medical Journal*, 323(7319), 968–969. <https://doi.org/10.1136/bmj.323.7319.968>
- Cheung, K. L., Wijnen, B. F. M., Hollin, I. L., Janssen, E. M., Bridges, J. F., Evers, S. M. A. A., & Hiligsmann, M. (2016). Using Best–Worst Scaling to Investigate Preferences in Health Care. *Pharmacoeconomics*, 34(12), 1195–1209. <https://doi.org/10.1007/s40273-016-0429-5>
- Christofides, N. J., Muirhead, D., Jewkes, R. K., Penn-Kekana, L., & Conco, D. N. (2006). Women’s experiences of and preferences for services after rape in South Africa: Interview study. *BMJ (Clinical Research Ed.)*, 332(7535), 209–213. <https://doi.org/10.1136/bmj.38664.482060.55>
- Cifter, A. S. (2017). Blood Pressure Monitor Usability Problems Detected Through Human Factors Evaluation. *Ergonomics in Design*, 25(3), 11–19. <https://doi.org/10.1177/1064804617691397>

- 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. <https://doi.org/10.1007/s40273-014-0170-x>
- Clarkson, P. J., Coleman, R., Keates, S., & Lebbon, C. (2013). *Inclusive Design: Design for the Whole Population*. Springer Science & Business Media.
- Coast, J., & Horrocks, S. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of Health Services Research & Policy*, 12(1), 25–30.
- Coleman, R., Clarkson, J., & Cassim, J. (2016). *Design for Inclusivity: A Practical Guide to Accessible, Innovative and User-Centred Design*. CRC Press.
- Connell, B. R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., Sanford, J., Steinfeld, E., Story, M., & Vanderheiden, G. (1997, April 1). *The Principles of Universal Design*. <https://design.ncsu.edu/research/center-for-universal-design/>
- Cote, V., Rao, A. H., & Sasangohar, F. (2019). Usability Study of the iProven BPM-337BT Blood Pressure Monitor and App. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), 1910–1914. <https://doi.org/10.1177/1071181319631311>
- Coyle, C. P., & Santiago, M. C. (2002). Healthcare utilization among women with physical disabilities. *Medscape General Medicine*, 4(3), 2.
- Croissant, C. (2021, February 16). *Why participants get banned*. <https://www.prolific.com/blog/why-participants-get-banned>
- de Bekker-Grob, E., Ryan, M., & Gerard, K. (2012). Discrete Choice Experiments in Health Economics: A Review of the Literature. *Health Economics*, 21, 145–172. <https://doi.org/10.1002/hec.1697>

- de Bekker-Grob, E. W., Donkers, B., Bliemer, M. C. J., Veldwijk, J., & Swait, J. D. (2020). Can healthcare choice be predicted using stated preference data? *Social Science & Medicine*, 246, 112736. <https://doi.org/10.1016/j.socscimed.2019.112736>
- de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample Size Requirements for Discrete-Choice Experiments in Healthcare: A Practical Guide. *The Patient - Patient-Centered Outcomes Research*, 8(5), 373–384. <https://doi.org/10.1007/s40271-015-0118-z>
- de Bekker-Grob, E. W., Swait, J. D., Kassahun, H. T., Bliemer, M. C. J., Jonker, M. F., Veldwijk, J., Cong, K., Rose, J. M., & Donkers, B. (2019). Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. *Value in Health*, 22(9), 1050–1062. <https://doi.org/10.1016/j.jval.2019.04.1924>
- de Corte, K., Cairns, J., & Grieve, R. (2021). Stated versus revealed preferences: An approach to reduce bias. *Health Economics*, 30(5), 1095–1123. <https://doi.org/10.1002/hec.4246>
- De Schauwer, E., Daelman, S., Vandebussche, H., Sergeant, S., Van de Putte, I., & Davies, B. (2021). Desiring and critiquing humanity/ability/personhood: Disrupting the ability/disability binary. *Disability & Society*, 36(2), 286–305. <https://doi.org/10.1080/09687599.2020.1735306>
- DeJonckheere, M., & Vaughn, L. M. (2019). Semistructured interviewing in primary care research: A balance of relationship and rigour. *Family Medicine and Community Health*, 7(2), e000057. <https://doi.org/10.1136/fmch-2018-000057>

- Deng, X., Yang, Y., & Jiang, W. (2023). Discrete choice models with Atanassov-type intuitionistic fuzzy membership degrees. *Information Sciences*, 622, 46–67.  
<https://doi.org/10.1016/j.ins.2022.11.127>
- DeShazo, J. R., & Fermo, G. (2002). Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. *Journal of Environmental Economics and Management*, 44(1), 123–143. <https://doi.org/10.1006/jeem.2001.1199>
- Domencich, T. A., & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis : a Charles River Associates Research Study*. North-Holland Publishing Company.
- Donahoe, G. F. (2021, June 2). *Estimates of Medical Device Spending in the United States*. AdvaMed. <https://www.advamed.org/member-center/resource-library/estimates-of-medical-device-spending-in-the-united-states/>
- Dubov, A., Ogunbajo, A., Altice, F. L., & Fraenkel, L. (2019). Optimizing access to PrEP based on MSM preferences: Results of a discrete choice experiment. *AIDS Care*, 31(5), 545–553.  
<https://doi.org/10.1080/09540121.2018.1557590>
- Espicom. (2009). *The medical device market: USA*.
- FDA. (2019). *Medical device databases*.
- Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic Press.
- Festinger, L. (1957). *A theory of cognitive dissonance* (pp. xi, 291). Stanford University Press.
- Flynn, T. N., & Marley, A. A. J. (2014). Best-worst scaling: Theory and methods. In S. Hess & A. Daly, *Handbook of Choice Modelling* (pp. 178–201). Edward Elgar Publishing.  
<https://doi.org/10.4337/9781781003152.00014>

- Gans, B. M., Mann, N. R., & Becker, B. E. (1993). Delivery of primary care to the physically challenged. *Archives of Physical Medicine and Rehabilitation*, 74(Suppl), s15–s19.
- Gardner-Bonneau, D. (2019). The United States' Journey to Achieve Accessibility of Medical Devices. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (pp. 1448–1455). Springer International Publishing. [https://doi.org/10.1007/978-3-319-96071-5\\_147](https://doi.org/10.1007/978-3-319-96071-5_147)
- Geisen, E. (2022, August 5). *Using Attention Checks in Your Surveys May Harm Data Quality*. Qualtrics. <https://www.qualtrics.com/blog/attention-checks-and-data-quality/>
- GeoPy. (2023). *Geopy* [Python]. geopy. <https://github.com/geopy/geopy>
- Glenn, B., Tieppo Francio, V., Westerhaus, B. D., Goree, J., Strand, N. H., Sparks, D., & Petersen, E. (2023). Accessibility and Ease of Use in Neuromodulation Devices. *Neuromodulation: Technology at the Neural Interface*. <https://doi.org/10.1016/j.neurom.2023.03.003>
- Glynn, L. G., Murphy, A. W., Smith, S. M., Schroeder, K., & Fahey, T. (2010). Interventions used to improve control of blood pressure in patients with hypertension. *Cochrane Database of Systematic Reviews*, 3. <https://doi.org/10.1002/14651858.CD005182.pub4>
- Google. (n.d.). *reCAPTCHA*. reCAPTCHA. Retrieved November 13, 2023, from <https://www.google.com/recaptcha/about/>
- Grabois, E. W., Nosek, M. A., & Rossi, C. D. (1999). Accessibility of primary care physicians' offices for people with disabilities. An analysis of compliance with the Americans With Disabilities Act. *Archives of Family Medicine*, 8(1), 44–51. <https://doi.org/10.1001/archfami.8.1.44>

- Grant, T. (2014). *The involvement of users in the design of home use medical devices: Challenges and incentives for change*. <https://www.semanticscholar.org/paper/The-involvement-of-users-in-the-design-of-home-use-Grant/5b56a8592d2ac96148e026f20ff4f917405e9eea>
- 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](https://doi.org/10.1016/S0191-2615(02)00046-2)
- Groemping, U., Amarov, B., & Xu, H. (2023). *DoE.base: Full Factorials, Orthogonal Arrays and Base Utilities for DoE Packages (1.2-3)* [Computer software]. <https://cran.r-project.org/web/packages/DoE.base/index.html>
- Gummer, T., Roßmann, J., & Silber, H. (2021). Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation. *Sociological Methods & Research*, 50(1), 238–264. <https://doi.org/10.1177/0049124118769083>
- Gwet, K. L. (2008). Intrarater reliability. In *Wiley Encyclopedia of Clinical Trials*. John Wiley & Sons.
- Haan, P. (2004). *Discrete choice labor supply: Conditional logit vs. Random coefficient models*. DIW Discussion Papers.
- Hahn, G. J., & Shapiro, S. S. (1967). *Statistical Models in Engineering*. Wiley.
- 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](https://doi.org/10.1016/S0148-2963(02)00352-1)

- Hanson, K., McPake, B., Nakamba, P., & Archard, L. (2005). Preferences for hospital quality in Zambia: Results from a discrete choice experiment. *Health Economics*, *14*(7), 687–701. <https://doi.org/10.1002/hec.959>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), Article 7825. <https://doi.org/10.1038/s41586-020-2649-2>
- Harris, K., & Keane, M. (1998). A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics*, *89*, 131–157.
- Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, *67*(8), 1688–1699. <https://doi.org/10.1016/j.jbusres.2014.02.015>
- Hayes, A. (2023, April 30). *MRS in Economics: What It Is and the Formula for Calculating It*. Investopedia. [https://www.investopedia.com/terms/m/marginal\\_rate\\_substitution.asp](https://www.investopedia.com/terms/m/marginal_rate_substitution.asp)
- Hedayat, A. S., Sloane, N. J. A., & Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer Science & Business Media.
- Helveston, J. P. (2023a). *cbcTools: Design and Evaluate Choice-Based Conjoint Survey Experiments (R Package) (0.4.0)* [R]. <https://jhelvy.github.io/cbcTools/>
- Helveston, J. P. (2023b). logitr: Fast Estimation of Multinomial and Mixed Logit Models with Preference Space and Willingness-to-Pay Space Utility Parameterizations. *Journal of Statistical Software*, *105*, 1–37. <https://doi.org/10.18637/jss.v105.i10>

Helveston, J. P. (2023c). *cbcTools: Usage*.

<https://jhelvy.github.io/cbcTools/articles/usage.html#assess-power>

Hemmert, G. A., Schons, L. M., Wieseke, J., & Schimmelpfennig, H. (2018). Log-likelihood-based pseudo-R<sup>2</sup> in logistic regression: Deriving sample-sensitive benchmarks. *Sociological Methods & Research*, 47(3), 507–531.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice Analysis: A Primer*. Cambridge University Press.

Hensher, D. A., Stopher, P. R., & Louviere, J. J. (2001). An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: An airline choice application. *Journal of Air Transport Management*, 7(6), 373–379. [https://doi.org/10.1016/S0969-6997\(01\)00031-X](https://doi.org/10.1016/S0969-6997(01)00031-X)

Hess, S., & Daly, A. (Eds.). (2014). Handbook of Choice Modelling. In *Handbook of Choice Modelling* (1st ed.). Edward Elgar Publishing.

<https://www.elgaronline.com/edcollbook/edcoll/9781781003145/9781781003145.xml>

Hoedemakers, M., Karimi, M., Jonker, M., Tsiachristas, A., & Rutten-van Mólken, M. (2022).

Heterogeneity in preferences for outcomes of integrated care for persons with multiple chronic diseases: A latent class analysis of a discrete choice experiment. *Quality of Life Research*, 31(9), 2775–2789. <https://doi.org/10.1007/s11136-022-03147-6>

Hollin, I. L., Paskett, J., Schuster, A. L. R., Crossnohere, N. L., & Bridges, J. F. P. (2022). Best–Worst Scaling and the Prioritization of Objects in Health: A Systematic Review.

*PharmacoEconomics*, 40(9), 883–899. <https://doi.org/10.1007/s40273-022-01167-1>

- Huertas-García, R., Nuñez-Carballosa, A., & Miravittles, P. (2016). Statistical and cognitive optimization of experimental designs in conjoint analysis. *European Journal of Management and Business Economics*, 25(3), 142–149.  
<https://doi.org/10.1016/j.redee.2015.10.001>
- IBM SPSS Statistics for Windows* (Version 28). (2021). [Windows]. IBM.
- IO. (2018, May 21). How To Design For The Web Like a Hero, Not an Asshole. *IO Digital*.  
<https://medium.com/i-o-digital/accessibility-talk-by-nicola-du-toit-steve-barnett-from-empathy-labradors-6e0cc4d0b5f3>
- ipify.org. (2019, October 31). *IP Geolocation Accuracy: How Reliable Is the Technology?* | Blog | *IP Geolocation API*. <https://geo.ipify.org>
- IPinfo. (2023). *The trusted source for IP address data, leading IP data provider*.  
<https://ipinfo.io/products/ip-geolocation-api>
- IPinfo Team. (2020, September 25). *Geolocation data: How accurate can it really be?* IPinfo Blog. <https://ipinfo.io/blog/ip-geolocation-accuracy/>
- IPlocation. (2018, December 20). *How accurate is IP-based Geolocation Lookup?*  
<https://www.iplocation.net/geolocation-accuracy>
- ISO 9241-11*. (2018). Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>
- ISO 9241-171*. (2008). Ergonomics of Human-System Interaction — Part 171: Guidance on Software Accessibility. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-171:ed-1:v1:en>

- Iwarsson, S., & Ståhl, A. (2003). Accessibility, usability and universal design—Positioning and definition of concepts describing person-environment relationships. *Disability and Rehabilitation*, 25(2), 57–66. <https://doi.org/10.1080/dre.25.2.57.66>
- Jaeger, S. R., Jørgensen, A. S., Aaslyng, M. D., & Bredie, W. L. P. (2008). Best–worst scaling: An introduction and initial comparison with monadic rating for preference elicitation with food products. *Food Quality and Preference*, 19(6), 579–588. <https://doi.org/10.1016/j.foodqual.2008.03.002>
- Jansen, S. J. T. (2011). The Multi-attribute Utility Method. In S. J. T. Jansen, H. C. C. H. Coolen, & R. W. Goetgeluk (Eds.), *The Measurement and Analysis of Housing Preference and Choice* (pp. 101–125). Springer Netherlands. [https://doi.org/10.1007/978-90-481-8894-9\\_5](https://doi.org/10.1007/978-90-481-8894-9_5)
- Johnson, F. R., 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: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 16(1), 3–13. <https://doi.org/10.1016/j.jval.2012.08.2223>
- 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/10.1016/j.jval.2018.07.876>
- Johnson, R., & Orme, B. (2003). *Getting the Most from CBC*. Sawtooth Software, Inc. <https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc>

- Jones, J. M. (2023, February 22). *U.S. LGBT Identification Steady at 7.2%*. Gallup.Com.  
<https://news.gallup.com/poll/470708/lgbt-identification-steady.aspx>
- Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2019). Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Economics*, 28(3), 350–363. <https://doi.org/10.1002/hec.3846>
- Kailes, J. I. (2007). The patient’s perspective on access to medical equipment. In J. M. Winters & M. F. Story (Eds.), *Medical instrumentation: Accessibility and usability considerations* (pp. 3–11). CRC Press.
- Kailes, J. I., Premo, B., & Richards, C. (2006). Toward a New Health Care Policy: Accessible Medical Equipment and Instrumentation. *Medical Instrumentation: Accessibility and Usability Considerations*, 41–57.
- Kasali, J., & Adeyemi, A. A. (2022). Model-Data Fit using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and The Sample-Size-Adjusted BIC. *Square : Journal of Mathematics and Mathematics Education*, 4(1), Article 1.  
<https://doi.org/10.21580/square.2022.4.1.11297>
- Kessels, R. (2016). Homogeneous versus heterogeneous designs for stated choice experiments: Ain’t homogeneous designs all bad? *Journal of Choice Modelling*, 21, 2–9.  
<https://doi.org/10.1016/j.jocm.2016.08.001>
- Kjaer, T. (2005). *A review of the discrete choice experiment-with emphasis on its application in health care*. Syddansk Universitet Denmark.

- Kjær, T., & Gyrd-Hansen, D. (2008). Preference heterogeneity and choice of cardiac rehabilitation program: Results from a discrete choice experiment. *Health Policy*, *85*(1), 124–132. <https://doi.org/10.1016/j.healthpol.2007.07.002>
- Kraus, L., Lauer, E., Coleman, R., & Houtenville, A. (2017). *2017 Disability Statistics Annual Report*.
- Kroes, E. P., & Sheldon, R. J. (1988). Stated Preference Methods: An Introduction. *Journal of Transport Economics and Policy*, *22*(1), 11–25.
- Kuhfeld, W. F. (2006). Construction of efficient designs for discrete choice experiments. In *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*. Sage Publications. Sage Publications.
- Kuhfeld, W. F. (2010). *Marketing Research Methods in SAS*. SAS Institute Inc.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The Journal of Political Economy*, *74*(2), 132–157.
- Lancsar, E., Fiebig, D. G., & Hole, A. R. (2017). Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. *PharmacoEconomics*, *35*(7), 697–716. <https://doi.org/10.1007/s40273-017-0506-4>
- Lancsar, E., & Louviere, J. (2006). Deleting ‘irrational’ responses from discrete choice experiments: A case of investigating or imposing preferences? *Health Economics*, *15*(8), 797–811. <https://doi.org/10.1002/hec.1104>
- Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User’s Guide. *PharmacoEconomics*, *26*, 661–677. <https://doi.org/10.2165/00019053-200826080-00004>

- Lancsar, E., Louviere, J., & Flynn, T. (2007). Several methods to investigate relative attribute impact in stated preference experiments. *Social Science & Medicine*, *64*(8), 1738–1753. <https://doi.org/10.1016/j.socscimed.2006.12.007>
- Laver, K., Ratcliffe, J., George, S., Lester, L., & Crotty, M. (2013). Preferences for rehabilitation service delivery: A comparison of the views of patients, occupational therapists and other rehabilitation clinicians using a discrete choice experiment. *Australian Occupational Therapy Journal*, *60*(2), 93–100. <https://doi.org/10.1111/1440-1630.12018>
- Law, M., Cooper, B., Strong, S., Stewart, D., Rigby, P., & Letts, L. (1996). The Person-Environment-Occupation Model: A Transactive Approach to Occupational Performance. *Canadian Journal of Occupational Therapy*, *63*(1), 9–23. <https://doi.org/10.1177/000841749606300103>
- Lewis, J. R. (2012). Usability Testing. In *Handbook of Human Factors and Ergonomics* (pp. 1267–1312). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118131350.ch46>
- Lewis, J. R. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human-Computer Interaction*, *34*(7), 577–590. <https://doi.org/10.1080/10447318.2018.1455307>
- Lin, M., Shaewitz, D., Overton, C., & Smith, D.-M. (2018). *A hidden market: The purchasing power of working-age adults with disabilities*.
- Lin, Y., Peterson, G. L., & Rogerson, P. A. (1988). A nested urban recreation site choice model. *Leisure Sciences*, *10*(1), 1–15. <https://doi.org/10.1080/01490408809512172>

- Linley, W. G., & Hughes, D. A. (2013). Decision-Makers' Preferences for Approving New Medicines in Wales: A Discrete-Choice Experiment with Assessment of External Validity. *Pharmacoeconomics*, 31(4), 345–355. <https://doi.org/10.1007/s40273-013-0030-0>
- Louviere, J., Hensher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge University Press.
- Louviere, J. J. (2001). Choice Experiments: An Overview of Concepts and Issues. In J. Bennett & R. Blamey (Eds.), *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar Publishing.
- 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. [https://doi.org/10.1016/S1755-5345\(13\)70014-9](https://doi.org/10.1016/S1755-5345(13)70014-9)
- Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research*, 20(4), 350–367. <https://doi.org/10.2307/3151440>
- Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15(3), 215–233. [https://doi.org/10.1016/0022-2496\(77\)90032-3](https://doi.org/10.1016/0022-2496(77)90032-3)
- Luce, R. D. (2008). Luce's choice axiom. *Scholarpedia*, 3(12), 8077. <https://doi.org/10.4249/scholarpedia.8077>
- Magasi, S., & Marshall, H. (2021). Accessible Medical Diagnostic Equipment: A Rapid Review. *Archives of Physical Medicine and Rehabilitation*, 102(10), e113. <https://doi.org/10.1016/j.apmr.2021.07.457>

- Mangham, L. J., Hanson, K., & McPake, B. (2009). How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning, 24*(2), 151–158. <https://doi.org/10.1093/heapol/czn047>
- Mariel, P., Ayala, A. de, Hoyos, D., & Abdullah, S. (2013). Selecting random parameters in discrete choice experiment for environmental valuation: A simulation experiment. *Journal of Choice Modelling, 7*, 44–57. <https://doi.org/10.1016/j.jocm.2013.04.008>
- Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., Jacobsen, J. B., Liebe, U., Olsen, S. B., Sagebiel, J., & Thiene, M. (2021). Experimental Design. In P. Mariel, D. Hoyos, J. Meyerhoff, M. Czajkowski, T. Dekker, K. Glenk, J. B. Jacobsen, U. Liebe, S. B. Olsen, J. Sagebiel, & M. Thiene (Eds.), *Environmental Valuation with Discrete Choice Experiments: Guidance on Design, Implementation and Data Analysis* (pp. 37–49). Springer International Publishing. [https://doi.org/10.1007/978-3-030-62669-3\\_3](https://doi.org/10.1007/978-3-030-62669-3_3)
- Markwalder, A. (2005). *Disability rights advocates. In A call to action: A guide for managed care plans serving Californians with disabilities.*
- Marschak, J. (1959). *Binary Choice Constraints on Random Utility Indicators* (Cowles Foundation Discussion Paper 74). Cowles Foundation for Research in Economics, Yale University. <https://econpapers.repec.org/paper/cwlcwldpp/74.htm>
- Marshall, D., Bridges, J. F. P., Hauber, B., Cameron, R., Donnalley, L., Fyie, K., & Reed Johnson, F. (2010). Conjoint Analysis Applications in Health—How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008. *The Patient: Patient-Centered Outcomes Research, 3*(4), 249–256. <https://doi.org/10.2165/11539650-000000000-00000>

- Mazzanti, M. (2003). Discrete choice models and valuation experiments. *Journal of Economic Studies*, 30(6), 584–604. <https://doi.org/10.1108/01443580310504453>
- McClain, L., Medrano, D., Marcum, M., & Schukar, J. (2000). A Qualitative Assessment of Wheelchair Users' Experience with ADA Compliance, Physical Barriers, and Secondary Health Conditions. *Topics in Spinal Cord Injury Rehabilitation*, 6(1), 99–118. <https://doi.org/10.1310/ENAP-Y4E7-RG05-6YV5>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.
- McFadden, D. (1977). *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*. Cowles Foundation for Research in Economics, Yale University. <https://elischolar.library.yale.edu/cowles-discussion-paper-series/707>
- McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3), 351–378.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470. [https://doi.org/10.1002/1099-1255\(200009/10\)15:5<447::AID-JAE570>3.0.CO;2-1](https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1)
- Mendonca, R. (2010). *Accessibility Labeling and its Effects on Medical Device Purchase Decisions by Consumers with Disabilities* [Dissertation]. University of Wisconsin-Milwaukee.
- Mendonca, R., & Smith, R. O. (2007). *Validity analysis: MED-AUDIT (Medical Equipment Device-Accessibility and Universal Design Information Tool)*. RESNA 30th International Conference on Technology and Disability: Research, Design, Practice and Policy.
- Mendonca, R., Smith, R. O., Lemke, M., & Winters, J. (2008). *Expert Validation of the MED-AUDIT (Medical Equipment Device-Accessibility & Universal Design Information Tool)*.

Proceedings of the RESNA 31st International Conference on Campaigning for Assistive Technology, Designing Policy, Promoting Research and Strengthening Practice.

Meyerhoff, J., Oehlmann, M., & Weller, P. (2015). The Influence of Design Dimensions on Stated Choices in an Environmental Context. *Environmental and Resource Economics*, 61(3), 385–407. <https://doi.org/10.1007/s10640-014-9797-5>

Microsoft Corporation. (2023a). *Microsoft Excel* [Windows]. Microsoft Corporation. <https://office.microsoft.com/excel>

Microsoft Corporation. (2023b). *Visual Studio Code* [Windows]. Microsoft Corporation. <https://code.visualstudio.com/>

Milte, R., Ratcliffe, J., Miller, M. D., Whitehead, C., Cameron, I. D., & Crotty, M. (2013). What are frail older people prepared to endure to achieve improved mobility following hip fracture? A Discrete Choice Experiment. *Journal of Rehabilitation Medicine*, 45(1), Article 1. <https://doi.org/10.2340/16501977-1054>

Mott, D. J., Chami, N., & Tervonen, T. (2020). Reporting Quality of Marginal Rates of Substitution in Discrete Choice Experiments That Elicit Patient Preferences. *Value in Health*, 23(8), 979–984. <https://doi.org/10.1016/j.jval.2020.04.1831>

Mühlbacher, A., & Johnson, F. R. (2016). Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. *Applied Health Economics and Health Policy*, 14(3), 253–266. <https://doi.org/10.1007/s40258-016-0232-7>

Naik Panvelkar, P., Armour, C., & Saini, B. (2010). Community Pharmacy-Based Asthma Services—What Do Patients Prefer? *Journal of Asthma*, 47(10), 1085–1093. <https://doi.org/10.3109/02770903.2010.514638>

National Science Foundation. (2020). *I-CORPS - NSF Innovation Corps*.

<https://new.nsf.gov/funding/initiatives/i-corps>

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

Willingness to pay estimates derived from a discrete choice experiment. *Social Science & Medicine*, 70(9), 1317–1325. <https://doi.org/10.1016/j.socscimed.2009.12.027>

NIH. (2022, October 3). *The Basics*. National Institutes of Health. <https://www.nih.gov/health-information/nih-clinical-research-trials-you/basics>

NobelPrize.org. (2023). *Daniel L. McFadden – Facts*. Nobel Prize Outreach AB.

<https://www.nobelprize.org/prizes/economic-sciences/2000/mcfadden/facts/>

North Carolina Office on Disability and Health. (2007). *Removing barriers to health care: A guide for health professionals*. The University of North Carolina.

<https://fpg.unc.edu/publications/removing-barriers-health-care-guide-health-professionals>

Nosek, M. (2000). Overcoming the odds: The health of women with physical disabilities in the United States. *Archives of Physical Medicine and Rehabilitation - ARCH PHYS MED*

*REHABIL*, 81(2), 135–138. <https://doi.org/10.1053/apmr.2000.0810135>

Nosek, M., & Howland, C. (1997). Breast and cervical cancer screening among women with physical disabilities. *Archives of Physical Medicine and Rehabilitation*, 78(12, Supplement

5), S39–S44. [https://doi.org/10.1016/S0003-9993\(97\)90220-3](https://doi.org/10.1016/S0003-9993(97)90220-3)

Nosek, M., Young, M., Rintala, D., Howland, C., Foley, C., & Bennett, J. (1995). Barriers to

Reproductive Health Maintenance Among Women with Physical Disabilities. *Journal of Women's Health*, 4(5), 505–518. <https://doi.org/10.1089/jwh.1995.4.505>

- Noto, S., Murata, T., Saito, S., Watanabe, T., & Kobayashi, M. (2023). Preferences for Rehabilitation in Persons with a History of Stroke: A Discrete Choice Experiment. *Patient Preference and Adherence*, *17*, 1611–1620. <https://doi.org/10.2147/PPA.S416699>
- NVivo (Version 13). (2020). [Windows]. QSR International.
- Oliver, D., Deal, K., Howard, M., Qian, H., Agarwal, G., & Guenter, D. (2019). Patient trade-offs between continuity and access in primary care interprofessional teaching clinics in Canada: A cross-sectional survey using discrete choice experiment. *BMJ Open*, *9*(3), e023578. <https://doi.org/10.1136/bmjopen-2018-023578>
- Orme, B. (2019). *Sample Size Issues for Conjoint Analysis Studies*. <https://sawtoothsoftware.com/resources/technical-papers/sample-size-issues-for-conjoint-analysis-studies>
- Ostchega, Y., Fryar, C. D., Nwankwo, T., & Nguyen, D. T. (2020). *Hypertension prevalence among adults aged 18 and over: United States, 2017–2018*. (364; NCHS Data Brief). National Center for Health Statistics.
- Patient Protection and Affordable Care Act, Pub. L. No. 111–148 (2010). <https://www.govinfo.gov/content/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf>
- Peacock, G., Iezzoni, L. I., & Harkin, T. R. (2015). Health Care for Americans with Disabilities—25 Years after the ADA. *The New England Journal of Medicine*, *373*(10), 892–893. <https://doi.org/10.1056/NEJMp1508854>
- Pearce, A., Harrison, M., Watson, V., Street, D. J., Howard, K., Bansback, N., & Bryan, S. (2021). Respondent Understanding in Discrete Choice Experiments: A Scoping Review. *The Patient*, *14*(1), 17–53. <https://doi.org/10.1007/s40271-020-00467-y>

- Pérez-Troncoso, D. (2020). *A step-by-step guide to design, implement, and analyze a discrete choice experiment* (arXiv:2009.11235). arXiv.  
<https://doi.org/10.48550/arXiv.2009.11235>
- Petrie, H., & Bevan, N. (2009). The evaluation of accessibility, usability and user experience. In *The Universal Access Handbook* (1st Edition). CRC Press.  
[www.crcpress.com/product/isbn/9780805862805](http://www.crcpress.com/product/isbn/9780805862805)
- Philips, H., Mahr, D., Remmen, R., Weverbergh, M., De Graeve, D., & Van Royen, P. (2012). Predicting the place of out-of-hours care—A market simulation based on discrete choice analysis. *Health Policy*, *106*(3), 284–290.  
<https://doi.org/10.1016/j.healthpol.2012.04.010>
- Pleskac, T. J. (2015). Decision and Choice: Luce’s Choice Axiom. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 895–900). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.43031-X>
- Posit team. (2023). *RStudio: Integrated Development Environment for R* (2023.6.0.421) [Computer software]. Posit Software. <http://www.posit.co/>.
- Powlik, J. J., & Karshmer, A. I. (2002). When accessibility meets usability. *Universal Access in the Information Society*, *1*(3), 217–222. <https://doi.org/10.1007/s10209-002-0020-z>
- Privitera, M. B., Evans, M., & Southee, D. (2017). Human factors in the design of medical devices – Approaches to meeting international standards in the European Union and USA. *Applied Ergonomics*, *59*, 251–263. <https://doi.org/10.1016/j.apergo.2016.08.034>

Prolific. (2023a, October 24). *Find, filter, favourite: How to select participants for AI tasks*.

