ABSTRACT

Stated-preference methods are increasingly used to obtain patient-preference information for use in healthcare decision making. However, there remains a paucity of literature transparently reporting on the application of good research principles through all phases of stated-preference study design.

In part one of this dissertation, we applied a novel 5-step framework to the development of a stated-preference instrument that could measure the treatment preferences of people with type 2 diabetes. The developed choice experiment contained 6 attributes (A1c decrease, blood glucose stability, low blood glucose, nausea, additional medicine, and cost).

In part two of this dissertation, we applied good research practices to the implementation of the developed discrete-choice experiment. Members of a United States online panel with type 2 diabetes completed a web-enabled, self-administered survey that elicited choices between treatment profile pairs. 552 participants (51% male) completed the survey. We found that patients with type 2 diabetes value both the benefits of their treatment, and the harms and treatment burden associated with treatment.

In part three of this dissertation, we sought to assess the impact of educational attainment on treatment preferences of patients with type 2 diabetes. 231 participants had completed high school or less, 156 participants had completed some college, and 165 participants had a college degree or more education. Participants with college or more were willing to pay more for A1C decreases than participants who had completed some college or high school or less.

In part four of this dissertation, we conducted a targeted literature review to identify tests for assessing validity and reliability of a DCE. We identified four domains for the validity of a DCE: measurement validity, preference reliability, decision processes, and choice rationality. These
domains consist of 14 components that can be identified using 24 possible tests of validity and reliability.

Using treatment preferences for type 2 diabetes as a case study, we demonstrated good research practices for the design, implementation, analysis, and evaluation of a stated-preference study. The applications of our approach will help other researchers conduct high quality stated-preference research for use in healthcare decision making processes.
COMMITTEE OF READERS

Advisor
John F.P. Bridges, PHD
Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health

Readers
Caleb Alexander, MD MS
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Jodi B. Segal, MD MPH
General Internal Medicine, Johns Hopkins School of Medicine

Albert Wu, MD MPH
Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health

Sydney E Morss Dy, MD MsC
Department of Health Policy and Management, Johns Hopkins School of Public Health

Tianjing Li, MD PhD MHS
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Craig Pollack MD MHA
General Internal Medicine, Johns Hopkins School of Medicine
ACKNOWLEDGEMENTS

There were many people involved in the successful completion of this dissertation. In the process of writing this dissertation, I gained a true appreciation for the extent to which research requires collaboration and team work.

I want to sincerely thank the Johns Hopkins Institute for Clinical and Translational Research (ICTR) Community Research Advisory Council (C-RAC) and members of the Diabetes Action Board (DAB) for their valuable contributions and engagement in the research study. I also thank the respondents who participated in the study.

I thank the entire PCORI stated-preference team for their valuable contributions to this project. This project contained many components, from literature reviews to conducting focus groups and interviews to statistical analyses. Without the PCORI stated-preference team this dissertation could not have been completed.

I want to thank the Johns Hopkins stated-preference team for providing a sounding board for ideas. Thank you for stepping up, not once, but twice, to administer surveys when I was traveling out of town. Our team helped create a supportive environment and made coming into the office every day fun. I want to particularly thank my office mate of three years, Mo Zhou, for answering all my econometric questions and for letting me distract you from your work.

I want to thank my committee members and dissertation readers for their advice and patience. Your expertise taught me about qualitative methods, survey design, regulatory science, and econometric choice models.

I want to thank my advisor John FP Bridges for bringing me onto the PCORI stated-preference methods project and for your dedicated mentorship. You have taught me everything I know about stated-preference methods. Thank you for providing numerous opportunities to learn, conduct
research, and present my work. Your passion for patient engagement has inspired my own.

I want to thank my parents, Willem Janssen and Dorien van Herpen. You have taught me the value of education and of hard work by leading through example. To my sisters, Simone and Annemieke Janssen, thank you for just sitting on the couch all day when you visited because I had to write and thank you for bursting into song with me at random.

Lastly, I want to thank my husband, Matt Zapf. We have hit quite a few milestones in these five years! You have kept me grounded with both feet in the real world throughout the entire graduate school process. Your unwavering support and confidence in me inspires me to try a little harder every day.

**Funding**

Research reported in this work was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-5946). The statements in work are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. Research reported in this work was also supported by the Johns Hopkins Center of Excellence in Regulatory Science and Innovation and the Food and Drug Administration (U01FD004977). The purpose of this funding is to enhance regulatory science through a unique FDA-Academic partnership. Ellen Janssen was also supported by LUNGevity (with support from Celgene) for work conducted during her dissertation.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ II

COMMITTEE OF READERS .............................................................................................................. IV

ACKNOWLEDGEMENTS................................................................................................................... V

LIST OF TABLES .............................................................................................................................. IX

LIST OF FIGURES ............................................................................................................................ X

CHAPTER ONE - BACKGROUND AND INTRODUCTION ............................................................... 1

DIABETES ........................................................................................................................................... 2
PATIENT PREFERENCES .................................................................................................................... 9
OVERVIEW OF THE DISSERTATION .............................................................................................. 12
REFERENCES ................................................................................................................................. 14

CHAPTER TWO - METHODS ........................................................................................................... 19

STATED-PREFERENCE METHODS ................................................................................................. 20
STATISTICAL ANALYSIS .................................................................................................................. 27
VALIDITY AND RELIABILITY OF STATED-PREFERENCE METHODS ............................................. 30
DATA COLLECTION .......................................................................................................................... 30
APPENDICES .................................................................................................................................... 34
REFERENCES .................................................................................................................................... 36

CHAPTER THREE - A FRAMEWORK FOR INSTRUMENT DEVELOPMENT OF A CHOICE
EXPERIMENT: AN APPLICATION TO TYPE 2 DIABETES ............................................................... 39

ABSTRACT ........................................................................................................................................ 40
INTRODUCTION ............................................................................................................................... 41
METHODS ....................................................................................................................................... 42
RESULTS .......................................................................................................................................... 48
DISCUSSION .................................................................................................................................... 62
CONCLUSION .................................................................................................................................... 65
REFERENCES .................................................................................................................................... 66

CHAPTER FOUR - CONDUCTING A DISCRETE-CHOICE EXPERIMENT STUDY FOLLOWING
RECOMMENDATIONS FOR GOOD RESEARCH PRACTICES: AN APPLICATION TO ELICITING
PATIENT PREFERENCES FOR DIABETES TREATMENTS ........................................................... 70

ABSTRACT ........................................................................................................................................ 71
INTRODUCTION ............................................................................................................................... 72
METHODS ....................................................................................................................................... 73
RESULTS .......................................................................................................................................... 81
DISCUSSION .................................................................................................................................... 85
CONCLUSION .................................................................................................................................... 90
APPENDICES .................................................................................................................................... 91
APPENDIX C ..................................................................................................................................... 91
APPENDIX D ..................................................................................................................................... 94
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Treatment for type 2 diabetes among people age 18 and older, United States</td>
<td>5</td>
</tr>
<tr>
<td>2-1</td>
<td>Priors from pilot test used for Bayesian D-efficient experimental design</td>
<td>32</td>
</tr>
<tr>
<td>3-1</td>
<td>Recommended stages of instrument development process of a choice experiment and application to type 2 diabetes experiment example</td>
<td>42</td>
</tr>
<tr>
<td>3-2</td>
<td>Adjustments in attributes and levels included in the BWS and DCE experiments through two rounds of revisions</td>
<td>49</td>
</tr>
<tr>
<td>3-3</td>
<td>Summary characteristics of DCE studies included in the evidence review</td>
<td>51</td>
</tr>
<tr>
<td>3-4</td>
<td>Demographic and clinical characteristics of pretest and pilot participants</td>
<td>56</td>
</tr>
<tr>
<td>3-5</td>
<td>Results of the pretest interviews (n=25)</td>
<td>58</td>
</tr>
<tr>
<td>4-1</td>
<td>Application of the general framework for stated-preference methods</td>
<td>74</td>
</tr>
<tr>
<td>4-2</td>
<td>Demographic and clinical characteristics of survey participants</td>
<td>82</td>
</tr>
<tr>
<td>4-3</td>
<td>Results of mixed logit analysis</td>
<td>86</td>
</tr>
<tr>
<td>4-4</td>
<td>Gaps in current guidance on stated-preference methods using the general framework</td>
<td>89</td>
</tr>
<tr>
<td>5-1</td>
<td>Characteristics of sample by educational attainment</td>
<td>109</td>
</tr>
<tr>
<td>5-2</td>
<td>Willingness to pay for diabetes medication</td>
<td>110</td>
</tr>
<tr>
<td>6-1</td>
<td>Articles included in targeted literature review with discussed tests of validity and reliability</td>
<td>123</td>
</tr>
<tr>
<td>6-2</td>
<td>Summary of tests for validity and reliability of a DCE</td>
<td>125</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE 1-1 Prevalence of type 2 diabetes by age and race/ethnicity, United States ....................... 7
FIGURE 1-2 Stated-preference study process .................................................................................13
FIGURE 2-1 Conceptual model: Balancing positive and negative attributes ...............................23
FIGURE 2-2 Example of a Discrete Choice Experiment choice task ..............................................24
FIGURE 3-1 Timeline of instrument development process ............................................................43
FIGURE 3-2 Evidence synthesis study selection results .................................................................50
FIGURE 3-3 Frequency and standardized relative attribute importance of diabetes medication
  attributes found in the DCE literature on preferences in type 2 diabetes .................................53
FIGURE 3-4 Example of a Discrete Choice Experiment Choice Task ..........................................59
FIGURE 3-5 Example of a Best-Worst Scaling Choice Task .......................................................60
FIGURE 3-6 Preference weights and relative attribute importance for diabetes medication attributes
  by type of preference experiment conducted ..............................................................................61
FIGURE 4-1 Example of a discrete-choice experiment choice task ..............................................76
FIGURE 4-2 Preference estimates for diabetes treatment related attributes ..................................84
FIGURE 5-1 Conceptual model of relationship between educational attainment, preferences, and
  health outcomes .......................................................................................................................105
FIGURE 5-2 Example of a discrete choice experiment choice task .............................................106
FIGURE 5-3 Budget allocation for attributes of diabetes medication by educational attainment ....111
FIGURE 6-1 Domains and tests to assess the validity and reliability of a DCE ...............................124
FIGURE 6-2 Illustration of the Scope test and Simple Scope test to test for level recoding ............134
CHAPTER ONE

- 

BACKGROUND AND INTRODUCTION
Diabetes

Diabetes is a set of metabolic diseases that results from defects in insulin action and/or secretion. It is characterized by chronic hyperglycemia and results in disturbances of carbohydrate, fat, and protein metabolism. Given the World Health Organization’s (WHO) criteria, diabetes mellitus is diagnosed by fasting plasma glucose level ≥126 mg/dL (7.0 mmol/L), 2-hour plasma glucose level ≥200 mg/dL (11.1 mmol/L) after a 75-g oral glucose load, or a glycated hemoglobin level of ≥6.5%. Impaired glucose tolerance, impaired fasting glucose, and glycated hemoglobin values between 5.7% and 6.4% are collectively associated with increased risk of diabetes development and are often known as prediabetes. Diabetes is associated with long-term micro-vascular and macro-vascular complications that result in damage and sometimes failure of various organs, especially the eyes, kidneys, nerves, heart, and blood vessels.