<https://www.prolific.com/blog/find-filter-favourite-how-to-select-participants-for-ai-tasks>

Prolific. (2023b, November 1). *Approvals, rejections & returns*. Prolific. [https://researcher-](https://researcher-help.prolific.com/hc/en-gb/articles/360009092394-Approvals-rejections-returns)

[help.prolific.com/hc/en-gb/articles/360009092394-Approvals-rejections-returns](https://researcher-help.prolific.com/hc/en-gb/articles/360009092394-Approvals-rejections-returns)

Prolific. (2023c, November 13). *Verifying your identity FAQ*. Participant Help Centre.

<https://participant-help.prolific.com/hc/en-gb/articles/360017282699-Verifying-your-identity-FAQ>

Python Software Foundation. (2023). *Python* (3.12) [Computer software]. Python Software

Foundation. <https://www.python.org/>

Quaife, M., Terris-Prestholt, F., Di Tanna, G. L., & Vickerman, P. (2018). How well do discrete

choice experiments predict health choices? A systematic review and meta-analysis of external validity. *The European Journal of Health Economics*, *19*(8), 1053–1066.

<https://doi.org/10.1007/s10198-018-0954-6>

Qualtrics. (n.d.-a). *Fraud Detection*. Retrieved November 9, 2023, from

<https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/>

Qualtrics. (n.d.-b). *Understanding Your Dataset*. Retrieved November 15, 2023, from

<https://www.qualtrics.com/support/survey-platform/data-and-analysis-module/data/download-data/understanding-your-dataset/>

Qualtrics. (2023). *Qualtrics* [Online]. Qualtrics. <https://www.qualtrics.com/>

- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rabbitt, E. (2019, August 6). *To Be More Inclusive, Support a Spectrum of Abilities*. ASAE. <https://www.asaecenter.org:443/en/resources/articles/foundation/2019/to-be-more-inclusive-support-a-spectrum-of-abilities>
- Radley, A., van der Pol, M., & Dillon, J. F. (2019). Application of a discrete choice experiment approach to support the design of a hepatitis C testing service in primary care. *International Journal of Drug Policy*, *65*, 1–7. <https://doi.org/10.1016/j.drugpo.2018.12.008>
- Ratcliffe, J., Bekker, H. L., Dolan, P., & Edlin, R. (2009). Examining the attitudes and preferences of health care decision-makers in relation to access, equity and cost-effectiveness: A discrete choice experiment. *Health Policy*, *90*(1), 45–57. <https://doi.org/10.1016/j.healthpol.2008.09.001>
- Ray, P. (1973). Independence of Irrelevant Alternatives. *Econometrica*, *41*(5), 987–991. <https://doi.org/10.2307/1913820>
- Raymaker, D. M., McDonald, K. E., Ashkenazy, E., Gerrity, M., Baggs, A. M., Kripke, C., Hourston, S., & Nicolaidis, C. (2017). Barriers to healthcare: Instrument development and comparison between autistic adults and adults with and without other disabilities. *Autism : The International Journal of Research and Practice*, *21*(8), 972–984. <https://doi.org/10.1177/1362361316661261>

- Ricee, S. (2020, April 19). *What is Permanent Disability?* <https://diversity.social/permanent-disability/>
- Roh, H.-J., & Khan, A. M. (2013). Enhancing algorithmic base for discrete choice modelling. *KSCFE Journal of Civil Engineering*, 17(7), 1798–1809. <https://doi.org/10.1007/s12205-013-0318-9>
- Rossi, R. J. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference*. John Wiley & Sons.
- Rouvinen, S., & Matero, J. (2013). Stated preferences of Finnish private homeowners for residential heating systems: A discrete choice experiment. *Biomass and Bioenergy*, 57, 22–32. <https://doi.org/10.1016/j.biombioe.2012.10.010>
- Rubin, J. M. (2011, November 1). Can a computer generate a truly random number? *MIT School of Engineering*. <https://engineering.mit.edu/engage/ask-an-engineer/can-a-computer-generate-a-truly-random-number/>
- Ryan, M. (2000). Using conjoint analysis to elicit preferences for health care. *BMJ*, 320(7248), 1530–1533. <https://doi.org/10.1136/bmj.320.7248.1530>
- Ryan, M., & Wordsworth, S. (2000). Sensitivity of Willingness to Pay Estimates to the Level of Attributes in Discrete Choice Experiments. *Scottish Journal of Political Economy*, 47(5), 504–524. <https://doi.org/10.1111/1467-9485.00176>
- Sade, R. M. (2012). The graying of America: Challenges and Controversies. *The Journal of Law, Medicine, & Ethics*, 40(1), 6–9. <https://doi.org/10.1111/j.1748-720X.2012.00639.x>

- Sanchez, J., Byfield, G., Brown, T., LaFavor, K., Murphy, D., & Laud, P. (2000). Perceived Accessibility Versus Actual Physical Accessibility of Healthcare Facilities. *Rehabilitation Nursing, 25*(1), 6–9. <https://doi.org/10.1002/j.2048-7940.2000.tb01849.x>
- Sarrias, M., & Daziano, R. (2017). Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The gmnL Package. *Journal of Statistical Software, 79*, 1–46. <https://doi.org/10.18637/jss.v079.i02>
- Scarborough-Kaufman, C. (2019). Publicly-Researchable Accessibility Information: Problems, Prospects and Recommendations for Inclusion. *Social Inclusion, 7*(1), 164–172. <https://doi.org/10.17645/si.v7i1.1651>
- Schlereth, C., Eckert, C., & Skiera, B. (2012). Using discrete choice experiments to estimate willingness-to-pay intervals. *Marketing Letters, 23*(3), 761–776. <https://doi.org/10.1007/s11002-012-9177-2>
- Schopp, L., Sanford, T., Hagglund, K., Gay, J., & Coatney, M. (2002). Removing service barriers for women with physical disabilities: Promoting accessibility in the gynecologic care setting. *Journal of Midwifery & Women's Health, 47*(2), 74–79. [https://doi.org/10.1016/S1526-9523\(02\)00216-7](https://doi.org/10.1016/S1526-9523(02)00216-7)
- SelectUSA. (2019). *Medical technology spotlight: The medical technology industry in the United States*.
- Shields, G. E., Wright, S., Wells, A., Doherty, P., Capobianco, L., & Davies, L. M. (2021). Delivery preferences for psychological intervention in cardiac rehabilitation: A pilot discrete choice experiment. *Open Heart, 8*(2), e001747. <https://doi.org/10.1136/openhrt-2021-001747>

- Shinohara, K., Wobbrock, J. O., & Pratt, W. (2018). Incorporating Social Factors in Accessible Design. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 149–160. <https://doi.org/10.1145/3234695.3236346>
- Sloane, N. J. A. (2007). *Orthogonal Arrays*. <http://neilsloane.com/oadir/>
- Smith, R. O., Barnekow, K., Lemke, M. R., Mendonca, R., Winter, M., Schwanke, T., & Winters, J. M. (2007). Development of the Medical Equipment Device Accessibility and Universal Design Information Tool. In *Medical Instrumentation*. CRC Press.
- Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints*, 39(2), 17–26.
- 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>
- Spassiani, N., & Friedman, C. (2014). Stigma: Barriers to Culture and Identity for People With Intellectual Disability. *Inclusion*, 2, 329–341. <https://doi.org/10.1352/2326-6988-2.4.329>
- Story, M. F., Mueller, J., & Montoya-Weiss, M. (2002). *Evaluating the universal design performance of products*.  
[https://projects.ncsu.edu/ncsu/design/cud/pubs\\_p/pudperformproduct.htm](https://projects.ncsu.edu/ncsu/design/cud/pubs_p/pudperformproduct.htm)
- Story, M. F., Winters, J. M., Lemke, M. R., Barr, A., Omiatek, E., Janowitz, I., Brafman, D., & Rempel, D. (2010). Development of a method for evaluating accessibility of medical equipment for patients with disabilities. *Applied Ergonomics*, 42(1), 178–183.  
<https://doi.org/10.1016/j.apergo.2010.07.003>

- Street, D. J., Burgess, L., Viney, R., & Louviere, J. (2008). Designing Discrete Choice Experiments for Health Care. In M. Ryan, K. Gerard, & M. Amaya-Amaya (Eds.), *Using Discrete Choice Experiments to Value Health and Health Care* (pp. 47–72). Springer Netherlands.  
[https://doi.org/10.1007/978-1-4020-5753-3\\_2](https://doi.org/10.1007/978-1-4020-5753-3_2)
- Szinay, D., Cameron, R., Naughton, F., Whitty, J. A., Brown, J., & Jones, A. (2021). Understanding Uptake of Digital Health Products: Methodology Tutorial for a Discrete Choice Experiment Using the Bayesian Efficient Design. *Journal of Medical Internet Research*, 23(10), e32365. <https://doi.org/10.2196/32365>
- Telser, H., & Zweifel, P. (2007). Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics*, 39(1), 69–78.  
<https://doi.org/10.1080/00036840500427858>
- Thai, T., Bliemer, M., Chen, G., Spinks, J., de New, S., & Lancsar, E. (2023). Comparison of a full and partial choice set design in a labeled discrete choice experiment. *Health Economics*, 32(6), 1284–1304. <https://doi.org/10.1002/hec.4666>
- The pandas development team. (2023). *pandas-dev/pandas: Pandas (v2.1.3)* [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.10107975>
- Theofanos, M. F., & Redish, J. (Ginny). (2003). Bridging the gap: Between accessibility and usability. *Interactions*, 10(6), 36–51. <https://doi.org/10.1145/947226.947227>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286.  
<https://doi.org/10.1037/h0070288>
- Thurstone, L. L. (1931). The Indifference Function. *The Journal of Social Psychology*, 2(2), 139–167. <https://doi.org/10.1080/00224545.1931.9918964>

- Timmermans, H. (2001). Spatial Choice Models. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14768–14771). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/02520-1>
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Trapero-Bertran, M., Rodríguez-Martín, B., & López-Bastida, J. (2019). What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. *PLOS ONE*, *14*(7), e0219905. <https://doi.org/10.1371/journal.pone.0219905>
- Tsafarakis, S., Grigoroudis, E., & Matsatsinis, N. (2011). Consumer choice behaviour and new product development: An integrated market simulation approach. *Journal of the Operational Research Society*, *62*(7), 1253–1267. <https://doi.org/10.1057/jors.2010.70>
- Ugba, E. R., & Gertheiss, J. (2023). A modification of McFadden's for binary and ordinal response models. *Communications for Statistical Applications and Methods*, *30*(1), 49–63. <https://doi.org/10.29220/CSAM.2023.30.1.049>
- Uhlig, K., Patel, K., Ip, S., Kitsios, G. D., & Balk, E. M. (2013). Self-Measured Blood Pressure Monitoring in the Management of Hypertension. *Annals of Internal Medicine*, *159*(3), 185–194. <https://doi.org/10.7326/0003-4819-159-3-201308060-00008>
- United States Access Board. (2010). *Standards for accessible medical diagnostic equipment*.
- United States Access Board. (2017). *VA Adopts New Standards for Medical Diagnostic Equipment*.
- United States Census Bureau. (2019). *American Community Survey—2018: ACS 1-Year Estimates Data Profiles*.

- U.S. Census Bureau. (2022). *ACS 1-Year Estimates Public Use Microdata Sample* [dataset]. American Community Survey. <https://data.census.gov/table/ACSSPP1Y2022.S0201?q=american+community+survey&y=2022>
- Usability.gov. (2013, December 16). *Usability and Accessibility: Looking at User Experience through Two Lenses*. Department of Health and Human Services. <https://www.usability.gov/get-involved/blog/2013/01/accessibility-and-usability.html>
- van den Broek-Altenburg, E., & Atherly, A. (2020). Using discrete choice experiments to measure preferences for hard to observe choice attributes to inform health policy decisions. *Health Economics Review*, 10(1), 18. <https://doi.org/10.1186/s13561-020-00276-x>
- Van Der Pol, M., Currie, G., Kromm, S., & Ryan, M. (2014). Specification of the Utility Function in Discrete Choice Experiments. *Value in Health*, 17(2), 297–301. <https://doi.org/10.1016/j.jval.2013.11.009>
- Vanderheiden, G., & Tobias, J. (1998). Barriers, Incentives and Facilitators for Adoption of Universal Design Practices by Consumer Product Manufacturers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(6), 584–588. <https://doi.org/10.1177/154193129804200613>
- Vanderheiden, G., & Tobias, J. (2000). Universal Design of Consumer Products: Current Industry Practice and Perceptions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(32), 6–19. <https://doi.org/10.1177/154193120004403206>

- Vass, C., Boeri, M., Karim, S., Marshall, D., Craig, B., Ho, K.-A., Mott, D., Ngorsuraches, S., Badawy, S. M., Mühlbacher, A., Gonzalez, J. M., & Heidenreich, S. (2022). Accounting for Preference Heterogeneity in Discrete-Choice Experiments: An ISPOR Special Interest Group Report. *Value in Health*, 25(5), 685–694.  
<https://doi.org/10.1016/j.jval.2022.01.012>
- Vass, C., Rigby, D., & Payne, K. (2017). The Role of Qualitative Research Methods in Discrete Choice Experiments: A Systematic Review and Survey of Authors. *Medical Decision Making*, 37(3), 298–313. <https://doi.org/10.1177/0272989X16683934>
- Veldwijk, J., Lambooi, M. S., de Bekker-Grob, E. W., Smit, H. A., & Wit, G. A. de. (2014). The Effect of Including an Opt-Out Option in Discrete Choice Experiments. *PLOS ONE*, 9(11), e111805. <https://doi.org/10.1371/journal.pone.0111805>
- Veltman, A., Stewart, D., Tardif, G., & Branigan, M. (2001). Perceptions of primary healthcare services among people with physical disabilities. Part 1: Access issues. *MedGenMed : Medscape General Medicine*, 3(2), 18.
- Viney, R., Lancsar, E., & Louviere, J. (2002). Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Review of Pharmacoeconomics & Outcomes Research*, 2(4), 319–326. <https://doi.org/10.1586/14737167.2.4.319>
- Vojáček, O., & Pecáková, I. (2010). Comparison of Discrete Choice Models for Economic Environmental Research. *Prague Economic Papers*, 19(1), 35–53.  
<https://doi.org/10.18267/j.pep.363>

- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*(2), 228–243. <https://doi.org/10.1037/a0027127>
- W3C Web Accessibility Initiative. (2022, March 31). *Introduction to Web Accessibility*. Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/fundamentals/accessibility-intro/>
- Walker, D. A., & Smith, T. J. (2016). Nine pseudo R2 indices for binary logistic regression models. *Journal of Modern Applied Statistical Methods, 15*(1), 848–854.
- Wang, Y., Wang, Z., Wang, Z., Li, X., Pang, X., & Wang, S. (2021). Application of Discrete Choice Experiment in Health Care: A Bibliometric Analysis. *Frontiers in Public Health, 9*. <https://www.frontiersin.org/articles/10.3389/fpubh.2021.673698>
- Watson, V., Luchini, S., Regier, D., & Schulz, R. (2020). Monetary analysis of health outcomes. In C. Guerriero (Ed.), *Cost-Benefit Analysis of Environmental Health Interventions* (pp. 73–93). Academic Press. <https://doi.org/10.1016/B978-0-12-812885-5.00004-4>
- Wheeler, R. E. (2009). *Comments on algorithmic design*. <https://cran.r-project.org/web/packages/AlgDesign/vignettes/AlgDesign.pdf>
- Wheeler, R. E., & Braun, J. M. (2022). *Package ‘AlgDesign.’* <https://cran.r-project.org/package=AlgDesign>
- Wongpakaran, N., Wongpakaran, I., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen’s Kappa and Gwet’s ACq when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Medical Research Methodology, 13*(61). <https://doi.org/10.1186/1471-2288-13-61>

- Woo, J., Shin, J., Kim, H., & Moon, H. (2022). Which consumers are willing to pay for smart car healthcare services? A discrete choice experiment approach. *Journal of Retailing and Consumer Services*, 69, 103084. <https://doi.org/10.1016/j.jretconser.2022.103084>
- Xie, Y., Zhao, K., & Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis & Prevention*, 47, 36–44. <https://doi.org/10.1016/j.aap.2011.12.012>
- Zhang, D. (2017). A coefficient of determination for generalized linear models. *The American Statistician*, 71(4), 310–316.
- Zickafoose, J. S., DeCamp, L. R., & Prosser, L. A. (2015). Parents' Preferences for Enhanced Access in the Pediatric Medical Home: A Discrete Choice Experiment. *JAMA Pediatrics*, 169(4), 358–364. <https://doi.org/10.1001/jamapediatrics.2014.3534>
- Zoom Meetings*. (2022). [Windows]. Zoom Video Communications Inc. <https://zoom.us/>

## VII. Appendices

### Appendix A: Syntax – Choice Matrix Design in R

```
library(cbcTools)
library(logitr)

##### GLOBAL VARIABLES #####

Nresp = 200 #identify number of respondents
Nchoice = 16 #global variable for the number of choice sets
Nalt = 4 #global variable for the number of alternatives per choice set
Nbreaks = 20 #global variable used in //POWER ANALYSIS SIMULATION//
HeadN <- Nalt*Nchoice #secondary variable to only print out the first respID set in
the DB-Efficient design, other respIDs are just repetition of the same design.

Seed = 7770
set.seed(Seed) #set the seed for randomization.

##### GENERATING FULL-FACTORIAL DESIGN #####

#create full-factorial design based on attributes and Levels
profiles <- cbc_profiles(
  cost = seq(20, 30, 5), #generates sequence of numbers between 20 and 30 inclusive,
5 by 5. --Ratio Data
  brand = c("HomeHealth", "Omron"), #--Nominal Data
  UserRating = seq(3.2, 4.8, 0.8), #generates sequence of numbers between 3.2 and 4.8
inclusive, 0.8 by 0.8. --Ratio Data
  UAscore = c("Unavailable", "High", "Medium", "Low") #--Nominal Data
)
profiles

##### RESTRICTIONS #####

rstrct_profiles <- cbc_restrict(
  profiles,
  cost == 20 & brand == "Omron" & UserRating == 4.8 & UAscore == "High", #exclude
dominant alternative with the best levels for each attribute
)
rstrct_profiles
```

```

##### DB-EFFICIENT DESIGN #####

#create DB-efficient design from restricted full-factorial design
design_dbeff <- cbc_design(
  profiles = rstrct_profiles,
  n_resp = Nresp,
  n_alts = Nalt, #number of alternatives in each choice set
  n_q = Nchoice, #number of "questions" or choice sets
  n_start = 10, #numeric value indicating the number of random start designs
  priors = list(
    cost = -0.13,
    brand = 0, #prior from Phase I did not achieve significance
    UserRating = 3.2,
    UAscore = c(5.2, 1.9, -1.8)
  ), #using priors from Phase I
  max_iter = 10000,
  method = "Modfed",
  keep_db_error = TRUE,
  parallel = TRUE
)

DBerr <- as.numeric(design_dbeff$db_err) #the output of cbc_design is a list, reads
the actual value of db_err
DF_design_dbeff <- as.data.frame(design_dbeff$design) #the output of cbc_design is a
list, reads the actual design and creates a dataframe
DF_design_dbeff <- subset(DF_design_dbeff, select = -blockID) #removes the blockID
column from the design, as there is only one block
FirstDesign <- head(DF_design_dbeff, HeadN) #only saves the design for the first
respID, the rest of the respIDs are repetitions of the first design
FirstDesign

##### BALANCE & OVERLAP #####

cbc_balance(FirstDesign) #examine the balance of the DB-efficient design
cbc_overlap(FirstDesign) #examine the overlap across in the choice sets

##### CHOICE SIMULATION #####

#simulate output using random choices
rndchoices <- cbc_choices(
  design = DF_design_dbeff,
  obsID = "obsID",
  priors = list(

```

```

    cost = -0.13,
    brand = 0, #prior from Phase I did not achieve significance
    UserRating = 3.2,
    UAscore = c(5.2, 1.9, -1.8)
  )
)

rndchoices <- as.data.frame(rndchoices) #cbc_choices output is a List as of v0.4.0,
converts the output to a dataframe

##### POWER ANALYSIS SIMULATION #####

#perform simulated power analysis for the simulated choices above, using "logitr"
library to estimate for a MNL model
power <- cbc_power(
  data = rndchoices,
  pars = c("cost", "brand", "UserRating", "UAscore"),
  outcome = "choice",
  obsID = "obsID",
  nbreaks = Nbreaks, #number of partitions of choice data. Partitions will increase
incrementally the number of participants to perform power analysis simulation.
  n_q = Nchoice,
  return_models = TRUE #necessary for the summary() function below.
)
summary(power[[Nbreaks]]) #provides a summary of the last iteration of power analysis
(number corresponds with "nbreaks" above).

#repeat of the simulated power analysis in a way that "plot(power)" can be generated
(exclude "return_models = TRUE). Make sure the values are the same as above.
power <- cbc_power(
  data = rndchoices,
  pars = c("cost", "brand", "UserRating", "UAscore"),
  outcome = "choice",
  obsID = "obsID",
  nbreaks = Nbreaks, #number of partitions of choice data. Partitions will increase
incrementally in size to perform power analysis.
  n_q = Nchoice,
)
plot(power) #plots Standard Error for coefficient estimates vs Sample Size.

##### OUTPUT DESIGN #####

```

```

#save only the columns needed in the next stages
FinalDesign <- subset(FirstDesign, select = c("profileID", "qID", "altID", "cost",
"brand", "UserRating", "UAscore"))

#convert factors to characters or numbers so they can be manipulated with no error
FinalDesign$cost <- as.numeric(FinalDesign$cost)
FinalDesign$brand <- as.character(FinalDesign$brand)
FinalDesign$UserRating <- as.numeric(FinalDesign$UserRating)
FinalDesign$UAscore <- as.character(FinalDesign$UAscore)

#rename the values according to the requirements of the next stages
FinalDesign["UAscore"][FinalDesign["UAscore"] == "High"] <- "High (8/10)"
FinalDesign["UAscore"][FinalDesign["UAscore"] == "Medium"] <- "Medium (6/10)"
FinalDesign["UAscore"][FinalDesign["UAscore"] == "Low"] <- "Low (4/10)"
FinalDesign["UAscore"][FinalDesign["UAscore"] == "Unavailable"] <- "Unavailable"

#convert the characters or numbers back to factors
FinalDesign$cost <- as.factor(FinalDesign$cost)
FinalDesign$brand <- as.factor(FinalDesign$brand)
FinalDesign$UserRating <- as.factor(FinalDesign$UserRating)
FinalDesign$UAscore <- as.factor(FinalDesign$UAscore)

##### ----- #####

t <- Sys.time() #get system time and date
replace <- gsub(":", "-", t) #replace : in time with - for filename restriction
purposes
newt <- gsub(" ", "_", replace) #replace spaces with underscore
newt <- substr(newt, 1, 19) #remove the precision for seconds (only keep date and
time up until seconds)
DBerrRound <- round(DBerr, 4)
csvName <- paste("RAW (DBerr", DBerrRound, ") ", "(Seed", Seed, ") ", "V06M01design
", newt, ".csv", sep = "") #create file name the includes the time of generation
write.csv(FinalDesign, csvName, row.names = FALSE) #write the efficient design to a
csv file

```

## Appendix B: Syntax – Preparation for Survey Implementation in Python

```
import pandas as pd
import datetime as dt
import os

choiceTaskCount = 16 #number of choice sets per respondent
altCount = 4 #number of alternatives per choice set
OrgFileName = "V06M01design.csv" #name of the csv file that contains the design

OrgDF = pd.read_csv(str(os.path.dirname(__file__)) + "\\ " + OrgFileName) #create a
dataframe from the design csv file

#create a text file in the current directory with a unique name for the output based
on the design and current date and time
CurrentDateAndTime = dt.datetime.now()
DnTApprop = CurrentDateAndTime.strftime("%Y%m%d_%H%M%S")
txtFileName = "HTML Tables for " + OrgFileName.strip(".csv") + " " + str(DnTApprop) +
".txt"
txtFileAddress = str(os.path.dirname(__file__)) + "\\ " + txtFileName
txtFile = open(txtFileAddress, "a")

#this function receives attribute-values for one alternative and creates an HTML
table for that alternative for use in Qualtrics
def HTMLTableCreator(name, cost, brand, usrnt, accinf):
    HTMLTable = str("""<table align="center" border="1" cellpadding="1"
cellspacing="1"
style="height:200px;width:150px;">""")
        + "\n"
    + """
<thead>"""
        + "\n"
    +
""" <tr>"""
        + "\n"
    + """ <th scope="col" style="text-align: center; vertical-align: middle;"><span
style="color:#dddddd;"><strong>""" + "Name: BPM" + str(name) +
""" </strong></span></th>""" + "\n"
    +
""" </tr>"""
        + "\n"
```

```

+ "" ""
</thead>"" ""

+ "\n"
+ "" ""
<tbody>"" ""

+ "\n"
+
"" "" <tr>"" ""

+ "\n"
+ "" "" <td style="text-align: center; vertical-align: middle;"><span
style="color:#7f8c8d;">Cost:<strong> </strong></span><strong>"" "" + str(cost) +
"" "" </strong></td>"" "" + "\n"
+
"" "" </tr>"" ""

+ "\n"
+
"" "" <tr>"" ""

+ "\n"
+ "" "" <td style="text-align: center; vertical-align: middle;"><span
style="color:#7f8c8d;">Brand:</span> <strong>"" "" + str(brand) +
"" "" </strong></td>"" "" + "\n"
+
"" "" </tr>"" ""

+ "\n"
+
"" "" <tr>"" ""

+ "\n"
+ "" "" <td style="text-align: center; vertical-align: middle;"><span
style="color:#7f8c8d;">User Rating:</span> <strong>"" "" + str(usrrt) + "" ""
Stars</strong></td>"" "" + "\n"
+
"" "" </tr>"" ""

+ "\n"
+
"" "" <tr>"" ""

+ "\n"

```

```

+ """ <td style="text-align: center; vertical-align: middle;"><span
style="color:#7f8c8d;">Usability/Accessibility Score:</span> <strong>""" +
str(accinf) + """</strong></td>""" + "\n"
+
""" </tr>"""

+ "\n"
+ """
</tbody>"""

+ "\n"
+
"""</table>"""

+ "\n"

return(HTMLTable)

#this function receives a dataframe and creates a list of lists from it for easier
processing.
#the inner-most lists are alternatives, the next layer out are the choice tasks.
#alternative explanation of the function: list of(list of choice tasks(list of
alternatives))
def DFtoList(DF):
    DFList = []
    for task in range(choiceTaskCount):
        ChoiceTaskTemp= []
        for alt in range((int(task)*int(altCount)),
((int(task)*int(altCount))+int(altCount))): #this range function ensures that for
each task, the correct alts are being read, e.g. for task 4, alts 16 to 20
            AltTemp = []
            AltTemp.append(DF.at[alt, "profileID"])
            AltTemp.append(DF.at[alt, "cost"])
            AltTemp.append(DF.at[alt, "brand"])
            AltTemp.append(DF.at[alt, "UserRating"])
            AltTemp.append(DF.at[alt, "UAScore"])
            ChoiceTaskTemp.append(AltTemp)
        DFList.append(ChoiceTaskTemp)

    return(DFList)

#This function reads the list created from the data frame, passes each alt to
HTMLTableCreator, and formats the output in a text file for the ease of use
def OutputTextFile(InputList, OutputFile):
    for task in range(choiceTaskCount):