Diabetes classification

Traditionally diabetes mellitus has been classified into two primary types: autoimmune (type 1) and nonautoimmune (type 2). With an increased, though still limited, understanding of its etiology the classification of diabetes mellitus has advanced. Type 1 diabetes mellitus was originally classified as juvenile diabetes and then as insulin dependent diabetes. Now differences in antibody positive (type 1a) and antibody negative (type 1b) are recognized. Slower adult-onset forms of autoimmune related diabetes are now recognized separately from type 2 diabetes as latent autoimmune diabetes of adults [LADA] and may include distinctions for more subtle forms of immune involvement. I simplify the discussion and merely consider type 1 and type 2 diabetes. Since this study is interested in the patient experience of living with diabetes, particularly type 2 diabetes, and most people are still diagnosed with type 1 or type 2 diabetes, these two classifications are appropriate.
Type 1 diabetes is an autoimmune disorder in which the body’s immune system attacks the β-cells in the pancreas, the cells that produce insulin, which eventually eliminates the production of insulin. Type 1 diabetes is associated with a genetic predisposition: monozygotic twins have a 50% concordance rate and dizygotic twins have a concordance rate of 10%.\textsuperscript{8,9} Onset of type 1 diabetes symptoms occurs because of a combination of β-cell loss and dysfunction. Despite extensive research, the timeframe and nature of the autoimmune abnormalities and destruction of the β-cells is still unclear.\textsuperscript{1}

Type 2 diabetes is characterized by insulin resistance and insulin deficiency. Just like type 1 diabetes, type 2 diabetes is associated with a genetic predisposition and multiple genetic mutations have been identified.\textsuperscript{10} However, for diagnosis and treatment of type 2 diabetes, the focus is generally on environmental factors since it is often not possible to identify genetic abnormalities. In type 2 diabetes, chronic glucotoxicity and lipotoxicity have been associated with diminishing insulin secretion by decreasing the amount of insulin secretory granules in the pancreas or by decreasing the conversion of proinsulin to insulin.\textsuperscript{1} In persons with type 2 diabetes, β-cell function deteriorates with increasing hyperglycemia despite treatment.\textsuperscript{11} In autopsy studies, individuals with type 2 diabetes showed a reduced β-cell mass.\textsuperscript{12}

Type 2 diabetes is often associated with obesity and the metabolic syndrome, but 15% of whites and the majority of south Asians with type 2 diabetes are non-obese.\textsuperscript{13,14} Many individuals with obesity have insulin resistance and progress to diabetes, but some never develop overt diabetes; their β-cells continue to function, and they are able to regulate their glycemic levels by compensating for increased insulin resistance with increased insulin secretion.\textsuperscript{1} The heterogeneity of the type 2 diabetes disease process is documented in part by The Baltimore Longitudinal Study of Aging.\textsuperscript{15}
Treatment of diabetes

Diabetes treatment most often requires a combination of lifestyle adaptation and medication. The ADA and EASD recommend a patient-centered approach that incorporates patient preferences, cost and side effects of medications, effects on body weight, and risk of hypoglycemia. Lowering A1c levels below or around 7% has been shown to reduce macrovascular and microvascular complications of diabetes. Therefore, the recommended A1c goal for many nonpregnant, non-elderly adults is <7%. 16

Lifestyle interventions that include nutrition therapy and exercise are important in the management of diabetes. 17,18 Weight loss may be clinically beneficial for people with type 2 diabetes, especially when undertaken in early stages of the disease. 19 Diabetes education programs that included nutrition therapy, both at an individual and at a group level, have reported A1c decreases of 0.5–2% for type 2 diabetes. 20-22 The ADA and the American College of Sports Medicine summarize the evidence for the benefits of exercise in people with type 2 diabetes. 23 Exercise interventions have been shown to lower A1c by about 0.66% in people with type 2 diabetes, even when no significant weight loss occurred. 24 Higher levels of exercise intensity are associated with a greater improvement in A1c levels and fitness, 25 and a slowing decline in mobility among overweight patients with diabetes. 26

While diet and exercise are advocated in the management of diabetes, there is no evidence of the benefit of comprehensive lifestyle interventions alone on patient-oriented outcomes, all-cause mortality, cardiovascular, and microvascular outcomes. 27 There exist multiple types of glucose lowering medication: eight classes of oral medications, a class of subcutaneous medications, and insulin. 28-31 These treatments are associated with a range of side effects, including weight and blood pressure changes, kidney damage, gastrointestinal side effects, and cardiovascular risk. 28-31
A majority of diabetes patients use only an oral medication, and about 14% of diabetes patients use neither insulin nor an oral medication \(^{32}\) (Table 1-1).

**Table 1-1 Treatment for type 2 diabetes among people age 18 and older, United States**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of adults (millions)</th>
<th>Percentage (unadjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin only</td>
<td>2.9</td>
<td>14.0</td>
</tr>
<tr>
<td>Both insulin and oral medication</td>
<td>3.1</td>
<td>14.7</td>
</tr>
<tr>
<td>Oral medication only</td>
<td>11.9</td>
<td>56.9</td>
</tr>
<tr>
<td>Neither insulin nor oral medication</td>
<td>3.0</td>
<td>14.4</td>
</tr>
</tbody>
</table>


Treatment guidelines emphasize the need to incorporate patient preferences in achieving treatment goals. \(^{33-35}\) However, estimated rates of non-adherence for oral diabetes medications are high at 35.1%. \(^{36}\) In addition, many people with type 2 diabetes are non-adherent to their other medications such as statins (non-adherence at 41.8%). \(^{36}\) Younger patients (<65 years), patients new to therapy or on twice-daily doses, women, black or Hispanic patients, and patients with high Charlson Comorbidity index scores are more likely to be non-adherent. \(^{37,38}\) Higher out-of-pocket expenses are also associated with non-adherence \(^{38}\) and approximately 16% of participants report cost-related non-adherence. \(^{37}\)