```

```

        OutputFile.write("////////////////////////////////Choice Task " + str(task+1) +
"////////////////////////////////")
        OutputFile.write("\n")
        for alt in range(altCount):
            OutputFile.write("-----Alternative 0" + str(alt+1) + "-----")
            OutputFile.write("\n")
            WorkingTask = InputList[task]
            WorkingAlt = WorkingTask[alt]

            #identifying which levels belong to which attributes
            name = WorkingAlt[0]
            cost = WorkingAlt[1]
            brand = WorkingAlt[2]
            usrprt = WorkingAlt[3]
            accinf = WorkingAlt[4]

            #cleaning up the presentation of the data before creating HTML table
            if int(name)<10:
                name = "0" + str(name)
                cost = """$"" + str(cost)
                brand.strip("'")
                accinf.strip("'")

            OutputFile.write(HTMLTableCreator(name, cost, brand, usrprt, accinf))
            OutputFile.write("-----")
            OutputFile.write("\n")
        OutputFile.write("//////////////////////////////////////////
//////////////////////////////////////////")
        OutputFile.write("\n\n\n")

OutputTextFile(DFtoList(OrgDF), txtFile)
txtFile.close()
print("HTML tables created successfully!\nCheck the directory for the output text
file:\n" + txtFileAddress)

```

## Appendix C: Design – D<sub>z</sub>-Efficient Design for Feasibility Study

profileID	qID	altID	cost	brand	usrtrt	accinf
3	1	1	20	HomeHealth	4.8	High (8/10)
36	1	2	20	Omron	4.8	Low (4/10)
30	1	3	20	Omron	3.2	Medium (6/10)
69	1	4	20	HomeHealth	3.2	Unavailable
19	2	1	25	Omron	4.8	Medium (6/10)
28	2	2	25	HomeHealth	4	Medium (6/10)
4	2	3	25	HomeHealth	4.8	High (8/10)
48	2	4	20	Omron	3.2	Low (4/10)
31	3	1	25	Omron	3.2	Medium (6/10)
44	3	2	30	Omron	4	Low (4/10)
16	3	3	25	HomeHealth	3.2	High (8/10)
70	3	4	25	HomeHealth	3.2	Unavailable
42	4	1	20	Omron	4	Low (4/10)
8	4	2	30	Omron	4	High (8/10)
65	4	3	30	HomeHealth	4	Unavailable
27	4	4	20	HomeHealth	4	Medium (6/10)
33	5	1	20	HomeHealth	3.2	Medium (6/10)
60	5	2	20	Omron	4	Unavailable
45	5	3	20	HomeHealth	4	Low (4/10)
67	5	4	25	Omron	3.2	Unavailable
49	6	1	25	Omron	3.2	Low (4/10)
18	6	2	20	Omron	4.8	Medium (6/10)
58	6	3	25	HomeHealth	4.8	Unavailable

profileID	qID	altID	cost	brand	usrtr	accinf
70	6	4	25	HomeHealth	3.2	Unavailable
6	7	1	20	Omron	4	High (8/10)
63	7	2	20	HomeHealth	4	Unavailable
46	7	3	25	HomeHealth	4	Low (4/10)
25	7	4	25	Omron	4	Medium (6/10)
21	8	1	20	HomeHealth	4.8	Medium (6/10)
36	8	2	20	Omron	4.8	Low (4/10)
60	8	3	20	Omron	4	Unavailable
9	8	4	20	HomeHealth	4	High (8/10)
12	9	1	20	Omron	3.2	High (8/10)
41	9	2	30	HomeHealth	4.8	Low (4/10)
2	9	3	30	Omron	4.8	High (8/10)
33	9	4	20	HomeHealth	3.2	Medium (6/10)
14	10	1	30	Omron	3.2	High (8/10)
62	10	2	30	Omron	4	Unavailable
35	10	3	30	HomeHealth	3.2	Medium (6/10)
53	10	4	30	HomeHealth	3.2	Low (4/10)
9	11	1	20	HomeHealth	4	High (8/10)
16	11	2	25	HomeHealth	3.2	High (8/10)
49	11	3	25	Omron	3.2	Low (4/10)
24	11	4	20	Omron	4	Medium (6/10)
66	12	1	20	Omron	3.2	Unavailable
1	12	2	25	Omron	4.8	High (8/10)
51	12	3	20	HomeHealth	3.2	Low (4/10)

<b>profileID</b>	<b>qID</b>	<b>altID</b>	<b>cost</b>	<b>brand</b>	<b>usrprt</b>	<b>accinf</b>
22	12	4	25	HomeHealth	4.8	Medium (6/10)
1	13	1	25	Omron	4.8	High (8/10)
23	13	2	30	HomeHealth	4.8	Medium (6/10)
56	13	3	30	Omron	4.8	Unavailable
40	13	4	25	HomeHealth	4.8	Low (4/10)
63	14	1	20	HomeHealth	4	Unavailable
35	14	2	30	HomeHealth	3.2	Medium (6/10)
14	14	3	30	Omron	3.2	High (8/10)
42	14	4	20	Omron	4	Low (4/10)
43	15	1	25	Omron	4	Low (4/10)
34	15	2	25	HomeHealth	3.2	Medium (6/10)
56	15	3	30	Omron	4.8	Unavailable
5	15	4	30	HomeHealth	4.8	High (8/10)
26	16	1	30	Omron	4	Medium (6/10)
66	16	2	20	Omron	3.2	Unavailable
47	16	3	30	HomeHealth	4	Low (4/10)
15	16	4	20	HomeHealth	3.2	High (8/10)

## Appendix D: Design – D<sub>z</sub>-Efficient Design for Phase I

profileID	qID	altID	cost	brand	usrtr	accinf
45	1	1	20	HomeHealth	4	Low (4/10)
68	1	2	30	Omron	3.2	Unavailable
35	1	3	30	HomeHealth	3.2	Medium (6/10)
6	1	4	20	Omron	4	High (8/10)
58	2	1	25	HomeHealth	4.8	Unavailable
25	2	2	25	Omron	4	Medium (6/10)
53	2	3	30	HomeHealth	3.2	Low (4/10)
14	2	4	30	Omron	3.2	High (8/10)
6	3	1	20	Omron	4	High (8/10)
37	3	2	25	Omron	4.8	Low (4/10)
58	3	3	25	HomeHealth	4.8	Unavailable
24	3	4	20	Omron	4	Medium (6/10)
23	4	1	30	HomeHealth	4.8	Medium (6/10)
13	4	2	25	Omron	3.2	High (8/10)
52	4	3	25	HomeHealth	3.2	Low (4/10)
56	4	4	30	Omron	4.8	Unavailable
27	5	1	20	HomeHealth	4	Medium (6/10)
42	5	2	20	Omron	4	Low (4/10)
63	5	3	20	HomeHealth	4	Unavailable
6	5	4	20	Omron	4	High (8/10)
8	6	1	30	Omron	4	High (8/10)
48	6	2	20	Omron	3.2	Low (4/10)
65	6	3	30	HomeHealth	4	Unavailable

profileID	qID	altID	cost	brand	usrtr	accinf
33	6	4	20	HomeHealth	3.2	Medium (6/10)
32	7	1	30	Omron	3.2	Medium (6/10)
16	7	2	25	HomeHealth	3.2	High (8/10)
52	7	3	25	HomeHealth	3.2	Low (4/10)
67	7	4	25	Omron	3.2	Unavailable
34	8	1	25	HomeHealth	3.2	Medium (6/10)
54	8	2	20	Omron	4.8	Unavailable
13	8	3	25	Omron	3.2	High (8/10)
40	8	4	25	HomeHealth	4.8	Low (4/10)
69	9	1	20	HomeHealth	3.2	Unavailable
19	9	2	25	Omron	4.8	Medium (6/10)
30	9	3	20	Omron	3.2	Medium (6/10)
41	9	4	30	HomeHealth	4.8	Low (4/10)
38	10	1	30	Omron	4.8	Low (4/10)
66	10	2	20	Omron	3.2	Unavailable
15	10	3	20	HomeHealth	3.2	High (8/10)
5	10	4	30	HomeHealth	4.8	High (8/10)
70	11	1	25	HomeHealth	3.2	Unavailable
31	11	2	25	Omron	3.2	Medium (6/10)
16	11	3	25	HomeHealth	3.2	High (8/10)
49	11	4	25	Omron	3.2	Low (4/10)
3	12	1	20	HomeHealth	4.8	High (8/10)
18	12	2	20	Omron	4.8	Medium (6/10)
53	12	3	30	HomeHealth	3.2	Low (4/10)

profileID	qID	altID	cost	brand	usrst	accinf
62	12	4	30	Omron	4	Unavailable
63	13	1	20	HomeHealth	4	Unavailable
42	13	2	20	Omron	4	Low (4/10)
3	13	3	20	HomeHealth	4.8	High (8/10)
21	13	4	20	HomeHealth	4.8	Medium (6/10)
27	14	1	20	HomeHealth	4	Medium (6/10)
36	14	2	20	Omron	4.8	Low (4/10)
10	14	3	25	HomeHealth	4	High (8/10)
61	14	4	25	Omron	4	Unavailable
63	15	1	20	HomeHealth	4	Unavailable
42	15	2	20	Omron	4	Low (4/10)
46	15	3	25	HomeHealth	4	Low (4/10)
25	15	4	25	Omron	4	Medium (6/10)
60	16	1	20	Omron	4	Unavailable
45	16	2	20	HomeHealth	4	Low (4/10)
17	16	3	30	HomeHealth	3.2	High (8/10)
32	16	4	30	Omron	3.2	Medium (6/10)

## Appendix E: Design – Db-Efficient Design for Phase II

profileID	qID	altID	cost	brand	UserRating	UAScore
63	1	1	20	Omron	4	Low (4/10)
1	1	2	20	HomeHealth	3.2	Unavailable
71	1	3	30	Omron	4.8	Low (4/10)
44	1	4	30	HomeHealth	4	Medium (6/10)
48	2	1	20	HomeHealth	4.8	Medium (6/10)
22	2	2	20	Omron	3.2	High (8/10)
13	2	3	20	HomeHealth	4.8	Unavailable
59	2	4	30	Omron	3.2	Low (4/10)
19	3	1	20	HomeHealth	3.2	High (8/10)
51	3	2	20	Omron	4.8	Medium (6/10)
56	3	3	30	HomeHealth	3.2	Low (4/10)
14	3	4	25	HomeHealth	4.8	Unavailable
64	4	1	25	Omron	4	Low (4/10)
40	4	2	25	Omron	3.2	Medium (6/10)
8	4	3	25	HomeHealth	4	Unavailable
61	4	4	25	HomeHealth	4	Low (4/10)
66	5	1	20	HomeHealth	4.8	Low (4/10)
21	5	2	30	HomeHealth	3.2	High (8/10)
45	5	3	20	Omron	4	Medium (6/10)
18	5	4	30	Omron	4.8	Unavailable
45	6	1	20	Omron	4	Medium (6/10)
28	6	2	20	Omron	4	High (8/10)
13	6	3	20	HomeHealth	4.8	Unavailable

profileID	qID	altID	cost	brand	UserRating	UAScore
42	6	4	20	HomeHealth	4	Medium (6/10)
67	7	1	25	HomeHealth	4.8	Low (4/10)
17	7	2	25	Omron	4.8	Unavailable
20	7	3	25	HomeHealth	3.2	High (8/10)
46	7	4	25	Omron	4	Medium (6/10)
20	8	1	25	HomeHealth	3.2	High (8/10)
46	8	2	25	Omron	4	Medium (6/10)
70	8	3	25	Omron	4.8	Low (4/10)
54	8	4	20	HomeHealth	3.2	Low (4/10)
53	9	1	30	Omron	4.8	Medium (6/10)
16	9	2	20	Omron	4.8	Unavailable
36	9	3	20	HomeHealth	3.2	Medium (6/10)
32	9	4	25	HomeHealth	4.8	High (8/10)
69	10	1	20	Omron	4.8	Low (4/10)
7	10	2	20	HomeHealth	4	Unavailable
36	10	3	20	HomeHealth	3.2	Medium (6/10)
58	10	4	25	Omron	3.2	Low (4/10)
15	11	1	30	HomeHealth	4.8	Unavailable
71	11	2	30	Omron	4.8	Low (4/10)
44	11	3	30	HomeHealth	4	Medium (6/10)
24	11	4	30	Omron	3.2	High (8/10)
68	12	1	30	HomeHealth	4.8	Low (4/10)
18	12	2	30	Omron	4.8	Unavailable
54	12	3	20	HomeHealth	3.2	Low (4/10)

<b>profileID</b>	<b>qID</b>	<b>altID</b>	<b>cost</b>	<b>brand</b>	<b>UserRating</b>	<b>UAScore</b>
44	12	4	30	HomeHealth	4	Medium (6/10)
49	13	1	25	HomeHealth	4.8	Medium (6/10)
23	13	2	25	Omron	3.2	High (8/10)
17	13	3	25	Omron	4.8	Unavailable
3	13	4	30	HomeHealth	3.2	Unavailable
11	14	1	25	Omron	4	Unavailable
19	14	2	20	HomeHealth	3.2	High (8/10)
45	14	3	20	Omron	4	Medium (6/10)
67	14	4	25	HomeHealth	4.8	Low (4/10)
30	15	1	30	Omron	4	High (8/10)
54	15	2	20	HomeHealth	3.2	Low (4/10)
50	15	3	30	HomeHealth	4.8	Medium (6/10)
18	15	4	30	Omron	4.8	Unavailable
66	16	1	20	HomeHealth	4.8	Low (4/10)
10	16	2	20	Omron	4	Unavailable
6	16	3	30	Omron	3.2	Unavailable
65	16	4	30	Omron	4	Low (4/10)

## Appendix F: Phase I Survey

### Consent Form

#### University of Wisconsin-Milwaukee

#### Informed Consent to Participate in Research

**Study title:** Blood Pressure Monitor Purchasing Survey (Pilot)

**Researcher[s]:** University of Wisconsin-Milwaukee (UWM) researchers including Maysam M. Ardehali and Dr. Roger O. Smith invite you to take a survey. This survey is completely voluntary. There are no negative consequences if you don't want to take it. If you start the survey, you can always change your mind and stop at any time.

**What is the purpose of this study?** To understand how you purchase and use medical devices, and what factors impact your decision-making in the purchasing process. We will use hypothetical blood pressure monitors with hypothetical information to examine your preferences.

**What will I do?** You will only need to take an online survey.

#### Risks:

- **Some questions may seem personal or upsetting.** You can skip them or quit the survey at any time.
- Any time you share information online there are risks, such as **data being hacked or intercepted**. We're using a secure system to collect this data, but we can't completely eliminate this risk.
- Any time data is collected, there is a chance it could be seen or accessed by someone who shouldn't have access to it. We're minimizing this risk in the following ways:
  - Data are anonymous. We will only have access to your Prolific ID number.
  - We'll store all electronic data on a password-protected, encrypted computer.

**Estimated number of participants:** We will collect data from about 200 individuals.

**How long will it take?** This survey should take between 20 to 45 minutes to complete.

**Costs:** None

**Compensation:** \$12 per hour, administered via Prolific.

- Only participants who qualify for the survey and agree to participate in the study will be compensated.
- To qualify for this study, participants need to be fluent in English, live in the US, have no disabilities, and be at or above the age of 50.

**Future research:** De-identified data (all identifying information removed) may be shared with other researchers. You won't be told specific details about these future research studies.

**Where will data be stored?** On the servers for the online survey software (Qualtrics) and on researchers' computers (in encrypted and de-identified format).

**How long will it be kept?** We will keep de-identified information for as long as necessary for our projects and publications.

#### Who can see my data?

- We (the researchers) will have access to your de-identified data (no names, birthdate, or address). This is so we can analyze the data and conduct the study.
- We may share our findings in publications or presentations. If we do, the results will be de-identified. If we quote you, we will use pseudonyms (fake names).
- To verify your participation and issue your payment, Prolific will have access to your survey completion records, any responses you provide on the Prolific platform for screening, as well as your ID and associated personal information. Your personal information will not be relayed to us.

**Questions about the research, complaints, or problems:** If you have any questions, comments, or concerns about this study, you may email Dr. Roger O. Smith ([smithro@uwm.edu](mailto:smithro@uwm.edu)). Please include the name of the study in the subject.

**Questions about your rights as a research participant, complaints, or problems:** Contact the UWM IRB (Institutional Review Board) at 414-662-3544 / [irbinfo@uwm.edu](mailto:irbinfo@uwm.edu).

Please print or save this screen if you want to be able to access the information later.

**IRB #: 24.010**

**IRB Approval Date: 14/07/2023**

### **Agreement to Participate**

Your participation is completely voluntary, and you can withdraw at any time.

To take this survey, you must be fluent in English, reside in the US, be at least 50 years old, and identify as not having a disability. If you meet these criteria, please check the box below to start the survey.

- **I meet the criteria and consent to participate in this study.**
- **I do NOT meet the criteria stated above or do not want to participate.**

### **Prolific ID**

What is your Prolific ID?

*Please note that this response should auto-fill with the correct ID*

### **Screener Validation**

Are you at least 50 years old?

- Yes
- No

Would you describe yourself as having a long-term health condition or a disability?

- Yes
- No

### **Soft Screening 1**

Do you, or someone you know, use a blood pressure monitor? (Please select all that apply)

- Yes, I use a blood pressure monitor
- Yes, someone in my family or friends uses a blood pressure monitor
- No, neither myself nor anyone I know use blood pressure monitors

### **Soft Screening 2**

Can you imagine needing to purchase a blood pressure monitor for either yourself, or a friend or family member?

- No
- Yes

### **Mindset**

Consider needing to purchase a blood pressure monitor for yourself, or a friend or family member.

In each question, you will be presented with 4 types of blood pressure monitors (these are hypothetical products with hypothetical features).

In each scenario, please select the blood pressure monitor you consider as the best purchase.

- **Got it!**

For each blood pressure monitor, you will have access to the following information. Please take the time to review how the product information will be presented:

**Cost:** This will be the out of pocket cost, and ranges from \$20 to \$30 (average \$25).

**Brand:** Omron (a trusted brand) or HomeHealth (a newly established brand).

**User Rating:** Average rating out of 5 stars.

**Usability/Accessibility Score:** An independent expert evaluation of how easy the device will be to use. This score may be Unavailable, or range from Low (4/10) to High (8/10).

- **Got it!**

**Practice - Dominant Alternative Check (73, 03)**

**Practice Question:** Below, 4 different blood pressure monitors are shown. Notice the difference in **Cost, Brand, User Rating,** and **Usability/Accessibility Score.**

Please compare the products and choose the blood pressure monitor you are the most likely to purchase.

	<table border="1"> <tr><td><b>Name: BPM35</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM35</b>	Cost: <b>\$30</b>	Brand: <b>HomeHealth</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	<table border="1"> <tr><td><b>Name: BPM03</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.8 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM03</b>	Cost: <b>\$20</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.8 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM73</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.8 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM73</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>4.8 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM60</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM60</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>
<b>Name: BPM35</b>																								
Cost: <b>\$30</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								
<b>Name: BPM03</b>																								
Cost: <b>\$20</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.8 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM73</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.8 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM60</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								

Your choice:

**DCE**

Consider the following products:

	<table border="1"> <tr><td><b>Name: BPM68</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM68</b>	Cost: <b>\$30</b>	Brand: <b>Omron</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>	<table border="1"> <tr><td><b>Name: BPM45</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM45</b>	Cost: <b>\$20</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM06</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM06</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM35</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM35</b>	Cost: <b>\$30</b>	Brand: <b>HomeHealth</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>
<b>Name: BPM68</b>																								
Cost: <b>\$30</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								
<b>Name: BPM45</b>																								
Cost: <b>\$20</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM06</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM35</b>																								
Cost: <b>\$30</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								

Your choice:

Consider the following products:

	<table border="1"> <tr><td><b>Name: BPM53</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM53</b>	Cost: <b>\$30</b>	Brand: <b>HomeHealth</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM58</b></td></tr> <tr><td>Cost: <b>\$25</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.8 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM58</b>	Cost: <b>\$25</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.8 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>	<table border="1"> <tr><td><b>Name: BPM14</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM14</b>	Cost: <b>\$30</b>	Brand: <b>Omron</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM25</b></td></tr> <tr><td>Cost: <b>\$25</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM25</b>	Cost: <b>\$25</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>
<b>Name: BPM53</b>																								
Cost: <b>\$30</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM58</b>																								
Cost: <b>\$25</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.8 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								
<b>Name: BPM14</b>																								
Cost: <b>\$30</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM25</b>																								
Cost: <b>\$25</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								

Your choice:

Consider the following products:

<b>Name: BPM06</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM58</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM24</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM37</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM23</b>
Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM56</b>
Cost: <b>\$30</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM13</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM52</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM27</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM63</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM42</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM06</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM33</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM08</b>
Cost: <b>\$30</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM65</b>
Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM48</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM16</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM52</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM67</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM32</b>
Cost: <b>\$30</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM34</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM54</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM40</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM13</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM69</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM41</b>
Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM30</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM19</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM66</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM15</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM05</b>
Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM38</b>
Cost: <b>\$30</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM70</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM31</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM16</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM49</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM03</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM53</b>
Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM18</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM62</b>
Cost: <b>\$30</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

<b>Name: BPM42</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM63</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

<b>Name: BPM03</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM21</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM27</b>
Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>

<b>Name: BPM10</b>
Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>

<b>Name: BPM36</b>
Cost: <b>\$20</b>
Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>

<b>Name: BPM61</b>
Cost: <b>\$25</b>
Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

	<table border="1"> <tr><td><b>Name: BPM63</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM63</b>	Cost: <b>\$20</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>	<table border="1"> <tr><td><b>Name: BPM46</b></td></tr> <tr><td>Cost: <b>\$25</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM46</b>	Cost: <b>\$25</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM42</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM42</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM25</b></td></tr> <tr><td>Cost: <b>\$25</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM25</b>	Cost: <b>\$25</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>
<b>Name: BPM63</b>																								
Cost: <b>\$20</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								
<b>Name: BPM46</b>																								
Cost: <b>\$25</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM42</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM25</b>																								
Cost: <b>\$25</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								

Your choice:

Consider the following products:

	<table border="1"> <tr><td><b>Name: BPM45</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM45</b>	Cost: <b>\$20</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM17</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM17</b>	Cost: <b>\$30</b>	Brand: <b>HomeHealth</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM60</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM60</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>	<table border="1"> <tr><td><b>Name: BPM32</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM32</b>	Cost: <b>\$30</b>	Brand: <b>Omron</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>
<b>Name: BPM45</b>																								
Cost: <b>\$20</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM17</b>																								
Cost: <b>\$30</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM60</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								
<b>Name: BPM32</b>																								
Cost: <b>\$30</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								

Your choice:

### Choice Consistency Check - Corrsp. Q15 (08-48-65-33)

Consider the following products:

	<table border="1"> <tr><td><b>Name: BPM08</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>High (8/10)</b></td></tr> </table>	<b>Name: BPM08</b>	Cost: <b>\$30</b>	Brand: <b>Omron</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>High (8/10)</b>	<table border="1"> <tr><td><b>Name: BPM48</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>Omron</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Low (4/10)</b></td></tr> </table>	<b>Name: BPM48</b>	Cost: <b>\$20</b>	Brand: <b>Omron</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	<table border="1"> <tr><td><b>Name: BPM33</b></td></tr> <tr><td>Cost: <b>\$20</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>3.2 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Medium (6/10)</b></td></tr> </table>	<b>Name: BPM33</b>	Cost: <b>\$20</b>	Brand: <b>HomeHealth</b>	User Rating: <b>3.2 Stars</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	<table border="1"> <tr><td><b>Name: BPM65</b></td></tr> <tr><td>Cost: <b>\$30</b></td></tr> <tr><td>Brand: <b>HomeHealth</b></td></tr> <tr><td>User Rating: <b>4.0 Stars</b></td></tr> <tr><td>Usability/Accessibility Score: <b>Unavailable</b></td></tr> </table>	<b>Name: BPM65</b>	Cost: <b>\$30</b>	Brand: <b>HomeHealth</b>	User Rating: <b>4.0 Stars</b>	Usability/Accessibility Score: <b>Unavailable</b>
<b>Name: BPM08</b>																								
Cost: <b>\$30</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>High (8/10)</b>																								
<b>Name: BPM48</b>																								
Cost: <b>\$20</b>																								
Brand: <b>Omron</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Low (4/10)</b>																								
<b>Name: BPM33</b>																								
Cost: <b>\$20</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>3.2 Stars</b>																								
Usability/Accessibility Score: <b>Medium (6/10)</b>																								
<b>Name: BPM65</b>																								
Cost: <b>\$30</b>																								
Brand: <b>HomeHealth</b>																								
User Rating: <b>4.0 Stars</b>																								
Usability/Accessibility Score: <b>Unavailable</b>																								

Your choice:

### Follow-up Questions

Please rank the following items from the most important in your purchasing decision (1) to the least important (4):

- Brand
- User Ratings
- Usability/Accessibility Score
- Cost

Do you currently own any medical devices at home? (e.g. Blood Pressure Monitor, Thermometer, C-PAP, etc.)

- No
- Yes (please identify)

How familiar are you with disability and accessibility concepts?

Not familiar at all      Slightly familiar      Moderately familiar      Very familiar      Extremely familiar  
 How important was the availability of Usability/Accessibility Score in your decision to purchase a product?  
 Not at all important      Slightly important      Moderately important      Very important      Extremely important

**General Demographics**

What is your gender?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

Choose one or more races below with which you identify:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Pacific Islander
- White
- Other

To what age group do you belong?

Age Group:      Below 49   50 - 54   55 - 59   60 - 64   65 - 69   70 - 74   75 - 79   80 - 84   85 or above

In which state do you currently reside?

Alabama Alaska Arizona Arkansas California Colorado Connecticut Delaware District of Columbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada New Hampshire New Jersey New Mexico New York North Carolina North Dakota Ohio Oklahoma Oregon Pennsylvania Puerto Rico Rhode Island South Carolina South Dakota Tennessee Texas Utah Vermont Virginia Washington West Virginia Wisconsin Wyoming I do not reside in the United States

What was your total household income before taxes during the past 12 months?

- Less than \$25,000
- \$25,000-\$49,999
- \$50,000-\$74,999
- \$75,000-\$99,999
- \$100,000-\$149,999
- \$150,000 or more
- Prefer not to say

**Pilot Feedback Questionnaire**

**The purpose of the following questions is to evaluate your experience with the survey to help us improve the study.**

How satisfied were you with the length of the survey and the time it took to complete?

Extremely satisfied      Somewhat satisfied      Neither satisfied nor dissatisfied      Somewhat dissatisfied      Extremely dissatisfied

How much do you agree with the following statement: "The instructions were clear and easy to understand."

Strongly disagree      Somewhat disagree      Neither agree nor disagree      Somewhat agree      Strongly agree

When you were trying to select the best product in each group, was there any other factor that you would like to have seen to help you make a better decision? (other than Cost, Brand, User Rating, and Usability/Accessibility Score)

Overall, how easy was it to understand the purpose of the study?

Extremely easy      Somewhat easy      Neither easy nor difficult      Somewhat difficult      Extremely difficult

How easy was it to assume the mindset of an individual who is in the market for purchasing a blood pressure monitor?

Extremely easy      Somewhat easy      Neither easy nor difficult      Somewhat difficult      Extremely difficult

How easy was it to compare the hypothetical products and make a final decision in each scenario?

Extremely easy      Somewhat easy      Neither easy nor difficult      Somewhat difficult      Extremely difficult

Based on what you remember from the instructions, which one of the following is the newly established brand for blood pressure monitors?

- Omron
- HomeHealth
- I don't remember

Do you have a preferred order for seeing product information relating to Cost, Brand, User Ratings, and Availability of Accessibility/Usability Information? If so, feel free to rank these items below in the order you would've liked to see them.

- Cost
- Brand
- User Ratings
- Usability/Accessibility Information

What comments do you have on your overall experience? Was there anything you would like to see improved?

**Thank You!**

**We would like to sincerely thank you for taking part in this study! Please click the button below to be redirected back to Prolific and register your submission.**

## Appendix G: Phase II Survey

### Consent Form

#### University of Wisconsin-Milwaukee

#### Informed Consent to Participate in Research

**Study title:** Blood Pressure Monitor Purchasing Survey

**Researcher[s]:** University of Wisconsin-Milwaukee (UWM) researchers including Maysam M. Ardehali and Dr. Roger O. Smith invite you to take a survey. This survey is completely voluntary. There are no negative consequences if you don't want to take it. If you start the survey, you can always change your mind and stop at any time.