Metformin is considered the first line treatment for type 2 diabetes (if not contraindicated). \(^{39}\) Metformin is generally considered safe and effective, is inexpensive, and may reduce risk of cardiovascular events. \(^{40}\) About 20-30% of people that take metformin experience severe gastrointestinal side effects. \(^{41}\) In most people these side effects are temporary but less than 5% of people need to stop taking metformin because continuous side effects. \(^{41}\) When metformin does not result in appropriate glycemic control, another medication should be added. Overall, each new
class of noninsulin medications added to the treatment regimens lowers A1c with approximately 0.9–1.1%. Many patients with type 2 diabetes eventually require the use of insulin.

**Diabetes as a public health problem**

Diabetes is a chronic disease that poses a significant public health burden in the US and around the globe. Diabetes affects 29.1 million people (9.3%) in the US. Diabetes incidence has increased by 50% over the past decade. In 2012, 1.7 million people 20 years or older (or 7.8 per 1000 people) in the US were newly diagnosed with diabetes. Approximately 90-95% of adults with diabetes suffer from type 2 diabetes. The number of people more than 65 years of age with diabetes is projected to increase by 4.5-fold between 2005 and 2050. In the entire US population the number of people with diagnosed and undiagnosed diabetes will increase from 23.7 million to 44.1 million from 2009-2034.

Diabetes is associated with a number of microvascular and macrovascular complications. Microvascular complications include eye disease (retinopathy), kidney disease (nephropathy), and neural damage (neuropathy). Macrovascular complications include cardiovascular disease such as myocardial infarction and cerebrovascular disease such as strokes. Acute metabolic complications associated with mortality include diabetic ketoacidosis from chronic hyperglycemia and coma from hypoglycemia. Other chronic complications of diabetes include depression, dementia, and sexual dysfunction.

Complications of diabetes result in significant morbidity and costs and are associated with a reduced quality of life compared to the general population. Diabetic nephropathy represents the major cause of end-stage renal failure in Western societies. Diabetic retinopathy is the leading cause of blindness among adults aged 20–74 years, and most people with diabetes exhibit some retinopathy after 20 years of disease. More than 50% of people with diabetes eventually
develop neuropathy, and some populations with diabetes exhibit a lifetime risk of lower extremity amputation of 15%. People with diabetes have a risk of myocardial infarction equivalent to that of non-diabetic individuals with a history of myocardial infarction. This equates to a threefold increased risk compared with the general population. Cardiovascular disease accounts for more than half of the mortality seen in the diabetic population and diabetes is the sixth leading cause of death in the US.

Diabetes disproportionately affects vulnerable populations such as minorities and the elderly (Figure 1-1). Compared to non-Hispanic whites, minority populations suffer a higher prevalence and 50-100% greater burden of illness and mortality from diabetes. While ethnic minorities spend less on diabetes healthcare than whites, this difference seems to be mainly based on differences in access to care between whites and blacks or Hispanics.

Figure 1-1 Prevalence of type 2 diabetes by age and race/ethnicity, United States

![Bar chart showing prevalence of type 2 diabetes by age and race/ethnicity, United States](chart.png)

<table>
<thead>
<tr>
<th>Age</th>
<th>Number with diabetes (millions)</th>
<th>Percentage with diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-44</td>
<td>4.3</td>
<td>4.1</td>
</tr>
<tr>
<td>45-64</td>
<td>13.4</td>
<td>16.2</td>
</tr>
<tr>
<td>65 or older</td>
<td>11.2</td>
<td>25.9</td>
</tr>
</tbody>
</table>

In addition, prevalence of diabetes increases significantly with age. Because many older people meet ADA criteria for pre-diabetes, less than 30% of the US population older than 65 years of age have normal glucose levels than 25% of adults over 65 years old in the United States have diagnosed or undiagnosed diabetes, and approximately 40% of people with a known diagnosis of diabetes are older than 65 years.

**Economic Burden of Diabetes**

Diabetes, including diagnosed and undiagnosed diabetes, pre-diabetes, gestational diabetes mellitus (GDM) and complications resulting from diabetes, accounts for a significant economic burden. Total estimated cost of diagnosed diabetes in 2012 was $245 billion, which includes $176 billion in direct medical costs and $69 billion in reduced productivity. This presents a 41% increase from previous estimates of total costs of diabetes in 2007. The largest components of direct costs are hospital inpatient care (43%), medications to treat diabetes complications (18%), diabetes medications and supplies (12%), physician office visits (9%), and nursing/residential facility stays (8%).

The average annual medical cost per case of undiagnosed diabetes is $2,864, and for pre-diabetes is $443. People with diagnosed diabetes incur average medical expenditures of about $13,700 per year, of which about $7,900-9,677 is attributed to diabetes. People with diagnosed diabetes, on average, have medical expenditures approximately 2.3 times higher than what expenditures would be in the absence of diabetes. Better A1c control and complications are associated with between $67-$158 lower cost of diabetes treatment.

Generally, direct costs of diabetes treatment are higher than indirect costs of diabetes. Overall, people with diabetes are less likely to be employed and more likely to have work limitations. Workplace productivity effects range from no impact of diabetes on lost workdays to 3.2 lost
workdays depending on the population and disease severity. Lost earning range from no lost earnings to $21,392 a person a year. Total indirect costs of diabetes include costs because of absenteeism ($5 billion), decreased productivity at work for the employed ($20.8 billion), and decreased productivity for people not in the labor force ($2.7 billion). In addition, indirect costs include cost of diabetes-related disability ($21.6 billion) and lost productivity from mortality ($18.5 billion).

Compared to other countries, the USA has high direct and indirect cost of diabetes, even after controlling for GDP per capita. Care for people with diagnosed diabetes accounts for more than 1 in 5 healthcare dollars in the US. 62.4% of the cost for diabetes care in the US is provided by government insurance (including Medicare, Medicaid, and the military), 34.4% is paid for by private insurance, and 3.2% is paid for by the uninsured.

Annual diabetes-related spending in the US is expected to increase from $113 billion to $336 billion between 2009-2034. The Medicare eligible diabetes population is expected to rise from 8.2 million in 2009 to 14.6 million in 2034 with an associated increase of Medicare spending on diabetes from $45 billion to $171 billion. In addition, costs of diabetes, as well as its adverse labor market effects, increase over time and with disease severity, indicating that early investments into prevention and disease management may be able to reduce costs.

Patient preferences

Preferences and Diabetes

Due to chronic nature of the disease and the large range of treatment options and health outcomes, treatment preferences for diabetes are particularly sensitive and complex. While treatment options for type 2 diabetes have expanded, poor adherence to effective interventions persists. Even though patient preferences are seen as important by major diabetes societies,
little is actually known about patient preferences for treatment. Certain experts acknowledge the importance of patient preferences and the American Diabetes Association and the European Association for the Study of Diabetes recommend a patient-centered approach that incorporate the patient’s individual needs, values and preferences. Despite these viewpoints, various treatment guidelines on type 2 diabetes do not address balancing the relative priorities of different treatment goals, and do not explicitly incorporate patient preferences.

Several literature reviews of patient preferences in type 2 diabetes have been conducted. Identified studies that investigated trade-offs considered a wide range of attributes, including measures of blood glucose control, hypoglycemia, weight gain, gastrointestinal side effects, heart attack risk, water retention, treatment protocol, method of administration, costs, and need for self-monitoring or laboratory testing. Preferences towards drug administration were associated with previous experience with injectable diabetes medicine. However, Von Arx et al concluded that drug administration did not appear to be a strong driver of preferences.

Willingness to pay (WTP) estimates for glucose lowering medication or for avoidance of hypoglycemic events varied widely across studies. For example, patients were willing to accept a higher risk for fewer restrictions on diet. In addition, the importance of drug administration or weight changes has not been well established. Uncertainty persists firstly because few studies are of high quality. Second, subgroup analyses were limited, potentially because of sample size restrictions. Third most studies identified were industry funded; studies funded through other venues could significantly add to the existing literature.

Despite the growing literature exploring the treatment preferences of patients with type 2 diabetes, many applications inadequately engage patients and other stakeholders in the development of these instruments and uncertainty regarding patient-preference in type 2 diabetes still exists. By facilitating patient-centered decision making in type 2 diabetes,
improvements in adherence, satisfaction, and quality of life can be expected. Understanding patients’ preferences will provide valuable information to guide clinical practice and to allow the design of studies targeting treatment adherence and improved patient-centered outcomes.

**Patient preferences and health-care decision making**

Decision making in medicine can incorporate various perspectives depending on the decision context. Patient-centered outcomes research (PCOR) aims to elicit the patient’s viewpoint to generate evidence that can inform healthcare decision-makers. As part of the trend towards patient-centered care and research, regulatory agencies have emphasized that patient preferences are an important factor in healthcare decision processes such as health technology assessments (HTAs) or benefit-risk assessments. From a regulatory perspective, patients have experiential knowledge on their illness and/or health condition and can provide insights into living with the illness, technology or treatment.84

**Patient-preferences and regulatory benefit-risk assessment**

To make regulatory decisions, the Food and Drug Administration (FDA) evaluates the safety, effectiveness, and quality of the drugs. Among other provisions, the Prescription Drug User Fee Act (PDUFA) and the Food and Drug Administration Safety and Innovation Act (FDASIA) require FDA to develop and implement a structured approach to benefit-risk decision making surrounding human drug, biological, and medical device products. The Agency emphasizes the need for transparency in decision making, and aims to be as quantitative as possible in considering available data. The evidence should support that the health benefits of using the product outweigh the potential risks and should demonstrate an absence of unreasonable health risks associated with the use of products for the intended conditions.85
Patients have unique perspectives about the value of the probable benefits and the impact of potential risks of their medical treatments. In recognition of the patient experience, FDA’s Center for Drug Evaluation and Research (CDER) committed to establishing an initiative called Patient-Focused Drug Development. This initiative aims to obtain patients’ perspective on disease conditions and available treatments using a more systematic and expansive approach. In the first three years, public meetings on 20 disease areas are being conducted with participants from FDA review divisions, patient-advocacy communities, and other interested stakeholders. These meeting with give reviewers a better understanding of challenges patients face in various therapeutic areas, including possible barriers to treatment.

The first indication that quantitative evidence on patient preferences could be considered in FDA regulatory benefit-risk assessments came from guidance by the FDA Center for Devices and Radiological Health (CDRH). In this guidance, CDRH listed “patient tolerance for risk and perspective on benefit” and heterogeneity in risk-tolerance as an additional factor that CDRH may consider in regulatory reviews. Subsequently, CDRH approved a weight-loss device based on a patient risk-tolerance study. In addition, CDRH has now established the Patient Preference Initiative to provide the information, guidance, and framework necessary to incorporate patient preferences into the full spectrum of CDRH regulatory processes.