**What is the purpose of this study?** To understand how you purchase and use medical devices, and what factors impact your decision-making in the purchasing process. We will use hypothetical blood pressure monitors with hypothetical information to examine your preferences.

**What will I do?** You will only need to take an online survey.

#### Risks:

- **Some questions may seem personal or upsetting.** You can skip them or quit the survey at any time.
- Any time you share information online there are risks, such as **data being hacked or intercepted**. We're using a secure system to collect this data, but we can't completely eliminate this risk.
- Any time data is collected, there is a chance it could be seen or accessed by someone who shouldn't have access to it. We're minimizing this risk in the following ways:
  - Data are anonymous. We will only have access to your Prolific ID number.
  - As with any online activity, your IP address will also be automatically recorded. Your IP address can never be used by us to determine your identity or precise location, and we will never disclose it.
  - We'll store all electronic data on a password-protected, encrypted computer.

**Estimated number of participants:** We will collect data from about 200 individuals.

**How long will it take?** This survey should take between 20 to 45 minutes to complete.

**Costs:** None

**Compensation:** \$12 per hour, administered via Prolific.

- Only participants who qualify for the survey, agree to participate, and finish the survey will be eligible for compensation.

**Future research:** De-identified data (all identifying information removed) may be shared with other researchers. You won't be told specific details about these future research studies.

**Where will data be stored?** On the servers for the online survey software (Qualtrics) and on researchers' computers (in encrypted and de-identified format).

**How long will it be kept?** We will keep de-identified information for as long as necessary for our projects and publications.

#### Who can see my data?

- We (the researchers) will have access to your de-identified data (no names, birthdate, or address). This is so we can analyze the data and conduct the study.
- We may share our findings in publications or presentations. If we do, the results will be de-identified. If we quote you, we will use pseudonyms (fake names).
- To verify your participation and issue your payment, Prolific will have access to your survey completion records, any responses you provide on the Prolific platform for screening, as well as your ID and associated personal information. Your personal information will not be relayed to us.

**Questions about the research, complaints, or problems:** If you have any questions, comments, or concerns about this study, you may email Dr. Roger O. Smith ([smithro@uwm.edu](mailto:smithro@uwm.edu)). Please include the name of the study in the subject.

**Questions about your rights as a research participant, complaints, or problems:** Contact the UWM IRB (Institutional Review Board) at 414-662-3544 / [irbinfo@uwm.edu](mailto:irbinfo@uwm.edu).

Please print or save this screen if you want to be able to access the information later.

**IRB #: 24.010**

**IRB Approval Date: 11/03/2023**

### **Agreement to Participate**

Your participation is completely voluntary, and you can withdraw at any time.

If you consent to participate in this study, please check the box below to start the survey.

- **I consent to participate in this study.**
- **I do NOT want to participate in this study.**

### **Prolific ID**

What is your Prolific ID?

*Please note that this response should auto-fill with the correct ID*

### **Commitment Check**

Our research is likely to have a positive impact on the quality of healthcare for everyone in our society and communities. By participating in this survey, you are joining us in identifying the best ways to achieve this goal. It is important that you provide honest and thoughtful answers to the questions in this survey, so we can accurately reflect your opinions and priorities.

Do you commit to reading and following the instructions, and providing honest and thoughtful responses to the questions in this survey?

- I can't make any promises
- Yes, I will!
- No, I will not!

### **Screener Validation**

Which option below best describes your age range?

- 50 years old or above
- 35 to 49 years old
- 18 to 34 years old

Would you describe yourself as having a long-term health condition or a disability?

- Yes
- No

### **Soft Screening 1**

Do you, or someone you know, use a blood pressure monitor? (Please select all that apply)

- Yes, I use a blood pressure monitor
- Yes, someone in my family or friends uses a blood pressure monitor
- No, neither myself nor anyone I know use blood pressure monitors

### **Soft Screening 2**

Can you imagine needing to purchase a blood pressure monitor for either yourself, or a friend or family member?

- No
- Yes

### **Attention Check 1**

How much do you agree with the following statement:

I live with one or more unicorns.

Strongly Disagree

Disagree

Agree

Strongly Agree

**Mindset**

Consider needing to purchase a blood pressure monitor for yourself, or a friend or family member. In each question, you will be presented with 4 types of blood pressure monitors (these are hypothetical products with hypothetical features). In each scenario, please select the blood pressure monitor you consider as the best purchase.

- **Got it!**

For each blood pressure monitor, you will have access to the following information. Please take the time to review how the product information will be presented:

**Cost:** This will be the out of pocket cost, and ranges from \$20 to \$30 (average \$25).

**Brand:** Omron (a trusted brand) or HomeHealth (a newly established brand).

**User Rating:** Average rating out of 5 stars.

**Usability/Accessibility Score:** An independent expert evaluation of how easy the device will be to use. This score may be Unavailable, or range from Low (4/10) to High (8/10).

Based on the information above, which is the newly established Brand in the blood pressure monitor market?

- Sunshine
- I don't know
- Omron
- HomeHealth

**Practice - Dominant Alternative Check (73, 31)**

**Practice Question:** Below, 4 different blood pressure monitors are shown. Notice the difference in **Cost, Brand, User Rating, and Usability/Accessibility Score.**

Please compare the products and choose the blood pressure monitor you are the most likely to purchase.

<b>Name: BPM73</b>	<b>Name: BPM11</b>	<b>Name: BPM38</b>	<b>Name: BPM31</b>
Cost: <b>\$20</b>	Cost: <b>\$25</b>	Cost: <b>\$30</b>	Cost: <b>\$20</b>
Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

**DCE**

Consider the following products:

<b>Name: BPM63</b>	<b>Name: BPM01</b>	<b>Name: BPM71</b>	<b>Name: BPM44</b>
Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>
Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM59</b>	<b>Name: BPM13</b>	<b>Name: BPM22</b>	<b>Name: BPM48</b>
Cost: <b>\$30</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>
Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>
User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM14</b>	<b>Name: BPM51</b>	<b>Name: BPM19</b>	<b>Name: BPM56</b>
Cost: <b>\$25</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM64</b>	<b>Name: BPM40</b>	<b>Name: BPM61</b>	<b>Name: BPM08</b>
Cost: <b>\$25</b>	Cost: <b>\$25</b>	Cost: <b>\$25</b>	Cost: <b>\$25</b>
Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

<b>Name: BPM18</b>	<b>Name: BPM45</b>	<b>Name: BPM21</b>	<b>Name: BPM66</b>
Cost: <b>\$30</b>	Cost: <b>\$20</b>	Cost: <b>\$30</b>	Cost: <b>\$20</b>
Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM42</b>	<b>Name: BPM13</b>	<b>Name: BPM28</b>	<b>Name: BPM45</b>
Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>4.0 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM67</b>	<b>Name: BPM20</b>	<b>Name: BPM46</b>	<b>Name: BPM17</b>
Cost: <b>\$25</b>	Cost: <b>\$25</b>	Cost: <b>\$25</b>	Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

<b>Name: BPM46</b>	<b>Name: BPM70</b>	<b>Name: BPM54</b>	<b>Name: BPM20</b>
Cost: <b>\$25</b>	Cost: <b>\$25</b>	Cost: <b>\$20</b>	Cost: <b>\$25</b>
Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.0 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM16</b>	<b>Name: BPM36</b>	<b>Name: BPM53</b>	<b>Name: BPM32</b>
Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$30</b>	Cost: <b>\$25</b>
Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM36</b>	<b>Name: BPM07</b>	<b>Name: BPM69</b>	<b>Name: BPM58</b>
Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM15</b>	<b>Name: BPM71</b>	<b>Name: BPM24</b>	<b>Name: BPM44</b>
Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM18</b>	<b>Name: BPM68</b>	<b>Name: BPM44</b>	<b>Name: BPM54</b>
Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$20</b>
Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM49</b>	<b>Name: BPM03</b>	<b>Name: BPM17</b>	<b>Name: BPM23</b>
Cost: <b>\$25</b>	Cost: <b>\$30</b>	Cost: <b>\$25</b>	Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>3.2 Stars</b>
Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>High (8/10)</b>

Your choice:

Consider the following products:

<b>Name: BPM19</b>	<b>Name: BPM67</b>	<b>Name: BPM45</b>	<b>Name: BPM11</b>
Cost: <b>\$20</b>	Cost: <b>\$25</b>	Cost: <b>\$20</b>	Cost: <b>\$25</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

<b>Name: BPM54</b>	<b>Name: BPM50</b>	<b>Name: BPM30</b>	<b>Name: BPM18</b>
Cost: <b>\$20</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>
Brand: <b>HomeHealth</b>	Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>3.2 Stars</b>	User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.8 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

Consider the following products:

<b>Name: BPM66</b>	<b>Name: BPM65</b>	<b>Name: BPM06</b>	<b>Name: BPM10</b>
Cost: <b>\$20</b>	Cost: <b>\$30</b>	Cost: <b>\$30</b>	Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>3.2 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Low (4/10)</b>	Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>Unavailable</b>

Your choice:

**Choice Consistency Check - Corrsp. Q28 (45-28-13-42)**

Consider the following products:

<b>Name: BPM13</b>	<b>Name: BPM28</b>	<b>Name: BPM45</b>	<b>Name: BPM42</b>
Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>	Cost: <b>\$20</b>
Brand: <b>HomeHealth</b>	Brand: <b>Omron</b>	Brand: <b>Omron</b>	Brand: <b>HomeHealth</b>
User Rating: <b>4.8 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.0 Stars</b>	User Rating: <b>4.0 Stars</b>
Usability/Accessibility Score: <b>Unavailable</b>	Usability/Accessibility Score: <b>High (8/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>	Usability/Accessibility Score: <b>Medium (6/10)</b>

Your choice:

### Follow-up Questions

Please rank the following items from the most important in your purchasing decision (1) to the least important (4):

- Brand
- Usability/Accessibility Score
- Cost
- User Ratings

Do you currently own any medical devices at home? (e.g. Blood Pressure Monitor, Thermometer, C-PAP, etc.)

- No
- Yes (please identify)

How familiar are you with disability and accessibility concepts?

Not familiar at all      Slightly familiar      Moderately familiar      Very familiar      Extremely familiar

How important was the availability of Usability/Accessibility Score in your decision to purchase a product?

Not at all important      Slightly important      Moderately important      Very important      Extremely important

### Attention Check 2

How much do you agree with the following statement:

All living humans need to breathe.

Strongly Disagree      Disagree      Agree      Strongly Agree

### General Demographics

What is your gender?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

Choose one or more races below with which you identify:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Pacific Islander
- White
- Other

To what age group do you belong?

Age Group:                Below 49 50 - 54 55 - 59 60 - 64 65 - 69 70 - 74 75 - 79 80 - 84 85 or above

In which state do you currently reside?

Alabama Alaska Arizona Arkansas California Colorado Connecticut Delaware District of Columbia Florida Georgia Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada New Hampshire New Jersey New Mexico New York North Carolina North Dakota Ohio Oklahoma Oregon Pennsylvania Puerto Rico Rhode Island South Carolina South Dakota Tennessee Texas Utah Vermont Virginia Washington West Virginia Wisconsin Wyoming I do not reside in the United States

What was your total household income before taxes during the past 12 months?

- Less than \$25,000
- \$25,000-\$49,999
- \$50,000-\$74,999
- \$75,000-\$99,999
- \$100,000-\$149,999
- \$150,000 or more
- Prefer not to say

**Effort Check**

In 2 or more sentences, please describe what you think the purpose of this study is:

**Feedback Questionnaire**

How satisfied were you with the length of the survey and the time it took to complete?

Extremely satisfied    Somewhat satisfied    Neither satisfied nor dissatisfied    Somewhat dissatisfied    Extremely dissatisfied

How much do you agree with the following statement: "The instructions were clear and easy to understand."

Strongly disagree    Somewhat disagree    Neither agree nor disagree    Somewhat agree    Strongly agree

When you were trying to select the best product in each group, was there any other factor that you would like to have seen to help you make a better decision? (other than Cost, Brand, User Rating, and Usability/Accessibility Score)

Overall, how easy was it to understand the purpose of the study?

Extremely easy    Somewhat easy    Neither easy nor difficult    Somewhat difficult    Extremely difficult

How easy was it to assume the mindset of an individual who is in the market for purchasing a blood pressure monitor?

Extremely easy    Somewhat easy    Neither easy nor difficult    Somewhat difficult    Extremely difficult

How easy was it to compare the hypothetical products and make a final decision in each scenario?

Extremely easy    Somewhat easy    Neither easy nor difficult    Somewhat difficult    Extremely difficult

What other thoughts, comments, or feedback do you have for us?

**Thank You!**

**We would like to sincerely thank you for taking part in this study! Please click the button below to be redirected back to Prolific and register your submission.**

## Appendix H: Syntax – Fraud Detection in Python

```
import plotly.express as px
import plotly
import pandas as pd
import statistics as st
import ipinfo
import datetime as dt
import termcolor as tc
import numpy as np
import time
import os
from itertools import zip_longest
from prettytable import PrettyTable
from geopy.geocoders import Nominatim
from openai import OpenAI

#----- USER INPUT -----
SurveyData = "FraudCheck(Main-DESV06M01-11082023-0816AM).csv" # Qualtrics survey
results to be checked, CSV MUST BE IN UTF-8 FORMAT
ProlificDemog = "ProlificDemographicsReport(Main-DESV06M01-11082023-0816AM).csv" #
Prolific demographics for cross-referencing
Identifier = "Main-DESV06M01-11082023-0816AM" # Unique ID to specify the I/O
documents associated with this analysis

BypassIPinfoAPI = "y" # Bypass using {ipinfo} API for Location validation? y for yes,
anything else for no. Saves API access quota if not used.
BypassOpenAIAPI = "y" #Bypass using {OpenAI} API to analyze sentences? y for yes,
which means no report will be generated.

#===== QUALTRICS REPORT:
Q_ColNameChangeDict = {} # Will contain Qualtrics report column names that need to be
changed internally, key-values added below as necessary

# Column names in Qualtrics report for metrics of interest in fraud detection. Best
to copy the entire cell contents from the report.
# The items below are also a good overview of what factors were considered in
comprehensive fraud detection.
# If a column name needs to be changed internally, "Q_ColNameChangeDict.update()" is
used. Otherwise, only the item of interest is identified.

# >> Used previously in Phase I fraud detection
Q_ColNameChangeDict.update(IPAddress = "IPAddress") # Metadata
Q_ColNameChangeDict.update(ProlificID = "PROLIFIC_PID") # Metadata
```

```

Q_ColNameChangeDict.update(AgeChk = "Q8") # Screener validation for age
Q_ColNameChangeDict.update(DisChk = "Q9") # Screener validation for disability
Q_ColNameChangeDict.update(Ethnicity = "Q51") # Demographics section
Q_ColNameChangeDict.update(AgeRange = "Q52_1") # Demographics section
Q_ColNameChangeDict.update(State = "Q53") # Demographics section
# --- Dependent item checked based on the data above: Multiple Submissions from one
participant
# --- Dependent item checked based on the data above: Entries from participants who
failed screeners will be ignored

# >> New in Phase II fraud detection
Q_ColNameChangeDict.update(TotalDuration = "Duration (in seconds)") # Metadata
Q_ColNameChangeDict.update(HumanChance = "Q_RecaptchaScore") # Metadata used for bot
detection
Q_ColNameChangeDict.update(CommitmentCheck = "Q6") # Response to asking for
commitment to providing honest answers
Q_ColNameChangeDict.update(MindsetAssumption = "Q13") # Response to asking whether
one can assume the mindset for the study
Q_ColNameChangeDict.update(AttentionCheck1 = "Q15") # Attention check question with
obviously negative answer
Q_ColNameChangeDict.update(AttentionCheck2 = "Q48") # Attention check question with
obviously positive answer
Q_ColNameChangeDict.update(InstructionsComprehensionCheck = "Q20") # Answer to
question on instructions immediately visible above
Q_ColNameChangeDict.update(Sex = "Q50") # Demographics section, but survey asked for
"gender". Discrepancy in cross-validation is possible.
Q_ColNameChangeDict.update(StudyPurpose = "Q56") # Answer to the question asking for
2 sentences on the perceived study purpose. Character counts will be compared, not
content.
PageSubmissionTimesList = ["Q1_Page Submit", "Q3_Page Submit", "Q5_Page Submit",
"Q7_Page Submit",
                        "Q10_Page Submit", "Q12_Page Submit", "Q14_Page Submit",
"Q16_Page Submit",
                        "Q18_Page Submit", "Q21_Page Submit", "Q39_Page Submit",
"Q41_Page Submit",
                        "Q43_Page Submit", "Q47_Page Submit", "Q49_Page Submit",
"Q55_Page Submit",
                        "Q57_Page Submit", "Q61_Page Submit", "Q65_Page Submit",
"Q67_Page Submit"
                        ] # Metadata from timing feature
PageClickCountsList = ["Q1_Click Count", "Q3_Click Count", "Q5_Click Count",
"Q7_Click Count",
                        "Q10_Click Count", "Q12_Click Count", "Q14_Click Count",
"Q16_Click Count",

```

```

        "Q18_Click Count", "Q21_Click Count", "Q39_Click Count",
"Q41_Click Count",
        "Q43_Click Count", "Q47_Click Count", "Q49_Click Count",
"Q55_Click Count",
        "Q57_Click Count", "Q61_Click Count", "Q65_Click Count",
"Q67_Click Count"
    ] # Metadata from timing feature
TimeOnDCEInstructions = ["Q16_Page Submit", "Q18_Page Submit"] # Time spent on each
of the 2 instrcution pages
TimeOnDominantAlternativeCheck = "Q21_Page Submit" # Time spent on a single choice
task (dominant alternative)
TimeOnChoiceConsistencyCheck = "Q39_Page Submit" # Time spent on a single choice task
(choice consistency)

# --- Dependent item checked based on the data above: Approx. DCE time
# --- Dependent item checked based on the data above: Entries from participants who
failed screeners will be removed

#===== PROLIFIC REPORT:
P_ColNameChangeDict = {} # Will contain Prolific report column names that need to be
changed internally, key-values added below as necessary

# Column names in Prolific Demographics Report that need to be cross-referenced, and
their updated column names in the rest of the code

P_ColNameChangeDict.update(ProlificID = "Participant id")
P_ColNameChangeDict.update(DisChk = "Long-term health condition/disability")
P_ColNameChangeDict.update(Age = "Age")
P_ColNameChangeDict.update(Ethnicity = "Ethnicity simplified")
P_ColNameChangeDict.update(Sex = "Sex")
#-----

#----- FILE INITIALIZATION -----
SurveyDataDF = pd.read_csv(str(os.path.dirname(__file__)) + "\\\" + SurveyData) #
Reading Qualtrics data into a dataframe
ProlificDemogDF = pd.read_csv(str(os.path.dirname(__file__)) + "\\\" + ProlificDemog)
# Reading Prolific data into a dataframe

# Swapping key-values because that's how renaming the columns works
Q_ColNameChangeDict = {v: k for k, v in Q_ColNameChangeDict.items()}
P_ColNameChangeDict = {v: k for k, v in P_ColNameChangeDict.items()}

# Renaming the dataframes internally for the rest of the program
SurveyDataDF = SurveyDataDF.rename(columns=Q_ColNameChangeDict)
ProlificDemogDF = ProlificDemogDF.rename(columns=P_ColNameChangeDict)

```

```

SurveyDataDF.drop([0], axis=0, inplace=True) # Delete the second row of the survey
report internally because it contains the actual text of the questions

# Recording current date and time in an appropriate format for filename, to be used
later
CurrentDateAndTime = dt.datetime.now()
DnTApprop = CurrentDateAndTime.strftime("%Y%m%d_%H%M%S")

# Creates a unique Results folder in the parent folder containing the code to store
all results and avoid clutter in the parent directory
ResultsDirectoryName = "Results - " + str(Identifier) + " - " + str(DnTApprop)
ResultsDirectoryPath = str(os.path.dirname(__file__)) + "\\ " + ResultsDirectoryName
os.mkdir(ResultsDirectoryPath)

### If Bypass is set to yes (y): IP validation for country and state is bypassed, and
Latitude and Longitude info from Qualtrics is
# used (which is based on IP anyway). For Lat/Long based geolocation, "geopy" library
is used. This preserves the ipinfo API access quota.
# If and only if Bypass is NOT "y" the ipinfo API Access Token will be read from a
text file in the root folder, and the IP Handler will be initialized.
if BypassIPinfoAPI != "y":
    KeyLocation = str(os.path.dirname(__file__)) + "\\ " + "ipinfo-AccessToken.txt"
    Key = open(KeyLocation, 'r')
    AccessToken = Key.readline()
    Key.close()
    handler = ipinfo.getHandler(AccessToken) # Initializing the handler for
operations on IP later in the code
else:
    geoloc = Nominatim(user_agent="bas-frd-det") # Initializing the handler for
geolocation operations later

if BypassOpenAIAPI != "y":
    KeyLocation = str(os.path.dirname(__file__)) + "\\ " + "openai-apikey.txt"
    Key = open(KeyLocation, 'r')
    AccessToken = Key.readline()
    Key.close()
    client = OpenAI(api_key=AccessToken) # Initializing the OpenAI handler
#-----
#----- STANDARDIZING CONSOLE OUTPUT -----
# Standardizes the aesthetics of console output reports
# Reports a header for the report, highlights the header based on whether violations
exist, and reports a message under the header
def Reporter(Header, HeaderHighLight, Message):
    Header = str(Header) + ":"

```

```

HeaderHighlight = str(HeaderHighlight)
tc.cprint(Header, "black", HeaderHighlight, attrs=["bold"])
for line in Message:
    print("\t" + str(line))
print("\n")

# Determines if a heading in the console output should be highlighted yellow
(warning) or green (all clear)
# If violations are reported to exist (Bool = True), the highlight will be yellow,
otherwise green
def Highlight(Bool):
    if Bool:
        HeaderHighlight = "on_yellow"
    else:
        HeaderHighlight = "on_green"
    return HeaderHighlight

#-----
#----- BASIC DATA COMPREHENSION -----
# Finds the indices for specific value(s) in a dataframe column
def IndicesFinder(Dataframe, Column, ListOfValues):
    Indices = []
    for i in ListOfValues:
        index = list(Dataframe[Dataframe[str(Column)]==i].index.values)
        for v in index:
            Indices.append(v)
    return Indices

# Alphabetically sorts a dictionary based on keys
def SortDict(Dict):
    Keys = list(Dict.keys())
    Keys.sort()
    SortedDict = {k: Dict[k] for k in Keys}
    return dict(SortedDict)

# Converts a string such as "55-59" to a range comprehensible for the code, i.e. (55,
60)
def StringToRange(String):
    String = str(String)
    String.strip()
    Values = String.split("-")
    Start = int(Values[0])
    End = int(Values[1]) + 1
    Range = (Start, End)
    return Range

```

```

# If IP Geolocation is being bypassed, this function will be called upon to translate
latitude and longitude into location
def GeoLoc(Lat, Long):
    lat = str(lat)
    long = str(long)
    loc = geoloc.reverse(lat+", "+long)
    locRaw = loc.raw["address"]
    time.sleep(0.25) # Delay to avoid triggering API limits
    return locRaw

# Identifies outliers based on IQR. Values < Q1 - 1.5*IQR or > Q3 + 1.5*IQR are
considered outliers.
# Used when data from multiple columns need to be analyzed.
# Inputs: Dataframe where the data is located, List of the names of columns that
contain values to be assessed, String of column name which includes identifiers
# Output: {"ID": [Outlying values for that ID from identified columns]}
def OutliersByIQR_Multi(SourceDataframe, ColumnNamesList, IDColumnName):
    IDs = list(SourceDataframe[IDColumnName])
    IndividualDataList = [list(SourceDataframe.loc[idx, ColumnNamesList]) for idx in
SourceDataframe.index]
    IDsDataDict = {k:v for k, v in zip(IDs, IndividualDataList)}

    ColumnsList = [list(SourceDataframe[col]) for col in ColumnNamesList]
    Q1s = [pd.Series(list(map(float, ValuesList))).quantile(0.25) for ValuesList in
ColumnsList]
    Q3s = [pd.Series(list(map(float, ValuesList))).quantile(0.75) for ValuesList in
ColumnsList]
    IQRs = [Q3 - Q1 for Q3, Q1 in zip(Q3s, Q1s)]

    OutliersKeyValues = {}
    TempList = []
    for key in IDsDataDict.keys():
        for i in range(len(IDsDataDict[key])):
            Value = float(IDsDataDict[key][i])
            if Value < Q1s[i] - 1.5*IQRs[i]:
                TempList.append(str(ColumnNamesList[i]) + " " + str(round(Value, 2))
+ " <<< " + str(round(Q1s[i], 2)))
            if Value > Q3s[i] + 1.5*IQRs[i]:
                TempList.append(str(ColumnNamesList[i]) + " " + str(round(Value, 2))
+ " >>> " + str(round(Q3s[i], 2)))

        if len(TempList) > 0:
            OutliersKeyValues[key] = TempList

```

```

    TempList = []

    return OutliersKeyValues

# An adaptation of OutliersByIQR_Multi() to work with only one column for data
# Identifies outliers based on IQR. Values < Q1 - 1.5*IQR or > Q3 + 1.5*IQR are
considered outliers.
# Used when data from just one column need to be analyzed.
# Inputs: Dataframe where the data is located, Name of the column that contain values
to be assessed, String of column name which includes with identifiers
# Output: {"ID":[Outlying values for that ID from identified columns]}
def OutliersByIQR_Single(SourceDataframe, DataColumnName, IDColumnName):
    IDs = list(SourceDataframe[IDColumnName])
    IndividualDataList = [SourceDataframe.loc[idx, DataColumnName] for idx in
SourceDataframe.index]
    IDsDataDict = {k:[v] for k, v in zip(IDs, IndividualDataList)}

    ColumnsList = [list(SourceDataframe[DataColumnName])]
    Q1s = [pd.Series(list(map(float, ValuesList))).quantile(0.25) for ValuesList in
ColumnsList]
    Q3s = [pd.Series(list(map(float, ValuesList))).quantile(0.75) for ValuesList in
ColumnsList]
    IQRs = [Q3 - Q1 for Q3, Q1 in zip(Q3s, Q1s)]

    OutliersKeyValues = {}
    TempList = []
    for key in IDsDataDict.keys():
        for i in range(len(IDsDataDict[key])):
            Value = float(IDsDataDict[key][i])
            if Value < Q1s[i] - 1.5*IQRs[i]:
                TempList.append(str(DataColumnName) + " " + str(round(Value, 2)) + "
<<< " + str(round(Q1s[i], 2)))
            if Value > Q3s[i] + 1.5*IQRs[i]:
                TempList.append(str(DataColumnName) + " " + str(round(Value, 2)) + "
>>> " + str(round(Q3s[i], 2)))

        if len(TempList) > 0:
            OutliersKeyValues[key] = TempList

    TempList = []

    return OutliersKeyValues

#-----
#----- REMOVE: MULTIPLE ENTRIES -----

```

```

# Identifies if some respondents have more than one entry associated
# Reports whether multiples exist, and for which respondents
def MultipleEntryCheck(Q_Dataframe):
    MultiplesExist = bool(len(Q_Dataframe["ProlificID"]) !=
len(set(Q_Dataframe["ProlificID"])))
    if MultiplesExist == True:
        DuplicatesPID = [p for p in list(Q_Dataframe["ProlificID"]) if
list(Q_Dataframe["ProlificID"]).count(p) > 1]
        DuplicatesPID = list(set(DuplicatesPID))
    else:
        DuplicatesPID = []
    return MultiplesExist, DuplicatesPID

# This block is executed outside the function above because multiple entries from the
# same respondent is a major violation and
# all their entries need to be disqualified and excluded. The rest of the code after
# this block depends on unique entries.
MultiplesExist, DuplicatesPID = MultipleEntryCheck(SurveyDataDF)
MultipleEntryHeader = "Multiple Entries Report"
MultipleEntryHeaderHighlight = Highlight(MultiplesExist)
MultipleEntryMessage = []
if MultiplesExist:
    MultipleEntryMessage.append("Multiple entries detected from the following
respondents:")
    for i in DuplicatesPID:
        i = str(i)
        MultipleEntryMessage.append(i)
    MultipleEntryMessage.append("\n\tAll entries from the above respondents will be
ignored going forward.")
    Q_DuplicateIndices = IndicesFinder(SurveyDataDF, "ProlificID", DuplicatesPID)
    P_DuplicateIndices = IndicesFinder(ProlificDemogDF, "ProlificID", DuplicatesPID)
    SurveyDataDF.drop(Q_DuplicateIndices, inplace=True)
    ProlificDemogDF.drop(P_DuplicateIndices, inplace=True)
else:
    MultipleEntryMessage.append("No multiple entries detected.")

Reporter(MultipleEntryHeader, MultipleEntryHeaderHighlight, MultipleEntryMessage)
#-----

#----- REMOVE: DNF -----
### Removes respondents who did not finish the survey from both Prolific and
Qualtrics data. This could
# either be because they failed screening, or they ended the session without reaching
the end of survey