**Overview of the dissertation**

Due to chronic nature of the diabetes and the range of available treatment options and possible health outcomes, preferences for diabetes treatment are particularly sensitive and complex. Increasingly, treatment preferences are being recognized as critical in the evaluation and planning of quality healthcare. However, while several institutions and guidance now call for the use of patient preferences in diabetes treatment, surprisingly little is known about patient preferences for diabetes treatment.
The objective of this dissertation was to apply good-research practices throughout the stated-preference study process (Figure 1-2). Through measuring the treatment preferences of patients with type 2 diabetes, we present a case study on good research practices in stated-preference methods research.

**Figure 1-2 Stated-preference study process**

*Development*
- Attribute identification
- Choice task refinement

*Implementation*
- Experimental design
- Survey administration

*Analysis*
- Aggregate preference
- Preference heterogeneity

*Evaluation*
- Survey validity
- Survey reliability

*Chapter 2* of this dissertation gave an overview of stated-preference methods. In *Chapter 3*, I demonstrated the application of a novel framework for the instrument development of a choice experiment to the development of a DCE that measures treatment preferences of patients with type 2 diabetes. In *Chapter 4*, I demonstrated good research practices to the experimental design and statistical analysis of a DCE and it will examine preference heterogeneity for type 2 diabetes treatment using a mixed logit model. In *Chapter 5*, I examined preference heterogeneity among patients with type 2 diabetes with different educational attainment. In *Chapter 6*, I provided an overview of tests that assess the validity and reliability of a discrete choice experiment. Finally, in *Chapter 7*, I summarized the findings of the dissertation and discussed the limitations and policy implications of this research.
References


35. National Collaborating Centre for Chronic Conditions and the Centre for Clinical Practice at NICE. Type 2 diabetes: The management of type 2 diabetes. 2009;87.


85. Section 513(a) of the federal food, drug, and cosmetic act. 21 U.S.C. 360c(a).


CHAPTER TWO

- METHODS
Stated-Preference Methods

Stated-preference methods are qualitative and quantitative research methods that can be used to scientifically study the preferences, priorities, and values of patients and stakeholders. Stated-preference methods have been applied in market research, transportation, environmental policy, and health and provide a robust set of tools for quantifying peoples priorities, preferences, and values.

Choice models date back to Louis Thurston’s Law of Comparative Judgment which introduced the idea measuring outcomes through pairwise comparisons. Thurston also laid the groundwork for random utility theory by introducing the concept of a random variable to account for observed variability of responses. Lancaster coined the concept that goods are composed of various characteristics and that utility is derived from these characteristics rather than the good as a whole. Daniel McFadden extended Thurston’s theory of paired comparisons to account for comparisons among multiple options. He also developed the conditional logit model, the statistical foundation for stated-preference methods, for which he received the Nobel Prize in Economics in 2000.

Stated preference methods can be used to model different scenarios and offer a flexible mechanism to evaluate products, or interventions. Revealed preferences, which are obtained from the past behavior of consumers, are often complicated by constraints or selection mechanisms and might not be generalizable to other settings. In addition, stated-preference methods can be used to value treatments that are not on the market yet.

Random utility theory

Stated preference methods are derived under the premise of a utility maximizing consumer and apply random utility theory. Random-utility theory is grounded in both psychology and
Random-utility theory assumes that an individual’s utility consists of two parts: a deterministic component that can be estimated from observed characteristics and a random component resulting from unobserved characteristics. If we assume that the random component is randomly drawn from the same distribution in each individual, we can use the deterministic component of the equation to estimate overall utility. A technical explanation follows.

Random Utility Theory states that the utility ($U$) of individual $n$ for a particular object $i$ has two components, a deterministic/explainable component ($V_{ni}$), and a random component ($\varepsilon_{ni}$) that consists of unobserved characteristics. The utility individual $n$ obtains from for scenario $i$ is:

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

where $V_{ni}$ depends on characteristics of the object, $X_{ni}$, and characteristics of the individual $z_n$:

$$V_{ni} = x_{ni}' \beta + z_{ni}' \gamma \quad (2)$$

The probability that person $n$ chooses object $i$ over object $j$ is:

$$P_{ni} = \Pr(U_{ni} > U_{nj}) = \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) = \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) \quad (3)$$

From equation 3, statistical models can then be used to estimate preferences (see statistical analysis section). This approach has been the basis for many of the seminal applications of conjoint analysis in health\textsuperscript{14-16} and can easily be adapted to include more than two choice objects.
Qualitative methods for stated preferences

Stated-preference instruments collect and analyze quantitative data. However, qualitative methods are an important part of designing and testing a stated-preference instrument. Qualitative methods, such as pretesting of survey instruments, is common to achieve identification and refinement of attributes and is recommended for stated preference instruments. A consensus report published by the International Society for Pharmacoeconomics and Outcome Research provides broad guidance on how to conduct good conjoint-analysis research in healthcare. This study will explore a combination of methods, including literature review, semi-structured interviews, and patient-engagement to develop and test the instrument.

Quantitative methods for stated preferences

Stated-preferences methods revolve around the concept that goods or services (in this case treatments and medications) are composed of different characteristics, or attributes. These attributes are taken into consideration by people and positive and negative weights are assigned. Participants make repeated choices about the goods and services. Based on these choices, inferences regarding preferences for the attributes relative to each other can be made. In addition, the tradeoffs people are willing to make between attributes can be estimated. These weights differ by individual based on demographic characteristics, values, and experiences. A person balances the positive and negative attributes and makes a decision whether or not to consume the good or service Figure 2-1.
Stated-preference methods have several advantages over other preference-elicitation techniques that involve direct engagement of patients and stakeholders. First, they can be used to derive weights for individual attributes of the product and to calculate rates at which patients are willing to trade between attributes. Second, the preferences of different sub-groups of patients and stakeholders can be identified and compared using stratified analysis. Third, the validity, reliability and generalizability of the priorities and preferences of patients and stakeholders can be evaluated statistically.