```

```

# and returning to Prolific automatically with the right "Completion Code"
(C1B97WAI). This essentially
# keeps only the "complete" entries in the both datasets.
def CheckDNF(P_Dataframe):
    P_PidCompletionCode = {k:v for k, v in zip(List(P_Dataframe["ProlificID"]),
P_Dataframe["Completion code"])}
    P_PidDNF = [i for i in P_PidCompletionCode.keys() if str(P_PidCompletionCode[i])
!= "C1B97WAI"]

    Header = "DNF Report"
    HeaderHighlight = Highlight(len(P_PidDNF) > 0)
    Message = []
    if len(P_PidDNF) > 0:
        Message.append("The following respondents did not finish the survey and were
removed from the analysis:")
        for pid in P_PidDNF:
            Message.append(str(pid))
        Message.append("\n\tTotal Removed: " + str(len(P_PidDNF)))
    else:
        Message.append("All participants finished the survey.")

    Reporter(Header, HeaderHighlight, Message)

    return P_PidDNF

ListOfDNF = CheckDNF(ProlificDemogDF)
if len(ListOfDNF) > 0:
    Q_IndicesDNF = IndicesFinder(SurveyDataDF, "ProlificID", ListOfDNF)
    P_IndicesDNF = IndicesFinder(ProlificDemogDF, "ProlificID", ListOfDNF)
    SurveyDataDF.drop(Q_IndicesDNF, inplace=True)
    ProlificDemogDF.drop(P_IndicesDNF, inplace=True)

NumberOfReminingParticipants = len(SurveyDataDF)
tc.cprint("      Participants Remaining in the Analysis: " +
str(NumberOfReminingParticipants) + "      \n", "black", "on_white", attrs=["bold"])
#-----

#----- DATA CROSS-VALIDATION -----
### Each validation function needs to output: 1) a LIST of offending participant
PIDs, 2) a DICTIONARY of {PID: [Stated Value, Observed Value]}. Usually if the
validation is
### taking place between Prolific info and Qualtrics data, Stated Value is what is
reported on prolific, and Observed Value is what is reported on Qualtrics. Or in the
case of

```

```

### IP-State validation, Stated is what is reported on Qualtrics, and Observed is
what is extracted from IP information.

# Determines if the age reported to Prolific is consistent with the age range
reported in Qualtrics survey
# If violations exist, reports the Prolific ID of violators, their entries on
Prolific, and their responses on Qualtrics
def AgeCrossValidation(P_Dataframe, Q_Dataframe):
    P_PidAgeDict = {P_Pid: P_age for P_Pid, P_age in
                    zip(list(P_Dataframe["ProlificID"]),
                        list(P_Dataframe["Age"]))
                    }

    Q_PidAgerangeDict = {Q_Pid: Q_agerange for Q_Pid, Q_agerange in
                         zip(list(Q_Dataframe["ProlificID"]),
                             list(Q_Dataframe["AgeRange"]))
                         }

    UntouchedPidAgerange = {k:v for k, v in Q_PidAgerangeDict.items()} # Some of the
values need to change as seen below, this reports the original responses instead.

    for key in Q_PidAgerangeDict.keys():
        if Q_PidAgerangeDict[key] == "85 or above":
            Q_PidAgerangeDict[key] = "85 - 99"
        if Q_PidAgerangeDict[key] == "Below 49":
            Q_PidAgerangeDict[key] = "0 - 49"
        if Q_PidAgerangeDict[key] is np.nan:
            Q_PidAgerangeDict[key] = "0 - 0"

    Q_PidConvertedRange = {}
    for key in Q_PidAgerangeDict.keys():
        Q_PidConvertedRange[key] = StringToRange(Q_PidAgerangeDict[key])

    PidAgeAgeRange = {}
    TempList = []
    for Pid in P_PidAgeDict.keys():
        Pid = str(Pid)
        TempList.append(P_PidAgeDict[Pid])
        TempList.append(Q_PidConvertedRange[Pid])
        PidAgeAgeRange[Pid] = TempList
        TempList = []

    CrossValidationResults = {}
    for Pid in PidAgeAgeRange:
        Pid = str(Pid)

```

```

        Start, End = PidAgeAgeRange[Pid][1]
        CrossValidationResults[Pid] = int(PidAgeAgeRange[Pid][0]) in range(Start,
End)

    FailedAgeChk = [Pid for Pid in CrossValidationResults.keys() if
CrossValidationResults[Pid] == False]

    OutputPidAgeAgerange = {Pid:[P_PidAgeDict[Pid], UntouchedPidAgerange[Pid]] for
Pid in FailedAgeChk}

    for pid in OutputPidAgeAgerange.keys():
        if OutputPidAgeAgerange[pid][1] is np.nan:
            OutputPidAgeAgerange[pid][1] = "UNDISCLOSED"

    Header = "Age Cross Validation Report"
    HeaderHighlight = Highlight(len(FailedAgeChk) > 0)
    Message = []
    if len(FailedAgeChk) > 0:
        Message.append("The following respondents have failed age cross validation
between Prolific demographics report and Qualtrics survey results:")
        for i in FailedAgeChk:
            p = str(i) + "\t" + str(OutputPidAgeAgerange[i])
            Message.append(p)
    else:
        Message.append("All participants passed age cross validation between Prolific
demographics report and Qualtrics survey results.")

    Reporter(Header, HeaderHighlight, Message)

    return FailedAgeChk, OutputPidAgeAgerange

# Determines if the disability status reported to Prolific matches the input in
Qualtrics
# If violations exist, reports the Prolific ID of violators, their entries on
Prolific, and their responses on Qualtrics
def DisabilityCrossValidation(P_Dataframe, Q_Dataframe):
    P_PidDisDict = {P_Pid: P_Dis for P_Pid, P_Dis in
                    zip(List(P_Dataframe["ProlificID"]),
                        List(P_Dataframe["DisChk"]))
                    }

    Q_PidDisDict = {Q_Pid: Q_Dis for Q_Pid, Q_Dis in
                    zip(List(Q_Dataframe["ProlificID"]),
                        List(Q_Dataframe["DisChk"]))
                    }

```

```

PidDis = {}
TempList = []
for Pid in P_PidDisDict.keys():
    Pid = str(Pid)
    TempList.append(P_PidDisDict[Pid])
    TempList.append(Q_PidDisDict[Pid])
    PidDis[Pid] = TempList
    TempList = []

FailedDisChk = [Pid for Pid in PidDis.keys() if PidDis[Pid][0] != PidDis[Pid][1]]

Header = "Disability Status Cross Validation Report"
HeaderHighlight = Highlight(len(FailedDisChk) > 0)
Message = []
if len(FailedDisChk) > 0:
    Message.append("The following respondents have failed disability status cross
validation between Prolific demographics report and Qualtrics survey results:")
    for i in FailedDisChk:
        p = str(i) + "\t" + str(PidDis[i])
        Message.append(p)
else:
    Message.append("All participants passed disability status cross validation
between Prolific demographics report and Qualtrics survey results.")

Reporter(Header, HeaderHighlight, Message)

return FailedDisChk, PidDis

# Determines if the ethnicity reported on Prolific matches the ethnicity identified
in Qualtrics
# Due to mismatch between the list of ethnicities on Prolific vs Qualtrics, Qualtrics
ethnicities are translated to match
# Failing this can be potentially harmless, mostly because of the mismatch between
the two platforms. Nonetheless, failures need to
# be checked to ensure there is no substantial violation.
def EthnicityCrossValidation(P_Dataframe, Q_Dataframe):
    P_PidEthDict = {P_Pid: P_Eth for P_Pid, P_Eth in
                    zip(list(P_Dataframe["ProlificID"]),
                        list(P_Dataframe["Ethnicity"]))
                    }

    Q_PidEthDict = {Q_Pid: Q_Eth for Q_Pid, Q_Eth in
                    zip(list(Q_Dataframe["ProlificID"]),
                        list(Q_Dataframe["Ethnicity"]))
                    }

```

```

    }

EthDict = {
    "White": "White",
    "Black or African American": "Black",
    "Asian": "Asian",
    "American Indian or Alaska Native": "Other",
    "Native Hawaiian or Pacific Islander": "Other",
    "Other": "Other"
}

for Pid in P_PidEthDict.keys():
    if Q_PidEthDict[Pid] in EthDict.keys():
        Q_PidEthDict[Pid] = EthDict[Q_PidEthDict[Pid]]
    else:
        Q_PidEthDict[Pid] = "Mixed"

PidEth = {}
TempList = []
for Pid in P_PidEthDict.keys():
    Pid = str(Pid)
    TempList.append(P_PidEthDict[Pid])
    TempList.append(Q_PidEthDict[Pid])
    PidEth[Pid] = TempList
    TempList = []

FailedEthChk = [Pid for Pid in PidEth.keys() if PidEth[Pid][0] != PidEth[Pid][1]]

Header = "Ethnicity Cross Validation Report"
HeaderHighlight = Highlight(len(FailedEthChk) > 0)
Message = []
if len(FailedEthChk) > 0:
    Message.append("The following respondents have failed ethnicity cross
validation between Prolific demographics report and Qualtrics survey results:")
    for i in FailedEthChk:
        p = str(i) + "\t" + str(PidEth[i])
        Message.append(p)
else:
    Message.append("All participants passed ethnicity cross validation between
Prolific demographics report and Qualtrics survey results.")

Reporter(Header, HeaderHighlight, Message)

return FailedEthChk, PidEth

```

```

# Determines if the IP recorded by Qualtrics is in the US, and reports the violators
and their country based on IP
def IPinUS(P_Dataframe, Q_Dataframe):
    P_PidList = List(P_Dataframe["ProlificID"])

    if BypassIPinfoAPI == "y":
        tc.cprint("IP-Based geolocation is bypassed and instead, Latitude and
Longitude data from Qualtrics is being used.",
                "black", "on_red", attrs=["bold"])

        Q_PidLatLongDict = {Q_Pid: [Lat, Long] for Q_Pid, Lat, Long in
                            zip(List(Q_Dataframe["ProlificID"]),
                                List(Q_Dataframe["LocationLatitude"]),
                                List(Q_Dataframe["LocationLongitude"]))
                            }

        QualifyingLocs = {k: v for (k, v) in Q_PidLatLongDict.items() if k in
P_PidList}

        ipCountry = {}
        for Pid in P_PidList:
            Lat, Long = QualifyingLocs[Pid]
            Location = GeoLoc(Lat, Long)
            Country = str(Location["country_code"])
            Country = Country.upper()
            ipCountry[Pid] = Country
        else:
            Q_PidIPDict = {Q_Pid: Q_ip for Q_Pid, Q_ip in
                            zip(List(Q_Dataframe["ProlificID"]),
                                List(Q_Dataframe["IPAddress"]))
                            }

            ipCountry = {}
            for Pid in P_PidList:
                ipAddress = Q_PidIPDict[Pid]
                ipDetails = handler.getDetails(ipAddress)
                Country = str(ipDetails.country)
                ipCountry[Pid] = Country

        FailedUSChk = [Pid for Pid in ipCountry.keys() if ipCountry[Pid] != "US"]

        Header = "US-Based Respondents Validation Report"
        HeaderHighlight = Highlight(len(FailedUSChk) > 0)
        Message = []
        if len(FailedUSChk) > 0:

```

```

        Message.append("The following respondents may not be from the US, based on
their IPs:")
        for i in FailedUSChk:
            p = str(i) + "\t" + str(ipCountry[i])
            Message.append(p)
        else:
            Message.append("All participants appear to be from the US, based on their
IPs.")

Reporter(Header, HeaderHighlight, Message)

return FailedUSChk, ipCountry

# Determines if the State of residence identified in Qualtrics, matches the State
extracted from IP recorded in Qualtrics
# Reports the violators, what they identified as their State, and what their IP shows
to be their State
def IPStateValidation(Q_Dataframe):
    Q_PidStateDict = {Q_Pid: Q_ip for Q_Pid, Q_ip in
                      zip(list(Q_Dataframe["ProlificID"]),
                          list(Q_Dataframe["State"]))
                      }

    for pid in Q_PidStateDict:
        if Q_PidStateDict[pid] is np.nan: Q_PidStateDict[pid] = "UNDISCLOSED"

    if BypassIPinfoAPI == "y":
        tc.cprint("IP-Based geolocation is bypassed and instead, Latitude and
Longitude data from Qualtrics is being used.",
                  "black", "on_red", attrs=["bold"])

    Q_PidLatLongDict = {Q_Pid: [Lat, Long] for Q_Pid, Lat, Long in
                        zip(list(Q_Dataframe["ProlificID"]),
                            list(Q_Dataframe["LocationLatitude"]),
                            list(Q_Dataframe["LocationLongitude"]))
                        }

    PidIPState = {}
    TempList = []
    for Pid in Q_PidStateDict.keys():
        Pid = str(Pid)
        TempList.append(Q_PidStateDict[Pid])

        Lat, Long = Q_PidLatLongDict[Pid]
        Location = GeoLoc(Lat, Long)

```

```

        State = str(Location["state"])
        TempList.append(State)

        PidIPState[Pid] = TempList
        TempList = []

    else:
        Q_PidIPDict = {Q_Pid: Q_ip for Q_Pid, Q_ip in
                       zip(List(Q_Dataframe["ProlificID"]),
                           List(Q_Dataframe["IPAddress"]))
                       }

        PidIPState = {}
        TempList = []
        for Pid in Q_PidStateDict.keys():
            Pid = str(Pid)
            TempList.append(Q_PidStateDict[Pid])

            ipAddress = Q_PidIPDict[Pid]
            ipDetails = handler.getDetails(ipAddress)
            IPState = str(ipDetails.region)
            TempList.append(IPState)

            PidIPState[Pid] = TempList
            TempList = []

    FailedIPStateChk = [Pid for Pid in PidIPState.keys() if PidIPState[Pid][0] !=
                        PidIPState[Pid][1]]

    Header = "State and IP Validation Report"
    HeaderHighlight = Highlight(len(FailedIPStateChk) > 0)
    Message = []
    if len(FailedIPStateChk) > 0:
        Message.append("The following respondents have a mismatch between their
reported State and IP:")
        for i in FailedIPStateChk:
            p = str(i) + "\t" + str(PidIPState[i])
            Message.append(p)
    else:
        Message.append("All participant IPs match their self-reported State.")

    Reporter(Header, HeaderHighlight, Message)

    return FailedIPStateChk, PidIPState

```

```

#===== NEW IN PHASE II:

# Determines if there are significant outliers in the total amount of time spent on
the survey. Only those who spent less time are
# considered. The rejection criteria from Prolific is used (3 SD below Mean), since
this is grounds for rejection on Prolific.
def TotalDurationOutliers(Q_Dataframe):
    Q_PidTotalDurationDict = {Q_Pid: Q_TotalDuration for Q_Pid, Q_TotalDuration in
        zip(List(Q_Dataframe["ProlificID"]),
            List(Q_Dataframe["TotalDuration"]))
        } # Creates a dictionary of all PIDs and the Total
Duration of their respective surveys

    # Creates a list of integers of Total Durations to calculate the threshold for
rejection (Mean - 3*SD)
    TotalDurationsList = List(Q_PidTotalDurationDict.values())
    TotalDurationsList = [int(i) for i in TotalDurationsList]
    AverageTime = st.mean(TotalDurationsList)
    StDev = st.stdev(TotalDurationsList)
    Threshold = AverageTime - 3*StDev

    # Identifies PIDs that failed the requirements in a list
    FailedPids = [Pid for Pid in Q_PidTotalDurationDict.keys() if
int(Q_PidTotalDurationDict[Pid]) < Threshold]

    # Generates the report for this function
    Header = "Total Survey Duration Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following respondents spent less time than 3 Standard
Deviations from the Mean:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(Q_PidTotalDurationDict[pid])
            Message.append(p)
    else:
        Message.append("All participants have spent an acceptable amount of time on
the survey.")

    Reporter(Header, HeaderHighlight, Message)

    return FailedPids, Q_PidTotalDurationDict

# Evaluates reCaptcha scores and identifies violators. Qualtrics criteria is used:
Less than 0.5 is likely a bot.

```

```

def Captcha(Q_Dataframe):
    Q_PidCaptcha = {Q_Pid: Q_Captcha for Q_Pid, Q_Captcha in
                    zip(List(Q_Dataframe["ProlificID"]),
                        List(Q_Dataframe["HumanChance"]))
                    }

    FailedPids = [Pid for Pid in Q_PidCaptcha.keys() if float(Q_PidCaptcha[Pid]) <
0.5]

    Header = "Bot Detection Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following respondents are likely bots:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(Q_PidCaptcha[pid])
            Message.append(p)
    else:
        Message.append("All participants are likely to be human.")

    Reporter(Header, HeaderHighlight, Message)

    return FailedPids, Q_PidCaptcha

# Checks if the participants committed to being attentive, thoughtful, and honest.
def CommitmentCheck(Q_Dataframe):
    Q_PidCommitment = {Q_Pid: Q_Commitment for Q_Pid, Q_Commitment in
                      zip(List(Q_Dataframe["ProlificID"]),
                          List(Q_Dataframe["CommitmentCheck"]))
                      }

    FailedPids = [Pid for Pid in Q_PidCommitment.keys() if Q_PidCommitment[Pid] !=
"Yes, I will!"]

    Header = "Commitment Check Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following respondents did not commit to being attentive,
thoughtful, and honest:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(Q_PidCommitment[pid])
            Message.append(p)
    else:

```

```

        Message.append("All participants committed to being attentive, thoughtful,
and honest.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, Q_PidCommitment

# Checks whether participants reported they can assume being in the market to
purchase a BPM
def MindsetCheck(Q_Dataframe):
    Q_PidMindset = {Q_Pid: Q_Mindset for Q_Pid, Q_Mindset in
                    zip(List(Q_Dataframe["ProlificID"]),
                        List(Q_Dataframe["MindsetAssumption"]))
                    }

    FailedPids = [Pid for Pid in Q_PidMindset.keys() if Q_PidMindset[Pid] != "Yes"]

    Header = "Mindset Assumption Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following respondents could not assume the required
mindset:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(Q_PidMindset[pid])
            Message.append(p)
    else:
        Message.append("All participants could assume the required mindset.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, Q_PidMindset

# VERIFIED OPERATION, UNKNOWN IMPLEMENTATION
# Checks whether participants paid attention during checks 1 and 2.
def AttentionChecks(Q_Dataframe):
    Q_PidAttChk1and2 = {Q_Pid: [AttChk1, AttChk2] for Q_Pid, AttChk1, AttChk2 in
                        zip_longest(List(Q_Dataframe["ProlificID"]),
                                    List(Q_Dataframe["AttentionCheck1"]),
                                    List(Q_Dataframe["AttentionCheck2"]))
                        }

    FailedPidAttChk1and2 = {} # Only violations will be reported, not all key-values
like other functions
    TempList = []
    for Pid in Q_PidAttChk1and2.keys():

```

```

        if Q_PidAttChk1and2[Pid][0] != "Strongly Disagree" and
Q_PidAttChk1and2[Pid][0] != "Disagree":
            TempList.append("Attn. 1 " + str(Q_PidAttChk1and2[Pid][0]))

        if Q_PidAttChk1and2[Pid][1] != "Strongly Agree" and Q_PidAttChk1and2[Pid][1]
!= "Agree":
            TempList.append("Attn. 2 " + str(Q_PidAttChk1and2[Pid][1]))

    if len(TempList) > 0:
        FailedPidAttChk1and2[Pid] = TempList

    TempList=[]

FailedPids = list(FailedPidAttChk1and2.keys())

Header = "Attention Check 1 & 2 Report"
HeaderHighlight = Highlight(len(FailedPids) > 0)
Message = []
if len(FailedPids) > 0:
    Message.append("The following respondents failed attention check 1 and/or
2:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(FailedPidAttChk1and2[pid])
        Message.append(p)
else:
    Message.append("All participants passed attention checks.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidAttChk1and2

#          VERIFIED OPERATION, UNKNOWN IMPLEMENTATION
# Checks whether participants correctly answered the question about instructions.
def InstructionComprehensionDCE(Q_Dataframe):
    Q_PidInstCompDCE = {Q_Pid: Q_InstCompDCE for Q_Pid, Q_InstCompDCE in
                        zip(list(Q_Dataframe["ProlificID"]),
                            list(Q_Dataframe["InstructionsComprehensionCheck"]))
                        }

    FailedPids = [Pid for Pid in Q_PidInstCompDCE.keys() if Q_PidInstCompDCE[Pid] !=
"HomeHealth"]

    Header = "DCE Instruction Comprehension Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []

```

```

    if len(FailedPids) > 0:
        Message.append("The following respondents failed instruction comprehension
check:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(Q_PidInstCompDCE[pid])
            Message.append(p)
        else:
            Message.append("All participants passed instruction comprehension check.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, Q_PidInstCompDCE

# Checks whether participants identified the same sex across Prolific and Qualtrics
data.
# Qualtrics asked for gender, and there might be discrepancy here. But this can be
useful as
# yet another metric for data consistency check.
def SexCrossValidation(P_Dataframe, Q_Dataframe):
    P_PidSex = {P_Pid: P_Sex for P_Pid, P_Sex in
                zip(List(P_Dataframe["ProlificID"]),
                    List(P_Dataframe["Sex"]))
    }

    Q_PidSex = {Q_Pid: Q_sex for Q_Pid, Q_sex in
                zip(List(Q_Dataframe["ProlificID"]),
                    List(Q_Dataframe["Sex"]))
    }

    QandP_PidSex = {}
    TempList = []
    for Pid in Q_PidSex.keys():
        Pid = str(Pid)
        TempList.append(P_PidSex[Pid])
        TempList.append(Q_PidSex[Pid])
        QandP_PidSex[Pid] = TempList
        TempList = []

    FailedPids = [pid for pid in QandP_PidSex.keys() if QandP_PidSex[pid][0] !=
QandP_PidSex[pid][1]]

    Header = "Sex Cross Validation Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:

```

```

    Message.append("The following respondents failed sex cross validation between
Prolific and Qualtrics data:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(QandP_PidSex[pid])
        Message.append(p)
    else:
        Message.append("All participants passed sex cross validation between Prolific
and Qualtrics data.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, QandP_PidSex

# For all participants, determines if each page submission time is an outlier.
Interquartile Range method is used.
def IndividualPageTimes(Q_Dataframe):
    FailedPidPageTimes = OutliersByIQR_Multi(Q_Dataframe, PageSubmissionTimesList,
"ProlificID")
    FailedPids = List(FailedPidPageTimes.keys())

    Header = "Page Submission Outliers Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following participants have outlying page submission
times:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(FailedPidPageTimes[pid])
            Message.append(p)
    else:
        Message.append("No participant has outlying page submission times")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidPageTimes

# For all participants, determines if each page click count is an outlier.
Interquartile Range method is used.
def IndividualClickCounts(Q_Dataframe):
    FailedPidClickCounts = OutliersByIQR_Multi(Q_Dataframe, PageClickCountsList,
"ProlificID")
    FailedPids = List(FailedPidClickCounts.keys())

    Header = "Page Click Counts Outliers Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)

```

```

Message = []
if len(FailedPids) > 0:
    Message.append("The following participants have outlying page click count:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(FailedPidClickCounts[pid])
        Message.append(p)
else:
    Message.append("No participant has outlying page click counts")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidClickCounts

# For all participants, determines if the total time spent on DCE Instructions is an
# outlier. Interquartile Range method is used.
# Because this function required the addition of two values, the OutliersByIQR()
# function could not be used and it was copied and adapted.
def InstructionsTimeDCE(Q_Dataframe):
    SourceDataframe = Q_Dataframe
    ColumnNamesList = TimeOnDCEInstructions
    IDColumnName = "ProlificID"

    IDs = list(SourceDataframe[IDColumnName])
    IndividualDataList = [list(SourceDataframe.loc[idx, ColumnNamesList]) for idx in
SourceDataframe.index]
    IndivSumInstTime = [sum(list(map(float, TimeList))) for TimeList in
IndividualDataList]
    IDsDataDict = {k:[v] for k, v in zip(IDs, IndivSumInstTime)}

    ColumnsList = [IndivSumInstTime]
    Q1s = [pd.Series(list(map(float, ValuesList))).quantile(0.25) for ValuesList in
ColumnsList]
    Q3s = [pd.Series(list(map(float, ValuesList))).quantile(0.75) for ValuesList in
ColumnsList]
    IQRs = [Q3 - Q1 for Q3, Q1 in zip(Q3s, Q1s)]

    OutliersKeyValues = {}
    TempList = []
    for pid in IDsDataDict.keys():
        for i in range(len(IDsDataDict[pid])):
            Value = float(IDsDataDict[pid][i])
            if Value < Q1s[i] - 1.5*IQRs[i]:
                TempList.append("(Q16&18) DCE Inst. " + str(round(Value, 2)) + " <<<
" + str(round(Q1s[i], 2)))
            if Value > Q3s[i] + 1.5*IQRs[i]:

```

```

        TempList.append("(Q16&18) DCE Inst. " + str(round(Value, 2)) + " >>>
" + str(round(Q3s[i], 2)))

    if len(TempList) > 0:
        OutliersKeyValues[pid] = TempList

    TempList = []

FailedPidDCEInst = OutliersKeyValues

FailedPids = List(FailedPidDCEInst.keys())

Header = "Time on DCE Instructions Outliers Report"
HeaderHighlight = Highlight(len(FailedPids) > 0)
Message = []
if len(FailedPids) > 0:
    Message.append("The following participants have outlying time on DCE
Instructions:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(FailedPidDCEInst[pid])
        Message.append(p)
else:
    Message.append("No participant has outlying time on DCE Instructions")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidDCEInst

# For all participants, determines if the time spent on DCE Dominant Alternative is
an outlier. Interquartile Range method is used.
def DomAltTime(Q_Dataframe):
    FailedPidDomAlt = OutliersByIQR_Single(Q_Dataframe,
TimeOnDominantAlternativeCheck, "ProlificID")
    FailedPids = List(FailedPidDomAlt.keys())

    Header = "Time on DCE Dominant Alternative Outliers Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following participants have outlying time on DCE Dominant
Alternative:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(FailedPidDomAlt[pid])
            Message.append(p)
    else:

```

```

        Message.append("No participant has outlying time on DCE Dominant
Alternative")

    Reporter(Header, HeaderHighlight, Message)

    return FailedPids, FailedPidDomAlt

# For all participants, determines if the time spent on DCE Choice Consistency is an
outlier. Interquartile Range method is used.
def ChoiceConsTime(Q_Dataframe):
    FailedPidChoiceCons = OutliersByIQR_Single(Q_Dataframe,
TimeOnChoiceConsistencyCheck, "ProlificID")
    FailedPids = List(FailedPidChoiceCons.keys())

    Header = "Time on DCE Choice Consistency Outliers Report"
    HeaderHighlight = Highlight(len(FailedPids) > 0)
    Message = []
    if len(FailedPids) > 0:
        Message.append("The following participants have outlying time on DCE Choice
Consistency:")
        for pid in FailedPids:
            p = str(pid) + "\t" + str(FailedPidChoiceCons[pid])
            Message.append(p)
    else:
        Message.append("No participant has outlying time on DCE Choice Consistency")

    Reporter(Header, HeaderHighlight, Message)

    return FailedPids, FailedPidChoiceCons

# For all participants, determines if the approximate time spent on the DCE itself is
an outlier. Interquartile Range method is used.
# Approximation is done by subtracting all other recorded times from total duration.
def ApproxTimeDCE(Q_Dataframe):
    Q_Pids = List(Q_Dataframe["ProlificID"])
    Q_TotalDur = List(Q_Dataframe["TotalDuration"])
    Q_PageSubTime = [List(Q_Dataframe.loc[idx, PageSubmissionTimesList]) for idx in
Q_Dataframe.index]

    ApproxTimeDCE = [float(t) - sum(List(map(float, s))) for t, s in zip(Q_TotalDur,
Q_PageSubTime)]

    PidApproxTimeDCE = {k:[v] for k, v in zip(Q_Pids, ApproxTimeDCE)}

```

```

Q1 = [pd.Series(list(map(float, ApproxTimeDCE))).quantile(0.25)]
Q3 = [pd.Series(list(map(float, ApproxTimeDCE))).quantile(0.75)]
iqr = [Q3[0] - Q1[0]]

FailedPidTimes = {}
TempList = []
for pid in PidApproxTimeDCE.keys():
    for i in range(len(PidApproxTimeDCE[pid])):
        Value = float(PidApproxTimeDCE[pid][i])
        if Value < Q1[i] - 1.5*iqr[i]:
            TempList.append("Approx. DCE Time " + str(round(Value, 2)) + " <<< "
+ str(round(Q1[i], 2)))
        if Value > Q3[i] + 1.5*iqr[i]:
            TempList.append("Approx. DCE Time " + str(round(Value, 2)) + " >>> "
+ str(round(Q3[i], 2)))

    if len(TempList) > 0:
        FailedPidTimes[pid] = TempList

    TempList = []

FailedPids = list(FailedPidTimes.keys())

Header = "Approx. Time on DCE Outliers Report"
HeaderHighlight = Highlight(len(FailedPids) > 0)
Message = []
if len(FailedPids) > 0:
    Message.append("The following participants have outlying approximate time on
DCE:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(FailedPidTimes[pid])
        Message.append(p)
else:
    Message.append("No participant has outlying approximate time on DCE")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidTimes

# For all participants, determines if they identified the purpose of the study
appropriately, using GPT 3.5 Turbo engine.
# Failures still need to be checked manually, but in general, it appears to be
categorizing responses well.
def StudyPurpose(Q_Dataframe):

```

```

PidSentences = {k:v for k, v in zip(Q_Dataframe["ProlificID"],
Q_Dataframe["StudyPurpose"])}
    CanonicPurpose = "The purpose of this study is to understand how people purchase
and use blood pressure monitors, and what factors impact their decision-making in the
purchasing process"

PidRelevance = {}
if BypassOpenAIAPI != "y":
    for pid in PidSentences.keys():
        sentence = PidSentences[pid]
        response = client.chat.completions.create(
            model="gpt-3.5-turbo",
            messages=[{"role": "system",
                "content": "You are a helpful assistant and answer with
either yes or no."},
                {"role": "user",
                "content": f"Is this sentence '{sentence}' relevant to this
sentence: '{CanonicPurpose}'?"}],
            max_tokens=1,
            temperature=0
        )
        relevance = response.choices[0].message.content # Responses will be "Yes"
or "No", for relevant and irrelevant respectively
        PidRelevance[pid] = relevance
        time.sleep(1) # To avoid triggering API rate limits (should be at 3500
requests per minute, but sometimes fails)
    else:
        tc.cprint("OpenAI API is bypassed. No results will be available from Study
Purpose check.",
            "black", "on_red", attrs=["bold"])

# FailedPidSentence = {k:[v] for k, v in PidSentences.items() if PidRelevance[k]
== "No"}
# FailedPids = FailedPidSentence.keys()

FailedPids = [pid for pid in PidRelevance.keys() if PidRelevance[pid] == "No"]
FailedPidSentence = {}
for pid in FailedPids:
    FailedPidSentence[pid] = PidSentences[pid]

Header = "Stated Study Purpose Relevance Report"
HeaderHighlight = Highlight(len(FailedPids) > 0)
Message = []
if len(FailedPids) > 0:

```

```

    Message.append("The following participants likely misidentified the purpose of
the study:")
    for pid in FailedPids:
        p = str(pid) + "\t" + str(FailedPidSentence[pid])
        Message.append(p)
    elif BypassOpenAI_API == "y":
        Message.append("(No Data to Show)")
    else:
        Message.append("No participant has misidentified the study purpose.")

Reporter(Header, HeaderHighlight, Message)

return FailedPids, FailedPidSentence
#-----

#----- VISUALIZING RESPONDENT GEO-LOCATIONS -----
# Maps the location of respondents based on the Latitude and Longitude data from
Qualtrics, extracted from their IP by the platform
# Qualtrics extracted these Latitudes and Longitudes based on IP, and according to
their documentation, in the US there is a 95-99%
# chance for accuracy at City Level geolocation based on IP.
# This is just for visualizing the approximate location of respondents, and not for
validating their entries.
def IPMap(Q_Dataframe):
    RawGeoDataFrame = {
        "Pid": [pid for pid in Q_Dataframe["ProlificID"]],
        "Lat": [float(lat) for lat in Q_Dataframe["LocationLatitude"]],
        "Long": [float(lat) for lat in Q_Dataframe["LocationLongitude"]]
    }

    GeoDataFrame = pd.DataFrame(RawGeoDataFrame)
    #for more mapbox styles, see: https://plotly.github.io/plotly.py-
docs/generated/plotly.express.scatter_mapbox.html
    figa = px.scatter_mapbox(GeoDataFrame, Lat="Lat", Lon="Long", mapbox_style="open-
street-map",
                            hover_name="Pid", zoom=4, center=dict(lat=37.5, lon=-
97.5))
    figa.update_traces(marker={"size":50,
                              "color": "crimson",
                              "opacity": 0.32
                              })

    SaveFigureAs = str(ResultsDirectoryPath) + "\\ " + "aRespondent Locations - " +
str(Identifier) + ".png"
    figa.write_image(SaveFigureAs, width=2160, height=1440, scale=4)

```

```

plotly.offline.plot(figa)

figb = px.scatter_geo(GeoDataFrame, Lat="Lat", Lon="Long", hover_name="Pid") # ,
mapbox_style="carto-darkmatter", zoom=4
figb.update_traces(marker={"size":50,
                           "color": "orangered",
                           "opacity": 0.5
                          })

figb.update_layout(
    geo = dict(
        scope='usa',
        showLand = True,
        landcolor = "black",
        subunitcolor = "dimgrey",
        subunitwidth = 2.5
    ),
)

SaveFigureAs = str(ResultsDirectoryPath) + "\\ " + "bRespondent Locations - " +
str(Identifier) + ".png"
figb.write_image(SaveFigureAs, width=2160, height=1440, scale=4)
plotly.offline.plot(figb)
#-----
# Creating reports for each of the violation categories
# If a new validation function is added above, areas denoted by "# <<<<<<<<<" need
to be updated.
AgeViolators, AgeValues = AgeCrossValidation(ProlificDemogDF, SurveyDataDF)
DisViolators, DisValues = DisabilityCrossValidation(ProlificDemogDF, SurveyDataDF)
EthViolators, EthValues = EthnicityCrossValidation(ProlificDemogDF, SurveyDataDF)
CountryViolators, CountryValues = IPinUS(ProlificDemogDF, SurveyDataDF)
StateViolators, StateValues = IPStateValidation(SurveyDataDF)

# New in Phase 2
DurationViolators, DurationValues = TotalDurationOutliers(SurveyDataDF)
CaptchaViolators, CaptchaValues = Captcha(SurveyDataDF)
CommitmentViolators, CommitmentValues = CommitmentCheck(SurveyDataDF)
MindsetViolators, MindsetValues = MindsetCheck(SurveyDataDF)
AttentionViolators, AttentionValues = AttentionChecks(SurveyDataDF)
InstructionComprehensionViolators, InstructionComprehensionValues =
InstructionComprehensionDCE(SurveyDataDF)

```





```

        if v in StudyPurposeViolators: TempListViolations.append("StudyPurpose");
TempListValues.append(StudyPurposeValues[v])

    ViolatorsDict[v] = [TempListViolations, TempListValues]
    TempListViolations = []
    TempListValues = []
ViolatorsDict = SortDict(ViolatorsDict)

NonViolatorsDict = {}
for r in P_RemainingPids:
    if r not in ViolatorsDict.keys(): NonViolatorsDict[r] = ["-", "-"]
NonViolatorsDict = SortDict(NonViolatorsDict)

PidViolationsReport = {**ViolatorsDict, **NonViolatorsDict} # Merges the
violators and non-violators dictionaries, perserving order.

# If summary report is requested, the PidViolationsReport is formatted as a Pretty
Table and saved as txt.
if CreateSummaryText == "y":
    ViolationsTable = []
    TempStrV0 = ""
    TempStrV1 = ""
    for k, v in ViolatorsDict.items():
        # The "for" loops below just make the presentation look cleaner by removing
items from lists and writing/concatenating them as a string
        for i in range(len(v[0])):
            TempStrV0 += str(v[0][i]) + " - "
        TempStrV0 = TempStrV0[:-3]
        for i in range(len(v[1])):
            TempStrV1 += str(v[1][i]) + " - "
        TempStrV1 = TempStrV1[:-3]
        ViolationsTable.append([str(k), TempStrV0, TempStrV1])
        TempStrV0 = ""
        TempStrV1 = ""

    ReportTable = PrettyTable(["Prolific ID", "Violation Categories", "Violations"])
    ReportTable.add_rows(ViolationsTable)
    ReportTable.align["Violation Categories"] = "r"
    ReportTable.align["Violations"] = "l"

    txtFileName = "Violations Summary - " + str(Identifier) + " - " + str(DnTApprop)
+ ".txt"
    txtFileAddress = str(ResultsDirectoryPath) + "\\ " + txtFileName
    txtFile = open(txtFileAddress, "a")

```

```

    if BypassIPinfoAPI == "y": txtFile.write(">>> IP Geolocation was bypassed,
latitude and longitude data from Qualtrics was used instead.\n\n")
    if BypassOpenAIAPI == "y": txtFile.write(">>> OpenAI was bypassed, no results
from Study Purpose Check are included in this report.\n\n")

    if MultiplesExist:
        txtFile.write("Multiple entries detected from the following respondents:\n")
        for i in DuplicatesPID: txtFile.write("\t"+str(i)+"\n")
        txtFile.write("Multiple entries from the same respondents will be excluded
from the data.\n\n")

    if len(ListOfDNF) > 0:
        txtFile.write("These respondents did not finish the survey:\n")
        for i in ListOfDNF: txtFile.write("\t"+str(i)+"\n")
        txtFile.write("Unfinished entries will be excluded from the data.\n\n")

    underline = ""
    for i in range(len("Participants Remaining in the Analysis: " +
str(NumberOfReminingParticipants))): underline += "="
    txtFile.write("Participants Remaining in the Analysis: " +
str(NumberOfReminingParticipants) + "\n")
    txtFile.write(str(underline) + "\n\n")

    txtFile.write(str(ReportTable))

    txtFile.close()

    tc.cprint("\n" + "Violations report created successfully.", "black", "on_cyan",
attrs=["bold"])
    tc.cprint("\t" + str(txtFileAddress), "black", "on_light_grey")

# If updated survey data is requested, a copy of the original survey data dataframe
is created and columns containing violation categories and violations
# are added. The updated dataframe is then saved as CSV.
if CreateUpdatedSurveyCSV == "y":
    PidsToDiscard = [i for i in SurveyDataDF["ProlificID"] if i not in
PidViolationsReport.keys()] # Prolific IDs inelligible for pay, to be removed from
the survey data
    IndicesToDiscard = IndicesFinder(SurveyDataDF, "ProlificID", PidsToDiscard)
    UpdatedSurveyDF = SurveyDataDF.copy(deep=True)
    UpdatedSurveyDF.drop(IndicesToDiscard, inplace=True)
    RelocateCol = UpdatedSurveyDF.pop("ProlificID") # Along with the next line
relocates Prolific ID to the 1st column
    UpdatedSurveyDF.insert(0, "ProlificID", RelocateCol)

```

```

UpdatedSurveyDF.insert(1, "ViolationCategories", "") # Along with the next line
creates two new empty columns for violations data
UpdatedSurveyDF.insert(2, "Violations", "")

# Writes violations data under corresponding columns for each Prolific ID
TempStrViolCats = ""
TempStrViols = ""
for Pid in UpdatedSurveyDF["ProlificID"]:
   RowIndex =
List(UpdatedSurveyDF[UpdatedSurveyDF["ProlificID"]==Pid].index.values)

    # The "for" loop just makes the presentation look cleaner by removing items
from lists and writing/concatenating them as a string
    for i in range(len(PidViolationsReport[Pid][0])):
        TempStrViolCats += str(PidViolationsReport[Pid][0][i]) + " - "
        TempStrViolCats = TempStrViolCats[:-3]
        UpdatedSurveyDF.loc[RowIndex, "ViolationCategories"] = TempStrViolCats

    # The "for" loop just makes the presentation look cleaner by removing items
from lists and writing/concatenating them as a string
    for i in range(len(PidViolationsReport[Pid][1])):
        TempStrViols += str(PidViolationsReport[Pid][1][i]) + " - "
        TempStrViols = TempStrViols[:-3]

    UpdatedSurveyDF.loc[RowIndex, "Violations"] = TempStrViols

    TempStrViolCats = ""
    TempStrViols = ""

# Creates a new dictionary that includes the sort order for the rows in the
updated dataframe. Otherwise, sorting the data based on Prolific ID while showing
# the violators at the top (still sorted by PID but not alphabetically for
violation categories), would be very challenging with default Pandas sorting methods.
SortOrderDict = {}
for i in range(len(List(PidViolationsReport.keys()))):
    SortOrderDict[List(PidViolationsReport.keys())[i]] = i
UpdatedSurveyDF.sort_values(by=["ProlificID"], inplace=True, key=lambda x:
x.map(SortOrderDict))

UpdatedCSVName = "ManualCheck - " + str(Identifier) + " - " + str(DnTApprop) +
".csv"
CSVFileAddress = str(ResultsDirectoryPath) + "\\ " + UpdatedCSVName
UpdatedSurveyDF.to_csv(CSVFileAddress, index=False)

```

```
tc.cprint("\n" + "Please review the responses manually before using the data in  
the next stages of processing.", "black", "on_red", attrs=["bold", "blink"])  
tc.cprint("CSV file containing UNIQUE responses deemed ELIGIBLE for payment by  
Prolific was created successfully.", "black", "on_cyan", attrs=["bold"])  
tc.cprint("\t" + str(CSVfileAddress), "black", "on_light_grey")
```

*### It is crucial to check the output CSV manually. Not all violators in fact  
need to be discarded (e.g. race on Prolific doesn't include Hispanic and some may  
### check "White" on prolific but in Qualtrics select Other and type Hispanic).  
Conversely, not all non-violators are necessarily legitimate respondents.*

*Comprehension*

*### and attention tests should be reviewed manually for all, and is very  
difficult to implement in code (e.g. reading an essay response and determining  
relevance).*

## Appendix I: Syntax – Internal Validity Report & Preprocessing in Python

```
import pandas as pd
import numpy as np
import plotly_express as px
import termcolor as tc
import plotly
import random
import datetime as dt
import sys
import os

#----- USER INPUT -----
-----
EfficientDesignMatrix = "V06M01design.csv" # CSV file containing the DB-Efficient
Design
SurveyChoiceData = "ValidatedData - Main.csv" # CSV file containing survey data
Identifier = "Main-DESV06M01-11082023-0816AM" # Unique identifier to reference the
output filename to the corresponding survey dataset
DominantAlts = [73, 31] # Product numbers for Dominant Alternatives
ConsistencyCheckRefColNum = 6 # DCE Question to use as reference for Choice
Consistency check. Corresponds with the order of DCE questions, not column name.

DurationColumnName = "TotalDuration" # Name of the column in the data that contains
the total duration of the survey
DomAltSurveyColumnName = "Q22_1" # Column name for Dominant Alternative check
question in the survey data
ConsChkSurveyColumnName = "Q40_1" # Column name for Consistency Check question in the
survey data
DCEQuestionsColumnNames = ["Q23_1", "Q24_1", "Q25_1", "Q26_1", "Q27_1", "Q28_1",
"Q29_1", "Q30_1",
"Q31_1", "Q32_1", "Q33_1", "Q34_1", "Q35_1", "Q36_1",
"Q37_1", "Q38_1"] # Column names for DCE questions in survey data

CreateReportFile = "y" # Create report TXT including internal validity test results
and duration information?
CreateMNLOutputFile = "y" # Create CSV file formatted based on the requirements of
the analysis code?

### Set ExcludeUnreliablePids to "n" for the first run on the data. After the first
run, if there are participants who failed Dominant Alternative or Choice
# Consistency checks, set ExcludeUnreliablePids to "y", and identify the offenders
using their ProlificIDs in DominantAlternativeOffenderPids and
```

```

# ChoiceConsistencyOffenderPids Lists below. If there are no offenders for one of the
categories, the corresponding list can be left empty.
# This is used to create a new report, identified with "IntValid - " prefix at the
beginning of the filenames, to run a separate analysis including
# only those who support internal validity.
ExcludeUnreliablePids = "n"

if ExcludeUnreliablePids == "y":
    DominantAlternativeOffenderPids = []
    ChoiceConsistencyOffenderPids = []

#-----
#-----

#----- INITIALIZATIONS -----
#-----

CurrentDateAndTime = dt.datetime.now()
DnTApprop = CurrentDateAndTime.strftime("%Y%m%d_%H%M%S")

# Reads required data into dataframes
EfficientDesignMatrixDF = pd.read_csv(str(os.path.dirname(__file__)) + "\\\" +
EfficientDesignMatrix) # Read the design matrix into a dataframe
SurveyChoiceDataDF = pd.read_csv(str(os.path.dirname(__file__)) + "\\\" +
SurveyChoiceData) # Read the survey data into a dataframe

# Finds the indices for specific value(s) in a dataframe column
# This function is placed here only because the next "if" statement needs it.
def IndicesFinder(Dataframe, Column, ListOfValues):
    Indices = []
    for i in ListOfValues:
        index = list(Dataframe[Dataframe[str(Column)]==i].index.values)
        for v in index:
            Indices.append(v)
    return Indices

### If there are ProlificIDs from previous runs that offended the internal validity
checks, this first confirms that dropping is done intentionally,
# then prints a confirmation, prints IDs that will be discarded, sets the prefix for
the filenames to "IntValid - ", and drops the entries from the survey
# dataframe. If the inclusion of those IDs was a mistake and confirmation is denied,
the script stops running altogether so the code can be fixed. If excluding
# individuals is not required, the code continues as normal and the prefix is set to
empty so it will not impact filenames going forward.
if ExcludeUnreliablePids == "y":

```

```

ConfirmationToExcludeUnreliablePids = str(input('ExcludeUnreliablePids' is set
to 'y' in the code.\n Type 'go' + Enter to confirm, anything else to quit: "))

if ConfirmationToExcludeUnreliablePids == "go":
    tc.cprint("Confirmed: Pre-identified unreliable participants below will be
excluded:", "black", "on_yellow", attrs=["bold"])
    IntValOffenderPids = List(set(DominantAlternativeOffenderPids +
ChoiceConsistencyOffenderPids))
    for pid in IntValOffenderPids: print("\n\t"+str(pid))
    print("\n")
    FilenamePrefix = "IntValid - "
    IntValOffenderIdx = IndicesFinder(SurveyChoiceDataDF, "ProlificID",
IntValOffenderPids)
    SurveyChoiceDataDF.drop(index=IntValOffenderIdx, inplace=True)
    SurveyChoiceDataDF.reset_index(inplace=True, drop=True) # Restarts the index
form zero so there are no missing indices after deletion
else:
    tc.cprint("Process terminated", "black", "on_red", attrs=["bold"])
    sys.exit()

else:
    FilenamePrefix = ""
    tc.cprint("All participants will be included in the analysis", "black",
"on_green", attrs=["bold"])

tc.cprint("\nNumber of included participants: " + str(len(SurveyChoiceDataDF)),
"black", "on_white", attrs=["bold"])
print("\n")

# Creates a unique Results folder in the parent folder containing the code to store
all results and avoid clutter in the parent directory
ResultsDirectoryName = str(FilenamePrefix) + "Results - " + str(Identifier) + " - " +
str(DnTApprop)
ResultsDirectoryPath = str(os.path.dirname(__file__)) + "\\ " + ResultsDirectoryName
os.mkdir(ResultsDirectoryPath)

# Cleans up the survey dataframe by only keeping relevant columns
ColumnsToKeep = List(DCEQuestionsColumnNames) # IMPORTANT: if List() is not added,
apparently Python mutates the original List
ColumnsToKeep.extend((DurationColumnName, DomAltSurveyColumnName,
ConsChkSurveyColumnName, "ProlificID"))
ColumnsToDrop = [col for col in List(SurveyChoiceDataDF.columns) if col not in
ColumnsToKeep]
SurveyChoiceDataDF.drop(columns=ColumnsToDrop, inplace=True)

```

```

#RenamingDict is a dictionary that contains the new column names for the data, and
corresponds with what the rest of the code will be looking for.
RenamingDict = {
    DurationColumnName: "DurationSec",
    DomAltSurveyColumnName: "DomAlt",
    ConsChkSurveyColumnName: str("ConsChk"+str(ConsistencyCheckRefColNum))
}

# Numbers DCE questions to avoid using their Qualtrics question labels
for q in range(len(DCEQuestionsColumnNames)):
    RenamingDict[DCEQuestionsColumnNames[q]] = str(q+1)
SurveyChoiceDataDF = SurveyChoiceDataDF.rename(columns=RenamingDict)

# Creates basic reports for validation. Allows to check whether the data were read
correctly. Presented on the console output.
NumParticipantRows = len(SurveyChoiceDataDF) # Determines the number of participants
by examining the length of the survey data (excluding headers)
NumDesignMatrixRows = len(EfficientDesignMatrixDF.index) # Determines the total
number of profiles (or alternatives) available in the efficient design
ChoiceTaskNum = int(max([i for i in EfficientDesignMatrixDF["qID"].values if (i != "
" and i != "\n")])) # Determines the number of choice tasks by determining the
largest number in the qID column in efficient design
print("Number of choice tasks detected:", ChoiceTaskNum)
AltNum = int(max([i for i in EfficientDesignMatrixDF["altID"].values if (i != " " and
i != "\n")])) # Determines the number of alternatives by determining the largest
number in the altID column in efficient design
print("Number of alternatives per choice task detected:", AltNum)
print("\n")

#----- FUNCTIONS -----
-----
# The alternatives in the Qualtrics data export are saved with a lot of unnecessary
spacing characters
# This function removes anything that is not alphanumeric so the useful text can be
manipulated.
def AlphaNumericOnly(data):
    data = str(data)
    OnlyANUMText = ""
    for char in data:
        if (char.isalnum()) == True:
            OnlyANUMText += str(char)
    return(str(OnlyANUMText))

# Creates a list of alternatives (product numbers) selected by a participant
def ParticipantChoices(ParticipantNumber):

```

```

ParticipantIndex = int(ParticipantNumber) - 1 # Participant numbers start at 1,
to get the index 1 is subtracted
ProductChoiceList = [] # Initialization of the choice list that will ultimately
contain the choices made by the specified participant
    for col in range(1, int(ChoiceTaskNum)+1): # Scans only the row that contains
participant's DCE choices
        ChoiceTemp = AlphaNumericOnly(SurveyChoiceDataDF.at[ParticipantIndex,
str(col)]) # Reads the alphanumeric characters in each cell
        ChoiceTemp = ChoiceTemp[7] + ChoiceTemp[8] # Only keeps the product number,
which is made up of the 7th and 8th characters of the alphanumeric string
        ChoiceTemp = int(ChoiceTemp) # Converts string of digits to an integer
number, and also removes of the 0 behind numbers less than 10, to match the profiles
in the choice matrix CSV
        ProductChoiceList.append(ChoiceTemp) # Adds the product number to the
participant's choice list
    return(ProductChoiceList)

### Splits the efficient design dataframe's choice tasks, then compares each
participant choice against the choice task, creates a list that identifies which
product was
# selected in the context of available alternatives. Presents results in long format
to match the desired output for MNL analysis.
def ParticipantChoiceColumn(ChoiceList):
    Profiles = [i for i in EfficientDesignMatrixDF["profileID"].values if (i != " "
and i != "\n")] # Reads the profiles contained within the efficient design
    ChoiceTasks = [Profiles[i:i+AltNum] for i in range(0, len(Profiles), AltNum)] #
Splits the profiles into lists of choice tasks, according to the number of
alternatives in choice tasks
    ChoiceColumn = [] # Initialization of the list that will ultimately contain 0s
and 1s corresponding to the profiles the participant has selected
    for i in range(ChoiceTaskNum): # Sweeps the lists of choice tasks
        WorkingChoiceTask = ChoiceTasks[i] # Identifies which choice task is up for
evaluation against participant choices
        WorkingParticipantChoice = ChoiceList[i] # Identifies which choice made by
the participant needs to be compared against the choice task
        for profile in WorkingChoiceTask: # For each alternative in the choice task,
appends a "1" to the list if it was selected by the participant, otherwise a "0"
            if profile == WorkingParticipantChoice:
                ChoiceColumn.append(1)
            else:
                ChoiceColumn.append(0)
    return(ChoiceColumn)

#----- DURATION REPORT -----

```

```

# Reports the duration average and standard deviation of the survey for all
# participants, and creates a box and whiskers plot
# If report file creation is enabled, a CSV containing raw data and the plot will be
# saved in the output folder.
def DurationReport():
    Duration = List(SurveyChoiceDataDF["DurationSec"])
    DurationMinutes = [round(i/60, 2) for i in Duration]
    AvgDur = round(np.average(DurationMinutes), 2)
    SDDur = round(np.std(DurationMinutes), 2)

    tc.cprint("Duration Report:\n", "black", "on_green", attrs=["bold"])
    print("\tAverage: " + str(AvgDur) + " minutes\n\tStandard Deviation: " +
str(SDDur) + "\n")
    DurationReport = "Duration Report:\n" + "\n\tAverage: " + str(AvgDur) + "
minutes\n\tStandard Deviation: " + str(SDDur) + "\n"

    DurationMinutesDF = pd.DataFrame({"Duration (minutes)": DurationMinutes,
"Participant": SurveyChoiceDataDF["ProlificID"]})

    MajorColor = "black"
    MinorColor = "slategrey"
    # MainPlotColor = random.choice(px.colors.qualitative.Dark24)
    MainPlotColor = "indianred"

    DurationFig = px.box(DurationMinutesDF, x="Duration (minutes)",
        hover_data="Participant", orientation="h", points="all",
        template="plotly_white", width=1000, height=250,
notched=True)
    DurationFig.update_layout(font_family="Calibri", font_size=16,
font_color=MajorColor, plot_bgcolor="whitesmoke",
        margin=dict(t=30, r=30), yaxis_title="")
    DurationFig.update_traces(marker_size=6, marker_symbol="diamond",
marker_color=MainPlotColor, line_width=2.5, line_color=MainPlotColor, jitter = 0.2)
    DurationFig.update_xaxes(showline=True, linecolor=MajorColor,
gridcolor=MinorColor, minor_griddash="dot", dtick=5, minor_ticks="inside",
range=(0,55))
    DurationFig.update_yaxes(showline=True, linecolor=MajorColor)
    # DurationFig.show()
    plotly.offline.plot(DurationFig)

    if CreateReportFile == "y":
        DurationCSV = str(FilenamePrefix) + "Duration" + "_" + str(DnTApprop) +
".csv"
        DurationFileAddress = str(ResultsDirectoryPath) + "\\\" + DurationCSV
        DurationMinutesDF.to_csv(DurationFileAddress, index=False)

```

```

        SaveFigureAs = str(ResultsDirectoryPath) + "\\ " + str(FilenamePrefix) +
"Duration - " + str(Identifier) + " - " + str(DnTApprop) + ".png"
        DurationFig.write_image(SaveFigureAs, width=1000, height=250, scale=2)

    return (DurationReport)

#----- DOMINANT ALT & CONSISTENCY CHECK -----
# Creates various reports on the performance of the participants with the Dominant
Alternative check
def DomAltReport():
    PidDomAltChoice = {k:AlphaNumericOnly(v)[7:9] for k, v in
                        zip(SurveyChoiceDataDF["ProlificID"],
SurveyChoiceDataDF["DomAlt"])}

    DAFailedPidChoice = {k:v for k, v in PidDomAltChoice.items() if int(v) not in
DominantAlts}
    DAPassedPidChoice = {k:v for k, v in PidDomAltChoice.items() if k not in
DAFailedPidChoice.keys()}

    FailedPids = list(DAFailedPidChoice.keys())

    DAScore = str(round((len(DAPassedPidChoice)/len(PidDomAltChoice))*100, 2)) +
"" "% ""

    DASelectionReport = {da:str(list(DAPassedPidChoice.values()).count(str(da))) + "
times" for da in DominantAlts}

    DAREport = "Dominant Alternative Selection Report:\n\n"
    DAREport += "\t>>> Failed IDs - Dominant Alternative: " + "Count = " +
str(len(FailedPids)) + "\n" + str(FailedPids) + "\n\n"
    DAREport += "\t>>> Successful Choice Counts Per Dominant Alternative: " +
str(DASelectionReport) + "\n\n"
    DAREport += "\t>>> Dominant Alternative Selection Rate: " + str(DAScore) + "\n"

    MessageColor = "on_green"
    if len(FailedPids) > 0: MessageColor = "on_yellow"
    tc.cprint("Dominant Alternative Selection Report:\n", "black", MessageColor,
attrs=["bold"])
    print("\t>>> Failed IDs - Dominant Alternative: " + "Count = " +
str(len(FailedPids)) + "\n" + str(FailedPids) + "\n\n" +