Stated-preference methods include direct assessment questions, threshold technique, conjoint analysis and discrete-choice experiments (DCE), and best-worst scaling (BWS). This study makes use of DCE. DCEs are one of the most common form of stated-preference methods, and guidelines have emerged for utilizing them. In a DCE, a respondent is repeatedly presented with two distinct profiles and is asked to select the best profile. Figure 2-2 gives an example of a
Experimental design

Experimental design of stated-preference instruments remains one of the most controversial aspects of the application of stated-preference methods.\textsuperscript{\text{26}} Experimental-design defines a systematic plan that determines the number of choice tasks and the variation in the attribute levels of the choice tasks. Efficient experimental designs maximize the precision of estimated choice-model parameters for a given number of choice questions.\textsuperscript{\text{26}} Many health studies have limited sample sizes due to resource constraints or the study of rare conditions that limit sample sizes to 100 to 300 respondents.\textsuperscript{\text{27}} In those circumstances, efficient experimental designs are critical to...
the success of the study. Three major components need to be considered in experimental design: model identification, statistical efficiency, and respondent efficiency.  

Model identification refers to the ability to obtain unbiased parameter estimates for every parameter in the model. To identify particular effects of interest, the experimental design must vary the relevant attribute levels within and across choice questions, include sufficient numbers of attribute-level combinations, and take into account if interaction effects need to get estimated.

Statistical efficiency refers to minimizing the confidence intervals around parameter estimates. Efficiency will be minimized by a full factorial design that present all possible combinations of attributes and levels. However, the number of choice tasks necessary for a full factorial design quickly increases; at 2 attributes and 2 levels the number of combinations will be four, but at 4 attributes and 3 levels this already increases to 81 combinations. Therefore, different designs need to be considered. Perfectly statistical efficient designs have three properties 1) balanced, meaning that each level appears equally often within an attribute, 2) orthogonal, meaning that each pair of levels appears equally often across all pairs of attributes within the design, and 3) minimal overlap, meaning that within a choice task pairs of levels should be the same a minimal amount of the time.

Response efficiency refers to measurement error that results from respondents’ inattention to the choice questions or other influences. There may be study-design trade-offs between maximizing statistical efficiency and maximizing response efficiency. Maximizing the overall precision of the estimates requires balancing these two sources of potential error.

Traditionally, the experimental design techniques used in health have focused on the a priori statistical efficiency of the experimental design, often trying to maximize the D-efficiency or D-optimality. Such methods focus on ensuring that the comparison of the profiles is done to
maintain orthogonality. As number of attributes and levels goes up, the number of choice tasks required for an orthogonal design tends increases as well. To give each respondent a manageable number of choice tasks, a respondent often only completes a subset, or block, of the experimental design. This blocking can introduce version bias, respondent inefficiency, and scale biases.  

Statistically efficient designs generally involve the use of uninformed priors (i.e., assuming the preference weights to be estimated are all zero). Because of this, they often compare options in which the preference would be obvious and no real tradeoff needs to occur (for example one option is very desirable while the other option is very undesirable). This lack of a trade-off within the task is a violation of the underlying theory and encourages respondents to answer with simplifying heuristics (e.g., focusing on only one attribute).

Modern experimental designs employ Bayesian techniques to maximize respondent efficiency. Respondent efficient designs devote less attention to the overall statistical efficiency and develop tasks that ensure that respondents are making real trade-offs. Such methods take into account the statistical models that will be used to estimate the choice model. These designs also require the use of assumptions, priors, on the preferences. Priors can be obtained from previous studies, or from a pilot study. This Bayesian approach requires that the efficiency of a design be evaluated over numerous different draws taken from the prior parameter distributions assumed in generating the design and therefore necessitates the need for simulation.
**Statistical analysis**

**Aggregate analysis/conditional logit model**

Statistical analyses of DCEs quantify the importance of each attribute level in an individual’s choices relative to the other levels of the same attribute. These estimates can be used to determine what tradeoffs participants are willing to make between levels. They can also be used to determine the importance of the attribute relative to the other attribute, given the levels included in the choice experiment.

The most common used methods to analyze choice data in health economics is the conditional logit model. The conditional logit model assumes that there are no consistently different responses between individuals. From random utility, once again let $n$ present the individual, and $i$ the choice scenario. Assume that $\varepsilon_n$ are independently and identically distributed and follow a Type 1 extreme value distribution, then $\varepsilon_{nj} - \varepsilon_{ni}$ follows a logistic distribution. The probability of person $n$ choosing scenario $i$ is (where $\sigma_n$ is a scale parameter that is generally normalized to 1):

$$P_{ni} = \frac{\exp(\sigma_n V_{ni})}{\exp(V_{ni}) + \exp(V_{nj})} \quad (4)$$

Then the odds of choosing scenario $i$ over scenario $j$ will be:

$$\frac{P_{ni}}{P_{nj}} = \frac{\exp(\sigma_n V_{ni})}{\exp(\sigma_n V_{nj})} = \exp(V_{ni} - V_{nj}) \quad (5)$$

This means that:

$$\ln \left( \frac{P_{ni}}{P_{nj}} \right) = (V_{ni} - V_{nj}) = (x'_{ni}\beta + z'_{ni}\gamma) - (x'_{nj}\beta + z'_{nj}\gamma) = \beta(x'_{ni} - x'_{nj}) \quad (6).$$
Accounting for preference heterogeneity

When the assumption of preference homogeneity is violated and preferences vary consistently across some respondents the conditional logit model may lead to biased results. Different techniques and models have been developed to account for preference heterogeneity.