```

```

        "\t>>> Successful Choice Counts Per Dominant Alternative: " +
str(DASelectionReport) + "\n\n" +
        "\t>>> Dominant Alternative Selection Rate: " + str(DAScore) + "\n\n"
    )

    return (DAREport)

# Creates various reports on the performance of the participants with the Consistency
Check
def ConsistencyCheckReport():
    PidChoiceCons = {pid:[AlphaNumericOnly(ref)[7:9], AlphaNumericOnly(cc)[7:9]] for
pid, ref, cc in
                                zip(SurveyChoiceDataDF["ProlificID"],
                                SurveyChoiceDataDF[str(ConsistencyCheckRefColNum)],
                                SurveyChoiceDataDF[str(RenamingDict[ConsChkSurveyCo
lumnName]])]
                                )
                                }

    CCFailedPidChoice = {k:v for k, v in PidChoiceCons.items() if v[0] != v[1]}

    CCPassedPidChoice = {k:v for k, v in PidChoiceCons.items() if k not in
CCFailedPidChoice.keys()}

    FailedPids = list(CCFailedPidChoice.keys())

    CCScore = str(round((len(CCPassedPidChoice)/len(PidChoiceCons))*100, 2)) +
"" "% ""

    CCREport = "Choice Consistency Report:\n\n"
    CCREport += "\t>>> Failed IDs - Choice Consistency: " + "Count = " +
str(len(FailedPids)) + "\n" + str(FailedPids) + "\n\n"
    CCREport += "\t>>> Failed IDs and Choices - Choice Consistency: " + "\n" +
str(CCFailedPidChoice) + "\n\n"
    CCREport += "\t>>> Choice Consistency Rate: " + str(CCScore) + "\n"

    MessageColor = "on_green"
    if len(FailedPids) > 0: MessageColor = "on_yellow"
    tc.cprint("Choice Consistency Report:\n", "black", MessageColor, attrs=["bold"])
    print("\t>>> Failed IDs - Choice Consistency: " + "Count = " +
str(len(FailedPids)) + "\n" + str(FailedPids) + "\n\n" +
        "\t>>> Failed IDs and Choices - Choice Consistency: " + "\n" +
str(CCFailedPidChoice) + "\n\n" +
        "\t>>> Choice Consistency Rate: " + str(CCScore) + "\n\n"
    )

```

```

return (CCReport)

#----- GENERATING MNL ANALYSIS DATAFRAME -----
-----
#Initialization of the dataframe that resembles the desired input for MNL analysis
MNLReadyDict = {"id": [], #Participant ID, ranges from 1 to Number Of Participants
                "obsID": [], #Observation ID, identifies choice tasks consecutively
                #throughout the whole data
                "alt": [], #Alternative ID, identifies the alternatives within each
                #choice task
                "choice": [], #Column consisting of 0s and 1s, identifying which
                #alternatives were not selected by the participant, and which alternatives were,
                #respectively.
                "cost": [], #Column that shows the same values as the original
                #efficient design, and identifies the "cost" for each alternative
                "brand": [], #Column that shows the same values as the original
                #efficient design, and identifies the "brand" for each alternative
                "UserRating": [], #Column that shows the same values as the original
                #efficient design, and identifies the "user rating" for each alternative
                "UAscore": [] #Column that shows the same values as the original
                #efficient design, and identifies the "accessibility/usability information" for each
                #alternative
                }

for ParticipantNumber in range(NumParticipantRows): #For each participant
    for i in range(NumDesignMatrixRows): #Repeat the participant number as many times
    as there are alternatives in the efficient design
        MNLReadyDict["id"].append(int(ParticipantNumber+1))

obsid = 0 #Initializing the variable that determines the observation ID
for i in range(1, (NumParticipantRows*NumDesignMatrixRows)+1, int(AltNum)): #For each
choice set repeat the same observation id, continuously, for as many observations as
there can be (number of participants * number of alternatives/profiles)
    obsid += 1
    for j in range(AltNum):
        MNLReadyDict["obsID"].append(int(obsid))

for i in range(1, (NumParticipantRows*NumDesignMatrixRows)+1, int(AltNum)): #For each
choice set, number the alternatives, repeatedly, for as many choice sets as there may
be
    for j in range(AltNum):
        MNLReadyDict["alt"].append(int(j+1))

```

```

for ParticipantNumber in range(NumParticipantRows): #For all available participants,
gets the list of binary choices per participant, and adds it to the "choice" cloumn
    ProperParticipantNumber = ParticipantNumber + 1
    Choice = ParticipantChoiceColumn(ParticipantChoices(ProperParticipantNumber))
    for ch in Choice:
        MNLReadyDict["choice"].append(ch)

    Cost = [i for i in EfficientDesignMatrixDF["cost"].values if (i != " " and i !=
"\n")] #Appned the "cost" column from the effcient design, repeatedly, for as many
participants as available
    for c in Cost:
        MNLReadyDict["cost"].append(c)

    Brand = [i for i in EfficientDesignMatrixDF["brand"].values if (i != " " and i !=
"\n")] #Same as above, for "brand"
    for b in Brand:
        MNLReadyDict["brand"].append(b)

    UserRating = [i for i in EfficientDesignMatrixDF["UserRating"].values if (i != "
" and i != "\n")] #Same as above, for "UserRating"
    for ur in UserRating:
        MNLReadyDict["UserRating"].append(ur)

    AccessInfo = [i for i in EfficientDesignMatrixDF["UAscore"].values if (i != " "
and i != "\n")] #Same as above, for "UAscore"
    for uas in AccessInfo:
        MNLReadyDict["UAscore"].append(uas)

#----- CREATING MNL ANALYSIS OUTPUT FILES -----
-----
DUR = DurationReport()
DAR = DomAltReport()
CCR = ConsistencyCheckReport()

if CreateReportFile == "y":
    TextFileName = str(FileNamePrefix) + "Report - " + str(Identifier) + "_" +
str(DnTApprop) + ".txt"
    TextFileAddress = str(ResultsDirectoryPath) + "\\" + TextFileName
    TextFileOut = open(TextFileAddress, "a")
    if ExcludeUnreliablePids == "y":
        TextFileOut.write("Pre-identified unreliable participants below were
excluded:\n" + str(IntValOffenderPids) + "\n\n")
    TextFileOut.write(str(
        DUR + "\n\n"

```

```

        + DAR + "\n\n"
        + CCR + "\n"
    ))
    TextFileOut.close()
    print("Reports prepared successfully!\nCheck the directory for the output TXT
file:\n" + TextFileAddress + "\n")

if CreateMNLOutputFile == "y":
    MNLReadyDF = pd.DataFrame(MNLReadyDict)
    CSVFileName = str(FileNamePrefix) + "MNL-Ready Survey Data" + " - " +
str(Identifier) + "_" + str(DnTApprop) + ".csv"
    CSVFileAddress = str(ResultsDirectoryPath) + "\\ " + CSVFileName
    MNLReadyDF.to_csv(CSVFileAddress, index=False)
    print("MNL ready data prepared successfully!\nCheck the directory for the output
CSV file:\n" + CSVFileAddress)

```

## Appendix J: Syntax – Multinomial Logit Analysis in R

```
library("logitr")
library("stargazer")
library("knitr")
library("plotly")
library("reshape2")
library("broom")
library("gtsummary")

#Open the CSV file containing the choice data from the survey, formatted according to
the requirements of Logitr function
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)
ChoiceData <- read.csv("IntValid - MNL-Ready Survey Data - Main-DESV06M01-11082023-
0816AM_20231112_000857.csv")
head(ChoiceData, 50)

#Use the "Unavailable" Level as reference for coding "UAscore" Levels
ChoiceData$UAscore <- factor(
  x = ChoiceData$UAscore,
  levels = c("Unavailable", "Low (4/10)", "Medium (6/10)", "High (8/10)")
)

#Predict the MNL model based on the choice data
MNL_Model <- logitr(
  data = ChoiceData,
  outcome = "choice",
  obsID = "obsID",
  pars = c("cost", "brand", "UserRating", "UAscore")
)

#Print the summary and performance details of the MNL model, save every table as a
CSV to aid formatting tables later
summary(MNL_Model)
tidy(MNL_Model)
glance(MNL_Model)

write.csv(coef(summary(MNL_Model)), "Model Coefficients - Summary.csv")
write.csv(tidy(MNL_Model), "Model Coefficients - Tidy.csv")
write.csv(glance(MNL_Model), "Model Performance Parameters - Glance.csv")

#Using gtsummary and stargazer for formatting the MNL model output
```

```

tbl_regression(MNL_Model, add_estimate_to_reference_rows = TRUE) %>%
italicize_levels() %>% bold_p(t = 0.05)
stargazer(coef(summary(MNL_Model)), title = "Coefficients of the MNL Model", type =
"html", out = "Stargazer - Model Coefficients Summary.html", summary = FALSE)

#Identify the Variance-Covariance Matrix, save it to multiple output types, and graph
it as a heatmap
VCOV <- vcov(MNL_Model)

COR <- cov2cor(VCOV)

write.csv(COR, "Parameter Correlation Matrix.csv")
write.csv(VCOV, "Model Variance-Covariance Matrix.csv")
stargazer(VCOV, title = "Variance-Covariance Matrix", type = "html", out = "Stargazer
- Model Variance-Covariance Matrix.html", summary = FALSE)

MeltedVCOV <- melt(VCOV)
VarCovPlot <- ggplot(MeltedVCOV, aes(Var1, Var2)) +
  geom_text(aes(Label = round(value, 2)), color = "black", size = 4) +
  geom_raster(aes(fill = value)) +
  scale_fill_gradient2(Low = "darkblue", mid = "white", high = "darkred") +
  ggtitle("Variance-Covariance Matrix") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, size = 10, hjust = 1),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        plot.title = element_text(hjust = 0.5)
        )
ggplotly(VarCovPlot)

#Create and show the residuals plot
Residuals <- residuals(MNL_Model)
ResidualsPlot <- plot_ly(Residuals, type = "scatter", mode = "markers", x = ~obsID,
y = ~residual, color = ~obsID, colors = "Dark2")%>%
  layout(title = list(text = "Residuals", x = 0.5, y = 0.99),
        xaxis = list(title = "Observation ID"),
        yaxis = list(title = "Residual"))
ResidualsPlot

#Predict choices using the MNL model, compare to actual choices, determine success
rate for prediction, and save the data as CSV
PredictedOutcomes <- predict(MNL_Model, type = "outcome", returnData = TRUE)

ChosenProducts <- subset(PredictedOutcomes, choice == 1)

```

```
ChosenProducts$PredictedCorrectly <- ChosenProducts$choice ==  
ChosenProducts$predicted_outcome  
PredictionSuccessRate <- (sum(ChosenProducts$PredictedCorrectly) /  
nrow(ChosenProducts))*100  
roundedPredSuccRt <- round(PredictionSuccessRate, 2)  
paste("Prediction Success Rate: ", as.character(roundedPredSuccRt), "%", sep = "")  
  
OutputPredCSVName <- paste("Outcome Predictions (SuccessRate ",  
as.character(roundedPredSuccRt), "%).csv", sep = "")  
write.csv(PredictedOutcomes, OutputPredCSVName)
```

## **Appendix K: Equivalent Text Descriptions (EqTDs)**

### ***Explanation of EqTDs***

This appendix includes Equivalent Text Descriptions (EqTDs) for each figure in this dissertation, enhancing accessibility by providing written descriptions for visually conveyed information. An EqTD typically comprises three parts: a Brief Description (providing a general overview), an Essential Description (summarizing the context), and a Detailed Description (explanation of visual features). For elements like images, charts, graphs, and symbols, all three descriptions are provided. However, for elements like spreadsheets and tables where the text is readable, only the Brief and Essential Descriptions are included. EqTDs play a crucial role in making visual information accessible and comprehensible, particularly for users who rely on alternative forms of information access.

**Figure 1**

**Brief Description.** Flowchart shows the significance of the dissertation research in reducing disparities between people with and without disabilities.

**Essential Description.** This flowchart shows the significance of this dissertation research in evidencing the impact of accessibility and usability information on the purchasing decisions of people with and without disabilities. The benefits of this information for PwD have been demonstrated previously. This study demonstrated the benefits of this information for PwoD and aimed to facilitate the evaluation of medical device accessibility and usability evaluation. This, in turn, will lead to better product design by the manufacturers and more informed purchasing decisions for all, thus leading to reduced disparities in health outcomes between PwoD and PwD caused by the unavailability of usability and accessibility information.

**Detailed Description.** The image presents a flowchart structured around a central theme of preferences and outcomes related to medical device accessibility and usability. At the top, there is a primary rectangular section divided into two smaller, side-by-side rectangles. The rectangle on the right is outlined in solid black with a white background, containing text that details preferences of People with Disabilities (PwD), emphasizing their inclination towards accessible information and higher usability for medical devices. This rectangle is marked with a green tick on its upper right corner, signifying affirmation. Adjacent to it on the left, the other rectangle, with a yellow background, mirrors the right one in content, indicating similar preferences for People without Disabilities (PwoD), but it is distinguished by an orange question mark in its upper right corner, indicating uncertainty or consideration.

Descending from these two rectangles are two vertical branches, each initiating with its own statement box, suggesting a cause-and-effect pathway. On the right, the pathway begins with a box that addresses the perspective of medical device manufacturers and leads to an outcome of improved accessibility of medical devices. On the left, the branch starts with an increase in the prevalence of medical device accessibility and usability evaluation, descending to a box that signifies the increased availability of such information, which then leads to informed purchasing decisions for both PwD and PwoD.

The two vertical branches of the flowchart ultimately join, merging their narrative into a final box at the bottom of the image. This box has a dotted outline and is labeled with a statement that suggests a reduction in health outcome disparities between PwD and PwoD, with a red bullseye target icon in its upper right corner. A dashed line connects the first boxes of the two branches, highlighting a reciprocal relationship between the availability of information and the motivation of manufacturers.

## **Figure 2**

**Brief Description.** Bar graph shows the commonness of medical devices used at home, as identified by PwD on the survey.

**Essential Description.** This bar graph shows what medical devices PwD use at home, as identified in the survey. The most common home medical devices are Blood Pressure Monitors and Blood Glucose Monitors, based on the data from the participants of the study.

**Detailed Description.** The image displays a horizontal bar graph detailing the count of various medical devices. Each bar represents a different type of medical device, with the length of the bar corresponding to its count. The background is white with faint grey vertical stripes that aid in reading the exact count of each item.

The bars are a deep shade of purple, and the graph is organized in descending order, starting from the top with the device having the highest count. At the top, Blood Pressure Monitors are the most numerous, totaling 46, closely followed by Blood Glucose Monitors at 44. PAP Therapy Devices and Weight Scales are next, with counts of 33 and 32, respectively. Other Devices are represented with a count of 31. Midway through the graph, the counts begin to decrease. Thermometers are listed with 24, Pulse Oximeters with 16, and Insulin Pumps with 14. Injection Syringes are less frequent, at a count of 8. COVID-19 Home Tests and Inhalers are equally represented, each with a count of five. Toward the bottom of the graph, the counts continue to decrease. Neurostimulators and Hearing Aids both have a count of four, similar to Smartphones/Tablets. Oxygen Compressors, Wheelchairs, and Medication Dispensers are the least numerous, each with a count of three.

### **Figure 3**

**Brief Description.** Bar graph shows the commonness of medical equipment used at healthcare facilities, as identified by PwD on the survey.

**Essential Description.** This bar graph shows what medical devices PwD use at healthcare facilities, as identified in the survey. The most commonly identified equipment were in the diagnostic category, based on the collected data from the participants.

**Detailed Description.** The image displays a horizontal bar chart with five categories, each represented by a distinct green bar against a white background with vertical gray lines to assist in quantifying. The chart is titled at the horizontal axis as 'Count.'

At the top of the chart, the 'Diagnostic' category has the longest bar, extending to the value of 43, indicating it has the highest count among the categories listed. Below it, the 'Positioning' category has a bar that reaches up to 12. The next two categories, 'Therapeutic' and 'Information Technology,' have equal counts, with their bars both stopping at the 9 mark. The category labeled 'Other Devices' has the shortest bar, indicating a count of seven.

Each bar's length is proportional to the count it represents, providing a visual representation of the quantity associated with each category. The bars are clearly labeled to the left with the category they represent, and the corresponding count is indicated at the end of each bar on the right.

**Figure 4**

**Brief Description.** Figure shows the components and subcomponents of DCE design to identify various study design factors.

**Essential Description.** This figure shows the breakdown of factors involved in the systematic design of a DCE, based on literature. The purpose of this breakdown is to provide a brief overview of the DCE components.

**Detailed Description.** The image presents a structured flowchart outlining the components involved in DCE design. The chart is organized into a series of connected boxes that branch out from the leftmost rectangular box labeled 'DCE Design,' which acts as the starting point. All boxes have black borders and are connected with black lines. The lines connecting the boxes gradually grow thinner from left to right, to visually denote the progression from general to specific details. The general layout is hierarchical and logical, moving from general to specific as one reads from left to right, illustrating a systematic approach to designing a DCE.

From the 'DCE Design' box, three primary branches extend to the right. The first branch leads to a box titled 'Establishing Attributes & Levels.' The second branch connects to a box titled 'Choice Tasks.' The third branch extends to another box called 'Experimental Design,' indicating another key phase of the experiment design. The 'Experimental Design' box further branches into five other boxes. The first box is 'Analytical Model,' followed by 'Main vs Interaction Effects' below, then 'Labeled vs Unlabeled,' 'Structure,' and lastly 'Internal Validity Assessment.' The fourth box, 'Structure,' is connected to three other boxes. From top to

bottom, these boxes are titled “Number of Choice Tasks & Blocking,” “Type of Choice Matrix Design,” and “Attribute-Level Balancing.”

### **Figure 5**

**Brief Description.** Image shows a collage of four screenshots from prominent e-tailers and highlights the most prevalent attributes of BPMs on the product pages.

**Essential Description.** This image aims to show that Cost, Brand, and User Rating are prominent product attributes that are present on established e-tailer websites. The purpose of this image is to emphasize the appropriateness of DCE attribute identification in the dissertation study.

**Detailed Description.** The image is a collage of four screenshots from various e-tailer websites, showcasing blood pressure monitors. Each screenshot is highlighted with colored markers emphasizing three main pieces of information: Cost, Brand, and User Rating. These markers have their own unique colors, and are all arrow shaped. The marker for cost is light green on a black background, brand is pink on black, and user rating is yellow on black. The top banners of the e-tailer websites are represented in color, to help distinguish the visual branding of various e-tailers. However, to reduce visual complexity and help the attribute markers stand out, the contents of pages appear desaturated and grayscale.

The top left screenshot is from Amazon, showing a blood pressure monitor with the brand name clearly displayed and the cost listed below the product image. The user rating is highlighted and appears to be positioned above the product title. The Amazon interface is recognizable at the top with the navy blue, orange, black, and white colors for various elements. The page contents appear in black font on white background.

The top right screenshot comes from the Walgreens website, featuring another blood pressure monitor. Similar to the first, the brand name is highlighted at the top, with the user

rating just below it. The cost is prominently displayed next to the product image. This interface has a mix of white, red, and blue elements, with a dark text on a light background for the contents.

The bottom left screenshot is from the Best Buy website, where a different blood pressure monitor is presented. As with the others, the brand name is marked at the top, and the user rating is indicated below the product image. The cost is highlighted and shown below the product description. The Best Buy website banner has a blue header with some yellow elements.

Finally, the bottom right screenshot is taken from the Walmart website, displaying yet another blood pressure monitor. The brand name is at the top, the user rating just beneath the product, and the cost at the bottom. The Walmart site features a light blue header, and some light orange elements.

## **Figure 6**

**Brief Description.** Boxplot shows the distribution of BPM cost across different e-tailers to help identify proper levels for the Cost attribute in DCE.

**Essential Description.** The boxplot in this figure shows the variations in cost for BPMs catalogued from various e-tailers. The purpose of this plot is to compare the BPM cost ranges across e-tailers and demonstrate the difficulty of identifying a shared range to determine realistic levels for the Cost attribute in the dissertation study.

**Detailed Description.** The image features a horizontal box plot graph displaying the cost distribution of BPMs catalogued from four different e-tailers: Amazon, Best Buy, Walmart, and Walgreens. The graph's background is white with vertical grey lines that help delineate the cost scale which runs along the bottom axis from \$0 to \$190. All plots are presented in green.

Each e-tailer is represented by a horizontal box plot on the graph. The box plots consist of a rectangular 'box' that shows the interquartile range, a line within the box that represents the median cost, and 'whiskers' that extend from the box to the minimum and maximum values within a certain range. Outliers are represented as individual dots that fall outside of the whiskers.

The 'Amazon' box plot is positioned at the top, showing a price range between approximately \$16 to \$74, with the median cost being around \$33. The 'Best Buy' plot beneath it has a larger box, indicating a high variability in cost range from \$35 to \$150, with the median approaching \$83. The 'Walmart' plot shows a range between \$13 and \$90, with the median around \$32. The 'Walgreens' box plot at the bottom shows a range of costs between \$16 and \$114, with the median cost at \$60.

### **Figure 7**

**Brief Description.** Histogram shows the cumulative distribution of User Rating for all BPMs catalogued from Amazon, Best Buy, Walmart, and Walgreens.

**Essential Description.** This histogram shows the distribution of User Ratings for all the BPMs catalogued from Amazon, Best Buy, Walmart, and Walgreens. The purpose of this figure is to demonstrate how the User Rating levels were realistically determined in the dissertation research.

**Detailed Description.** The image depicts a histogram with a series of vertical bars representing the distribution of user ratings across four e-tailers. The histogram is set against a light gray background, with the bars colored in a vibrant orange with black borders. The x-axis of the histogram is segmented into intervals that represent user rating scores, indicated as star ratings ranging from 1.0 to 5.0, with each interval captured within brackets, such as [1.0, 1.2], [1.2, 1.4], and so on, up to [4.8, 5.0]. The y-axis is a numerical scale indicating the frequency of ratings within each interval.

The bars vary in height, correlating with the number of user ratings that fall within each user rating interval. The majority of the bars are short, indicating fewer ratings in those categories. However, there is a prominent peak in the histogram where two bars stand significantly taller than the others, in the [4.2, 4.4] and [4.4, 4.6] intervals, suggesting a high concentration of ratings within the range. This peak is surrounded by smaller yet substantial bars, showing a concentration of ratings that tend to skew towards the higher end of the scale.

### **Figure 8**

**Brief Description.** Composite of six bar charts to compare the sociodemographics of the qualifying participants available on Prolific (PwOD above the age of 50) against the US population data, to demonstrate access potential to a representative sample.

**Essential Description.** This composite graph compares the characteristics of PwOD above the age of 50 between Prolific and the US population. This comparison is made on sociodemographic information such as ethnicity, sexual orientation, gender, employment status, housing, and annual household income. The purpose of this figure is to demonstrate the potential for recruiting a representative sample, and in all instances the comparisons confirm this potential.

**Detailed Description.** The figure is a composite of six bar charts comparing various demographic characteristics of PwOD aged 50 and older. There are three rows in total. The first row contains two comparisons, the second row contains three, and the last row contains one comparison. Each comparison shows the data from eligible Prolific participants on the left, and the US population data on the right.

In the first bar chart on the top left is titled "Ethnicity (PwD, 50+)." Each ethnic group's representation is shown as a percentage of the total, with blue for White, orange for Black, gray for Asian, yellow for Mixed, and light blue for Other. The comparison is made between Prolific eligible participants and data from the American Community Survey 2022. The blue bars, representing the White ethnicity, are the tallest by a large margin. The rest of the ethnicities hover at around or below 10% of the total.

The second chart, "Sexual Orientation (Mixed Populations)," compares the sexual orientation distribution of the Prolific dataset specifically for PwD over 50 with Gallup '22 data representing all US adults. The orientations shown are Heterosexual in green, Homosexual in blue, and Bisexual in yellow. The green bars are the tallest by a large margin, compared to the other bars. The other bars are all very close to 0%, however, they are not zero.

The third chart is "Gender (PwD, 50+)" which compares the gender distribution between men (in blue) and women (in orange). Comparison is made between Prolific and ACS 2022 data. Orange is taller than blue in both.

The fourth chart, "Employment (PwD, 50+)," compares employment status categories: Employed in blue, Unemployed in orange, and Not working in gray. Comparison between Prolific and ACS 2022 shows blue is the tallest bar, gray is the second tallest, and orange is the shortest bar for both.

The fifth chart is "Housing (PwD, 50+)," comparing types of housing such as Owned in blue, Rented in orange, and Other in gray. The comparison between Prolific and ACS 2022 shows that blue is the tallest, orange is the second tallest with a large margin compared to blue, and gray is the shortest and closer to 0% in both.

The final bar chart at the bottom, "Annual Household Income (PwD, 50+)," displays a wide range of income brackets from "Less than \$10K" to "More than \$150K." This chart compares the income distribution between Prolific and the ACS 2022 data, showing the percentage of the population falling within each income bracket. All income brackets below \$100K range between 0 to 10% with a similar distribution in comparison. Between \$100K and

\$150K, Prolific and ACS 2022 bars are close to each other, however, above \$150K the Prolific bar is shorter than the ACS 2022 bar by about 10%.

**Figure 9**

**Brief Description.** Screenshot shows an example of the Automatic Fraud Detection stage outcome, which is used in manual review to assess severity of the violations.

**Essential Description.** This screenshot shows the output of automatic fraud detection, which is then manually inspected to determine the severity of the violations committed by the participants and decide whether the violations are severe enough to warrant exclusion from the data analysis. The purpose of this figure is to present an example of the data quality examination rigor in the dissertation study to ensure maximum accuracy and quality of the findings.

**Detailed Description.** The image shows a screenshot of a table with the following column headers: Prolific ID, Violation Categories, and Violations. The IDs in the Prolific ID column are blurred for privacy. The right side of the image also fades to white to show that the right side of the table is excluded for brevity. The bottom of the image is also abruptly cut to show that only the top portion of the table is being shown.

The text is black in all cells. The backgrounds of rows are color coded to represent severity of violations. Red indicates severe violations that require exclusion from the analysis, orange denotes more than one mild violation that may not require exclusion, yellow represents a single mild violation that does not require exclusion, and green shows the absence of any violations, or the existence of negligible ones. The first row is red, the next seven rows are orange, and the next 11 rows are yellow. The rest of the rows are green.

**Figure 10**

**Brief Description.** Image shows the transformation of the wide format data from Qualtrics into long format data required for data analysis.

**Essential Description.** This image shows an example of how the wide format data obtained from Qualtrics was transformed into the long format data required by the data analysis module. The purpose of this figure is to demonstrate how this transformation took place and how the representation of the data changed.

**Detailed Description.** The image consists of two screenshots of a small portion of the survey data with different representations. The screenshot at the top shows the first six product selections of a participant, and each selection is shown with one cell filled with information about the product that was selected. These cells have a green background, and the rest of the table has a light-yellow background. Above this table, there is a red double arrow spanning the length of the table to highlight that the data is in the wide format. The text above this arrow reads “Qualtrics Output – Wide Format” in large black font.

The screenshot at the bottom shows the same selections from the same participant, however, the presentation is different and additional context has been added. In this table, instead of presenting the entire selection in one cell, each specific feature of the product has a dedicated column. In addition, the options that were present but not selected are also shown in additional rows for each choice task. All cells have a light-yellow background, except for cells that contain the selection information, which are highlighted in green, corresponding to the selection data from the wide format table. A purple double arrow is shown that covers the

height of the table to represent long format transformation of the same data. A label next to this arrow reads “MNL Analysis Input – Long Format” in bold black font.

A large arrow goes from the top table to the bottom table. The color of this arrow changes in a gradient from red to white to purple to denote the transformation of data representation.

## **Figure 11**

**Brief Description.** Flowchart shows a summary of the eight distinct stages of data collection and analysis from study design to data analysis.

**Essential Description.** This flowchart shows the components of the data collection and analysis pipeline. The purpose of this figure is to provide an overview of all the stages involved in the design and analysis of the dissertation research.

**Detailed Description.** The image shows a flowchart from design and implementation to data collection, quality control, and analysis. The flowchart format and the use of symbols convey the sequence of actions and the analytical tools utilized at each stage. Each stage is represented with chevron that includes the title of the stage in large black font. Below each chevron, a brief list of activities in that stage is provided in a rectangle, with black font on white background. On the top left corner of each chevron, a logo represents what software solution was primarily used in that stage.

The first stage is marked with the "Excel" logo and the chevron has a green background and is titled "Experimental Design Parameters." This stage includes tasks such as BPM (Blood Pressure Monitor) Market Research and Literature Review. Following this is the "Choice Matrix Design" stage with a light blue chevron and marked with an "R" symbol, signifying the use of the R programming language. It lists tasks like D-Efficient Method, D-error evaluation, Attribute Level Balance, and Power Analysis Simulation. Next is "Preparation for Survey Implementation" on a yellow chevron and marked by the "Python" logo. The only task listed underneath is Presentation Optimization. The fourth stage is "Data Collection" on a light grey chevron and marked with the Qualtrics (XM) symbol next to the Prolific logo. The tasks are listed as:

obtaining IRB Approval, Prolific Recruitment for participants, and the Qualtrics Survey Setup. Following data collection, there is a stage for "Automatic Fraud Detection & Data Quality Assessment," indicated by the "Python" icon on a yellow chevron. The tasks involve checking Data Consistency (Across Sources), Attention Checks, Effort Checks, Timings, and Comprehension Evaluation. After this stage is "Manual Review of Data Quality" on a green chevron, marked with the "Excel" logo. The tasks in this stage include Violations Review, Violations Severity Evaluation, Justifiability Evaluation, Identifying Rejections, and Data Quality Assessment. The subsequent stage is "Internal Validity Report & Preprocessing for Analysis," on a yellow chevron and marked with the "Python" logo. The tasks are listed as: wide to long format conversion, useful column retention, total duration report, Dominant Alternatives Report, and Choice Consistency report. The final stage is "Data Analysis" on a light blue chevron, marked with the "R" logo, indicating the use of R for this phase. The listed tasks are Model Estimation, Performance Documentation, and a Posteriori Analyses.

## **Figure 12**

**Brief Description.** Plot shows the decrease in parameter standard error simulations as the sample size increases in Phase I.

**Essential Description.** This plot shows that in Phase I, as the sample size increases, the simulated standard error for each study parameter decreases. By including this plot, the relationship between sample size and simulated parameter standard errors is examined. This plot serves to depict this relationship, and denote that beyond a certain sample size, the improvement in accuracy is not substantial anymore.

**Detailed Description.** The image depicts a plot of the relationship between sample size and standard error for various coefficients. The x-axis represents the sample size, which ranges from 0 to 20, and the y-axis represents the standard error, ranging from 0 to 0.6.

Different colored diamond-shaped points represent the standard errors of different coefficients as the sample size increases. The coefficients are identified in a legend titled "Coefficient," located on the right side of the graph. The coefficients include "Brand Omron," "Cost," "Usability Accessibility Score High (8/10)," "Usability Accessibility Score Low (4/10)," "Usability Accessibility Score Medium (6/10)," and "User Rating."

The plot points for "Brand Omron" are colored orange, "Cost" in brown, "Usability Accessibility Score High (8/10)" in green, "Usability Accessibility Score Low (4/10)" in turquoise, "Usability Accessibility Score Medium (6/10)" in blue, and "User Rating" in pink. The standard error value is at the highest for the smallest sample size, and as the sample size increases, all points gradually decrease toward zero. First, for sample sizes between two and ten, the gradual

decrease is rapid. However, these points appear to form flat horizontal lines as the sample size exceeds 15.

**Figure 13**

**Brief Description.** Horizontal bar chart shows the summary of violations in Phase I of the dissertation study after data quality inspection.

**Essential Description.** This figure shows the violations identified in Phase I of the dissertation study and helps provide a general overview of data quality.

**Detailed Description.** The image contains a horizontal bar chart that represents counts of various types of discrepancies found in the Phase I survey data. There are five categories of discrepancies displayed on the y-axis, which include Multiple Entry, Age Discrepancy, Disability Discrepancy, Ethnicity Discrepancy, Country Discrepancy, and State Discrepancy.

The x-axis represents the count of discrepancies, with a scale ranging from 0 to 20. Each category has a corresponding horizontal bar indicating the number of discrepancies detected in each category. The 'Multiple Entry' category has a diagonally striped bar in red that indicates two counts. The 'Age Discrepancy,' 'Disability Discrepancy,' and 'Country Discrepancy' categories all show a count of zero. The 'Ethnicity Discrepancy' category has the longest bar in green, showing three counts. Lastly, the 'State Discrepancy' category shows a count of one with a green bar. The background of the chart is light gray.

**Figure 14**

**Brief Description.** Composite of horizontal bar charts and US map shows the sociodemographics of the participants in Phase I and their geographical distribution across the US.

**Essential Description.** This composite figure shows the sociodemographic characteristics of the sample in Phase I of the dissertation research and denotes an overall representative sample.

**Detailed Description.** This figure consists of two main sections. The top section shows four horizontal bar charts in a two-by-two arrangement, and the bottom section shows a map of the US with large orange dots scattered across. In the top section, all bar charts are on a light gray background with a black font, and the x-axis ranges from 0 to 18. The chart on the top left is titled "Gender – Count" and shows two blue bars of the same size for Male and Female, each with nine participants. The chart on the top right is titled "Age Range – Count" and using yellow bars, shows that the 50-54 age range has the highest count at 11, followed by the 60-64 age range with four. The 55-59 and 65-69 age ranges have one individual each. The bar chart on the bottom left of this section is titled "Race – Count" with red bars. The longest bar represents White individuals, totaling 13, followed by Other and Black categories, each with 2. American Indian or Alaska Native and Asian categories both have one individual, and there are no individuals in the Native Hawaiian or Pacific Islander category. The last bar chart in this section, on the bottom right, uses green bars and is titled "Annual Household Income Range – Count." The \$50K-\$75K income range has the highest count at 5, followed by the \$100K-\$150K and

\$25K-\$50K ranges, each with four. The \$75K-\$100K range has 3, and both the <\$25K and Undisclosed categories have 1 individual each.

The bottom section of this figure shows a colored map of the US, overlaid with large orange circles. The majority of the orange circles are on the middle and right sections of the map, and a few are on the left side.

**Figure 15**

**Brief Description.** Stacked horizontal bar chart showing the distribution of participants' priority rankings of study attributes in Phase I.

**Essential Description.** This figure aims to visualize how the participants explicitly ranked the importance and priority of study attributes in their decision-making, in Phase I. The purpose of this figure is to show how most participants ranked User Rating and Usability/Accessibility Score as the most important factors in their purchasing decisions.

**Detailed Description.** The figure shows a horizontal stacked bar chart. The x-axis ranges from 0 to 18, and the y-axis shows the assigned rank with "First" at the top and "Fourth" at the bottom. User Rating is shown in yellow, Usability/Accessibility Score in pink, Cost in light green, and Brand in dark purple. For the first rank in importance, ten participants selected User Rating, six selected Usability/Accessibility Score, and two selected Cost. For the second rank in importance, five participants selected User Rating, nine selected Usability/Accessibility Score, three selected Cost, and one selected Brand. For the third rank, 3 selected User Rating, 2 selected Usability/Accessibility Score, 6 selected Cost, and 7 selected Brand. For the bottom rank, one selected Usability/Accessibility Score, 7 selected Cost, and 10 selected Brand.

**Figure 16**

**Brief Description.** Box plot shows the amount of time participants spent on the survey.

**Essential Description.** This figure shows how much time participants spent on the survey and identifies the median and outliers. Based on this plot, the survey did not take a substantial amount of time from the participants, and the majority spent between about 10 to 20 minutes.

**Detailed Description.** The horizontal box plot is blue on a light gray background. Underneath the box plot, individual data points are also shown in blue. The x-axis is labeled "Total Survey Duration (minutes)" with a range from 0 to 80 minutes. The middle rectangle ranges from about 9 minutes to 21 minutes, with the median line near 11. The whiskers extend to 7 on the left and about 23 on the right. There are 2 blue dots at about 45 minutes and 76 minutes that are outliers.

### **Figure 17**

**Brief Description.** Plot shows the decrease in parameter standard error simulations as the sample size increases in Phase II.

**Essential Description.** This plot shows that in Phase II, as the sample size increases, the simulated standard error for each study parameter decreases. By including this plot, the relationship between sample size and simulated parameter standard errors is examined. This plot serves to depict this relationship, and denote that beyond a certain sample size, the improvement in accuracy is not substantial anymore.

**Detailed Description.** The image depicts a plot of the relationship between sample size and standard error for various coefficients. The x-axis represents the sample size, which ranges from 0 to 200, and the y-axis represents the standard error, ranging from 0 to 0.6.

Different colored diamond-shaped points represent the standard errors of different coefficients as the sample size increases. The coefficients are identified in a legend titled "Coefficient," located on the right side of the graph. The coefficients include "Brand Omron," "Cost," "Usability Accessibility Score High (8/10)," "Usability Accessibility Score Low (4/10)," "Usability Accessibility Score Medium (6/10)," and "User Rating."

The plot points for "Brand Omron" are colored orange, "Cost" in brown, "Usability Accessibility Score High (8/10)" in green, "Usability Accessibility Score Low (4/10)" in turquoise, "Usability Accessibility Score Medium (6/10)" in blue, and "User Rating" in pink. The standard error value is at the highest for the smallest sample size, and as the sample size increases, all points gradually decrease toward zero. First, for sample sizes between 10 and 100, the gradual

decrease is rapid. However, these points appear to form flat horizontal lines as the sample size exceeds 150.

**Figure 18**

**Brief Description.** Horizontal bar chart shows the summary of violations in Phase II of the dissertation study after data quality inspection.

**Essential Description.** This figure shows the violations identified in Phase II of the dissertation study and helps provide a general overview of data quality.

**Detailed Description.** The image displays a horizontal bar chart that represents the count of various types of discrepancies or issues identified in a dataset. Each bar represents a different type of discrepancy, and the length of the bar indicates the count of occurrences for each type. The categories of discrepancies are listed on the y-axis, including Multiple Entry, Two Attention Checks Fail, Age Discrepancy, Commitment Check Fail, Country Discrepancy, Disability Discrepancy, Ethnicity Discrepancy, Exceptionally Fast Submission, Gender Discrepancy, Instruction Comprehension Fail, Mindset Assumption Fail, One Attention Check Fail, State Discrepancy, and Suspected Bot.

The x-axis represents the count of each type of discrepancy, with a scale ranging from 0 to approximately 175. The bars are color-coded, with the majority being green to indicate mild violations, while the bars for Instruction Comprehension Fail, Mindset Assumption Fail, and One Attention Check Fail are orange to indicate important violations. Multiple Entry and Two Attention Check Fail are diagonally striped in red, indicating severe violations that were excluded from the analysis.

Multiple Entry and Two Attention Check Fails have counts of four and one, respectively. Instruction Comprehension Fail, Mindset Assumption Fail, and One Attention Check Fail have

24, 14, and 2 counts respectively. Age Discrepancy has 4 counts, Ethnicity Discrepancy also has 4, Gender Discrepancy has 1, and State Discrepancy has 17. The rest are all zero.

**Figure 19**

**Brief Description.** Composite of horizontal bar charts and US map shows the sociodemographics of the participants in Phase II and their geographical distribution across the US.

**Essential Description.** This composite figure shows the sociodemographic characteristics of the sample in Phase I of the dissertation research and denotes an overall representative sample.

**Detailed Description.** This figure consists of two main sections. The top section shows four horizontal bar charts in a two-by-two arrangement, and the bottom section shows a map of the US with large orange dots scattered across. In the top section, the chart on the top left is titled "Gender - Count," and there are two yellow bars of nearly equal length. The bar representing females shows a count of 88, while the bar for males is slightly shorter, with a count of 86. The top right chart with purple bars and titled "Age Range - Count" depicts various age ranges with the 50-54 age group having the highest count at 57. The counts decrease progressively with higher age groups, with the 55-59 age group at 45, and the 60-64 at 29. The lowest counts are in the oldest age groups, with one each for those aged 75-79 and 80-84, and none for the >85 age group. The bottom left chart with pink bars and titled "Race - Count" shows a majority of participants identified as White, with a count of 159. This is followed by a smaller count for Black participants at 12, and the lower counts are for Asian and Other, both at 3. American Indian or Alaska Native has a count of 1, and there are no participants listed as Native Hawaiian or Pacific Islander. The bottom right chart with green bars titled "Annual Household Income - Count" illustrates a relatively even distribution among income ranges, with

the highest count at 35 for the \$50K-\$75K range. The \$25K-\$50K, \$75K-\$100K, \$100K-\$150K, and >\$150K ranges each have counts ranging from 31 to 32. There are 9 participants in the <\$25K income range, and 1 participant's income is undisclosed.

The bottom section of this figure shows a colored map of the US, overlaid with large pink circles. Most of the circles are on the right side of the map. The middle and left sections of the map, however, are also covered in circles. The area in the top middle is empty around the Dakotas, Wyoming, and Nebraska.

**Figure 20**

**Brief Description.** Stacked horizontal bar chart showing the distribution of participants' priority rankings of study attributes in Phase II.

**Essential Description.** This figure aims to visualize how the participants explicitly ranked the importance and priority of study attributes in their decision-making, in Phase II. The purpose of this figure is to show how most participants ranked User Rating and Usability/Accessibility Score as the most important factors in their purchasing decisions.

**Detailed Description.** The figure shows a horizontal stacked bar chart. The x-axis ranges from 0 to 174, and the y-axis shows the assigned rank with "First" at the top and "Fourth" at the bottom. User Rating is shown in red, Usability/Accessibility Score in yellow, Cost in light blue, and Brand in dark blue. For the first rank in importance, 109 participants selected User Rating, 25 selected Usability/Accessibility Score, 25 selected Cost, and 15 selected Brand. For the second rank in importance, 41 participants selected User Rating, 71 selected Usability/Accessibility Score, 31 selected Cost, and 31 selected Brand. For the 3<sup>rd</sup> rank, 19 selected User Rating, 48 selected Usability/Accessibility Score, 49 selected Cost, and 58 selected Brand. For the bottom rank, 5 selected User Rating, 30 selected Usability/Accessibility Score, 69 selected Cost, and 70 selected Brand.

**Figure 21**

**Brief Description.** Graph shows the six-class LC-MNL analysis parameter estimates and their variations, to show the class differences in utility assignment in Phase II.

**Essential Description.** This graph shows how the six classes identified by the LC-MNL analysis differ in the amount of utility they assign to various attributes of the DCE in Phase II. The objective of this figure is to show that six distinct classes were identified in the participants in Phase II, each with their own unique utility model.

**Detailed Description.** The graph has labels on the left for class specific coefficients. On the right of the graph, the coefficient values and their variations within a class are represented with a black dot and a bar extending to left and right, respectively. The figure is vertically sectioned into 6 slices, each containing class-specific data for one coefficient. The top section shows the class differences for Usability/Accessibility Score High, the second section shows Usability/Accessibility Score Medium, the third shows Usability/Accessibility Score Low, the fourth shows User Rating, the fifth shows Brand – Omron, and the sixth shows Cost. Class 1 in all sections is represented with a red line, Class 2 with blue, Class 3 with green, Class 4 with purple, class 5 with orange, and Class 6 with gray.

In the top section for Usability/Accessibility Score High, Class 1's parameter varies between 4 and 6, Class 2 between 6 and 9, Class 3 between 1 and 2, Class 4 between 2 and 4, Class 5 between 2 and 5, and Class 6 between 2 and 4. In the section for Usability/Accessibility Score Medium, Class 1's parameter varies between 3 and 4, Class 2 between 2 and 4, Class 3 between 0 and 1, Class 4 between 0.5 and 1.5, Class 5 between 1 and 4, and Class 6 between 1 and 2. In the section for Usability/Accessibility Score Low, Class 1's parameter varies between -1

and 0.5, Class 2 between -2 and -0.5, Class 3 between -0.5 and 1, Class 4 between -5 and -2, Class 5 between -1 and 0.5, and Class 6 between -1 and 0.5.

In the section for User Rating, Class 1's parameter varies between 2 and 3, Class 2 between 1 and 2, Class 3 between 0.5 and 1.5, Class 4 between 3 and 4, Class 5 between 4 and 8, and Class 6 between 3 and 4. In the section for Brand, the parameters for all classes are between 0 and 1, except for Class 6 which is between 3 and 4. In the bottom section for Cost, all parameters are near zero on the negative side, with very minimal variations.

## **Figure 22**

**Brief Description.** Survey data that differed with statistical significance between the six classes, presented in a collection of stacked and standard bar charts.

**Essential Description.** This collection of charts shows the survey items that emerged as differentiating between classes with statistical significance. Causation may not be assumed; however, this figure depicts how other items on the survey differed among the six classes identified in the LC-MNL analysis. The purpose of this figure is to provide insights into the distinction between the six classes to complement the data from class-specific utility models.

**Detailed Description.** This figure consists of five stacked bar charts and one standard bar chart, in three rows and two columns. Each chart compares the responses to one survey question denoted in the title, across the 6 identified classes.

The top left bar chart is stacked and titled “Familiarity with Accessibility & Disability (% Total Class).” In this chart, red represents “Not familiar at all,” yellow is for “Slightly familiar,” green is for “Moderately familiar,” light blue is for “Very familiar,” and purple is for “Extremely familiar.” For Class 1 yellow is the largest. For Class 2, purple does not exist, and the majority of the bar is in blue and green. For Class 3, yellow, green, and blue make up most of the chart and are almost in equal amounts, and purple and red are present with an almost equal but small amount. For Class 4, green covers the largest section, and blue is the smallest compared to other classes. For Class 5 purple does not exist, and blue and green make up most of the bar and almost in equal amounts. For Class 6, red and purple have the largest share compared to all other classes.

The top right bar chart is standard with red bars and is titled “Average Total Survey Duration (minutes).” Class 1 has the tallest bar with close to 16 minutes. Class 2 is approximately 11 minutes. Class 3 has the shortest bar with about 8 minutes. Class 4 is nearly 12 minutes. Class 5 is nearly 14 minutes. Class 6 is around 13 minutes.

The middle-left chart is stacked and titled “Satisfaction with the Survey Length (counts).” Dark red represents “Extremely dissatisfied,” red means “Somewhat dissatisfied,” yellow means “Neither satisfied nor dissatisfied,” light blue means “Somewhat satisfied,” and dark blue means “Extremely satisfied.” Class 1 has the second largest dark blue section, and no red or dark red. For Class 2, most of the bar is covered in Dark and light blue, and this class has the largest red section. Class 3 has the largest light blue section, and no red or dark red is present. Class 4 has the largest dark blue section, and ties with Class 6 in having the largest dark red section, although the dark red is a very small portion of the bar. Class 5 has the largest dark and light blue section combined, with only a sliver of yellow and no light or dark red. Class 6 has the largest yellow compared to the rest, and ties with Class 4 in the size of the dark red section.

The middle-right chart is stacked and titled “Average Attribute Rankings (% Total Class).” Cost is shown in purple, Brand in pink, User Rating in yellow, and Usability/Accessibility Score in green. For Class 1, User Rating and Usability/Accessibility Score almost tie in having the highest rank, and Cost and Brand almost tie in having the lowest ranks. Class 2 ranked Usability/Accessibility Score the highest, both within class and between classes. Class 3 assigned a better rank to Cost compared to all other classes, however, prioritized User Rating above Cost overall, by a small margin. Class 5 assigned the highest rank to User Rating compared to all

other classes, and within the class. Class 6, both within class and between classes, assigned the highest rank to Brand.

The bottom left chart is stacked and titled “Importance of Usability/Accessibility Score in choices (% Total Class).” Red denotes “Not at all important,” yellow is for “Slightly important,” green is for “Moderately important,” blue is for “Very important,” and purple is for “Extremely important.” Class 1 has a very small red section compared to other classes. The purple section is the second largest in this class, and blue ties with Classes 2 and 4 as the largest compared to other classes. Class 2 has the smallest green and the largest purple section compared to other classes. Class 4 has no purple, and ties with Classes 5 and 6 at having the largest red section. Class 5 has the largest green section and the smallest red section. In Class 5 purple is absent, green and yellow cover the majority of the bar and almost in equal amounts, and red is almost the largest, compared to other classes except Classes 3 and 6. Class 6 has the largest yellow section, and ties with Classes 3 and 5 in having the largest red.

The bottom right chart is stacked and titled “Stability Validity Test (% Total Class).” Blue is used to denote consistency, and red for inconsistency. For Class 1, almost 70% of the bar is blue. Class 2 has the second largest blue section at just below 90%. Class 3 has the largest red section at around 65%. Class 4, similar to Class 1, is around 70% blue. Class 5 is around 75% blue. Class 6 has the largest blue section at around 90%.

**Figure 23**

**Brief Description.** Box plot shows the total survey duration in minutes, mostly between 5 to 15 minutes.

**Essential Description.** This plot is used to illustrate the total survey duration time and identify the outliers. The majority of respondents spent between 5 to 15 minutes on the survey.

**Detailed Description.** The image features a horizontal box plot overlaid with individual data points, representing the total survey duration in minutes. The x-axis is labeled "Total Survey Duration (minutes)" and is scaled from 0 to 55 minutes. The plot is in dark blue against a light gray background. Individual data points, represented by dark blue diamond shapes, are scattered primarily below the 25-minute mark. The box ranges from 9 to 15 minutes, with the median line around 11 minutes. The left whisker extends left to 5 minutes, and the right whisker extends to about 23 minutes. About 6 outliers are visible beyond the endpoint of the right whisker.

**Figure 24**

**Brief Description.** Flowchart demonstrates the multifaceted benefits of Usability/Accessibility Assessment and the beneficiaries, as posited by the MAUVE hypothesis.

**Essential Description.** This figure intends to depict how Usability/Accessibility Assessment benefits many stakeholders, including PwOD, PwD, and medical device manufactures. Overall, this figure illustrates the MAUVE hypothesis, suggesting that merely by evaluating and presenting Usability/Accessibility Information, many stakeholders will benefit, and the health outcome disparities will be reduced between PwD and PwOD.

**Detailed Description.** The figure depicts a flowchart with a series of interconnected steps outlining a circular process. In this figure, dotted green arrows denote a connection that has been made, or can be logically assumed to exist, which requires further reinforcement. Dotted brown arrows denote connections that have not yet been established.

The flow starts with three groups represented in separate circles with black text on grey background: Designers & Manufacturers, Invested Experts & Stakeholders, and Consumers. The process begins with these groups contributing to assessment of medical devices for usability and accessibility, which is illustrated by a green dotted arrow connecting to a central rectangular node titled "Usability/Accessibility Assessment" with mauve background and white font. One pathway emerges downward from this node, as a dotted green arrow, connecting to a box titled "Usability/Accessibility Information" on a mauve background with white font. From this box, a green dotted arrow goes right, terminating in a box titled "Informed Purchase Decisions" with dark red background and white font. From this box, another dotted green

arrow terminates in an orange box on the right, titled in bold black font “Reduced Disparities Between PwD and PwoD.”

From the “Informed Purchase Decisions” box discussed earlier, two dotted green arrows emerge downward, connecting to two boxes with dark red backgrounds and white font, with PwD and PwoD written inside. Two dotted green arrows exit these boxes and converge downward into the dark red box titled “Satisfaction” in white font. From this box, a brown arrow goes left to a dark blue box that says “Profit Increase” in white font. Next, this box is connected to another similarly colored box on its left with a dotted brown arrow, which reads “Further Investment.” The next dotted brown arrow goes up to connect to another dark blue box with white font that reads “Design Improvement,” which has a dotted green arrow terminating in its right side coming from the mauve “Usability/Accessibility Information” box discussed previously. Ultimately, the “Design Improvement” box leads to Usability/Accessibility Assessment with a dotted brown arrow.

**Table 1**

**Brief Description.** Table shows a summary comparison between various SP methods.

**Essential Description.** This table is used to demonstrate the advantages of DCE over other prominent SP methods. These methods are listed in the leftmost column and compared across various factors shown on the top row.

**Table 2**

**Brief Description.** Table shows the number of BPMs found on e-tailers, and the number of catalogued ones.

**Essential Description.** This table demonstrates how many BPMs each e-tailer reported to have found in the search results, along with the number of unique BPMs that were catalogued for market analysis in the dissertation study. The e-tailers are listed in the leftmost column, and the top row identifies two columns named “Total BPMs Found” and “Unique BPMs Catalogued.” The purpose of this table is to demonstrate the thoroughness and practicality of the market research for the dissertation study.

**Table 3**

**Brief Description.** Table shows the realistic attributes and levels identified for this dissertation research.

**Essential Description.** This table shows the attributes of this dissertation research in the left, their levels in the middle, and their data type in the right column. The purpose of this table is to show a brief summary of the realistic attributes and levels selected for this study.

**Table 4**

Brief Description. Table shows the attribute level balance of the feasibility study when the choice matrix was created.

Essential Description. This table serves to demonstrate the acceptable attribute level balance of the feasibility study, considering that the goal of the D-Efficient design is to maximize the information and not balance optimization. The purpose of this table is to denote the attention paid to proper experiment design in this study. The four attributes are listed in the top row, and below each attribute the levels are listed on the left and the counts for each level are listed on the right.

**Table 5**

**Brief Description.** Table shows the attribute level balance of the Phase I study when the choice matrix was created.

**Essential Description.** This table serves to demonstrate the acceptable attribute level balance of the Phase I study, considering that the goal of the D-Efficient design is to maximize the information and not balance optimization. The purpose of this table is to denote the attention paid to proper experiment design in this study. The four attributes are listed in the top row, and below each attribute the levels are listed on the left and the counts for each level are listed on the right.

**Table 6**

**Brief Description.** Table shows the parameter estimates for the utility model in Phase I.

**Essential Description.** This table demonstrates the parameters estimated for the utility model of the Phase I study. The parameter name, estimate, standard error, 95% confidence interval, z-statistic, and p-value are listed in the columns from left to right.

**Table 7**

**Brief Description.** Table shows the typical MNL model performance indicators to evaluate the utility model identified in Phase I.

**Essential Description.** This table shows that all relevant performance metrics indicate a great model performance for the utility function estimated in Phase I. The indicators listed in the left column are commonly seen in literature to evaluate MNL model performance.

**Table 8**

**Brief Description.** Table demonstrates the variance-covariance matrix of the model parameter estimates in Phase I.

**Essential Description.** This table shows that the variances and covariances of parameter estimates for Phase I appear acceptable. The parameter names are listed in the second row and repeated in the first column. Each cell shows either the variance of the same parameter, or the covariance of two different parameters. Only the lower triangle of the matrix is shown since the upper triangle above the diagonal would show the same values for covariances.

**Table 9**

**Brief Description.** Table shows the attribute level balance of the Phase II study when the choice matrix was created.

**Essential Description.** This table serves to demonstrate the acceptable attribute level balance of the Phase II study, considering that the goal of the D-Efficient design is to maximize the information and not balance optimization. The purpose of this table is to denote the attention paid to proper experiment design in this study. The four attributes are listed in the top row, and below each attribute the levels are listed on the left and the counts for each level are listed on the right.

**Table 10**

**Brief Description.** Table shows the parameter estimates for the utility model in Phase II.

**Essential Description.** This table demonstrates the parameters estimated for the utility model of the Phase II study. The parameter name, estimate, standard error, 95% confidence interval, z-statistic, and p-value are listed in the columns from left to right.

**Table 11**

**Brief Description.** Table shows the typical MNL model performance indicators to evaluate the utility model identified in Phase II.

**Essential Description.** This table shows that all relevant performance metrics indicate a great model performance for the utility function estimated in Phase II. The indicators listed in the left column are commonly seen in literature to evaluate MNL model performance. However, despite a performant model, the prediction success rate was low, which triggered an a posteriori analysis to identify the reason.

**Table 12**

**Brief Description.** Table demonstrates the variance-covariance matrix of the model parameter estimates in Phase II.

**Essential Description.** This table shows that the variances and covariances of parameter estimates for Phase II appear acceptable. The parameter names are listed in the second row and repeated in the first column. Each cell shows either the variance of the same parameter, or the covariance of two different parameters. Only the lower triangle of the matrix is shown since the upper triangle above the diagonal would show the same values for covariances.

**Table 13**

**Brief Description.** Table shows the performance comparison between the MNL model for Phase II, and the LC-MNL model as the number of classes increased.

**Essential Description.** This table demonstrates that the AIC, BIC, and Log-Likelihood continually decreased as the number of classes in the LC-MNL model increased from two to six, compared to the MNL model for Phase II. The implication is that there are six classes that can be identified in the Phase II choice data, with distinct utility models for each class.

**Table 14**

**Brief Description.** Table shows the class-specific parameter estimates for each of the six classes identified by the LC-MNL analysis.

**Essential Description.** This table shows the class-specific parameters in the leftmost column, and the estimate, standard error, 95% confidence interval, z-statistic, and p-value are presented in the following columns from left to right. The purpose of this table is to identify the specific utility models for different segments of the Phase II participants, as it appeared that they did not share the same priorities in their purchasing decisions.

**Table 15**

**Brief Description.** Table compares the parameter estimates identified in this dissertation research with the PwD population, against the findings reported by Dr. Mendonca in a similar study from 2010 with PwD.

**Essential Description.** This table highlights the similarity in the impact of Usability/Accessibility Score on the purchasing decisions of PwD and PwoD, by comparing the findings of this study with Dr. Mendonca's 2010 study evaluating similar parameters. Data from Dr. Mendonca's study are reported for a BPM and an exam table, on the left side of the table. The parameter estimates from Phase II of this study, on the impact of Usability/Accessibility Score, are presented on the right. The comparison shows that while the magnitude of Usability/Accessibility Score parameter estimates are generally higher for PwD, overall, both PwD and PwoD found devices and equipment with low Usability/Accessibility Score the least desirable, and high Usability/Accessibility Score was far more desirable, with medium Usability/Accessibility Score in between.