Mixed logit model

The mixed logit model assumes that preferences are not fixed, but are random variables that are continuously distributed. The mixed logit model captures the main effects as well as the distribution of each attribute level. It is increasingly used in the analysis of DCEs in healthcare. If $f(\beta|\theta)$ is the density function of $\beta$, the in the mixed logit model, the probability of person $n$ choosing scenario $i$ is:

$$P_{ni} = \frac{\int \exp(x'_{ni}\beta)}{\exp(x'_{nj}\beta)} f(\beta|\theta) d\beta$$

which needs to be estimated using maximum simulated likelihood.

Stratification

Preference heterogeneity can be systematically examined via stratification. Stratification models assume that preferences are distributed in discrete groups. It separates study subjects into homogeneous groups based on observed characteristics (e.g., demographics) and estimates separate models or separate sets of coefficients ($\beta$’s) for each stratum. Stratification methods can be combined with any other type of model including the conditional logit and mixed logit models. In addition, many researchers are familiar with the concept of stratification.
Tests for equivalence of stratified models

When running a stratified model, the equivalence of preference parameters between the different models should be evaluated. All tests discussed below are outlined as Stata 13 (College Station, Texas) code in Appendix A.

If the stratification is done by running separate models for the different groups, the models for the groups can be compared in their entirety using a log-likelihood ratio test\cite{38} or a Wald test. Paired t-tests can be used to test equivalence between the groups for each individual parameter estimate. If the model is stratified using interaction variables, a Chow test can be used to test for equivalence of between the groups for each individual parameter estimate as well for overall preference models. These procedures test whether the stratified models are equivalent, a p valued below the alpha threshold indicates that preferences are different for the specified models.

In stated-preference models, the preference parameter is influenced by the scale parameter, which is inversely related to the error variance. The scale factor in a preference model is related to people’s choice consistency; the more consistent a person is in their choices, the higher their scale factor, and the more pronounced their preferences seem to be. This means that preference weights for different individuals might seem different only because the scale factor for these individuals is different. When evaluating the results of two stratified preference models, one therefore has to test whether observed preference differences are due to true underlying preference or due to the scale factor. This can be done using a two-step Swait-Louviere test described in Swait and Louviere (1993)\cite{39} and in Appendix A.

Limitations of stratification

Stratification requires a limiting assumption that preference heterogeneity can be accurately determined a priori by observed variables,\cite{40} an assumption that is not met empirically.\cite{41,42}
addition, separate models (or interacted models) have to be estimated for each sub-group, which generally limits the number of subgroups that can be considered simultaneously. Finally, given this inability to conduct stratification on multiple indicators simultaneously, preference differences can be misclassified.

Validity and reliability of stated-preference methods

Validity is the extent to which preference measured in a stated-choice experiment reflect the true preferences of participants. However, due to the hypothetical nature of the choices made, validating stated-preference results is difficult. While there is no agreement on what constitutes a valid patient-preference study, methods to evaluate the quality of preference data exist. These methods include evaluating the consistency of choices and the use of simplifying choice heuristics, examining test-retest reliability, assessing the predictive power of preference data to predict choices, comparing preference results obtained using different elicitation techniques, and examining the face validity of the results and how they fit within previously conducted preference studies. Currently no cut-off levels exist to determine when these test meet acceptable levels.

Data collection

The Internal Review Board of Johns Hopkins School of Public Health determined this study to be exempt from human subjects review (IRB 6001). The entire project was guided by the Diabetes Action Board (DAB), an advisory board that consisted of local Baltimore community members, people with type 2 diabetes, clinicians specializing in type 2 diabetes, regulators, stated-preference methods experts, and other stakeholders. The DAB convened five times between March 2014 and November 2016 for the purposes of this study.

This study included an instrument development stage (Paper 1), an instrument implementation
stage (Paper 2), and a data analysis stage (Paper 2-3). The instrument development stage made use of qualitative semi-structured pretesting interviews with participants with type 2 diabetes from the local Baltimore community. These interviews assessed the acceptability of the survey instrument to people with type 2 diabetes. The implementation stage was subcontracted to GfK Knowledgepanel (see Appendix B); they recruited national participants and implemented the pilot and final survey instruments.

**Recruitment and sample selection**

Local recruitment and data collection was done by investigators at Johns Hopkins and by local members of the DAB. Recruitment occurred within the community and did not make use of any clinical knowledge. Recruitment was conducted through posting of flyers and by making use of DAB members’ personal networks. People interested in participating in the pretest interview called a member of the research staff to set up an appointment. Participants had to be over 18, and speak English. All participants reported having type 2 diabetes during recruitment, but some participants reported not having diabetes during the interviews themselves. Twenty-five Baltimore community members were recruited to participate in the pretest interviews.

Recruitment for the pilot and national surveys was conducted by GfK Knowledgepanel, a panel that is nationally representative of the US population. Participants had to be over 18, have self-reported physician diagnosed diabetes, and speak English or Spanish. 50 patients with type 2 diabetes completed the pilot survey. 1103 patients with type 2 diabetes completed the main national survey. To allow for sub-group analysis, we oversampled patients with type 2 diabetes who are African American (n=254) and Latino (n=236) for the main survey. Oversampling is important as African Americans and Latinos have a high prevalence of diabetes \(^{44}\) and have lower rates of adherence to treatment. \(^{45}\) Most of this study focuses on the DCE that was completed by half the participants (pilot n=25, national n=552).
Design methodology and statistical analysis

For the pilot survey, an orthogonal design with 81 choice tasks was used. For each participant, 12 choice tasks were randomly chosen without replacement from the 81 available choice tasks.

For the national survey, NGene software was used to create a Bayesian D-efficient design. Priors for the Bayesian design were estimated from the pilot survey results (Table 2-1). Cost was assumed to be continuous and fixed. The other attribute levels were assumed to be categorical and follow a uniform distribution. The MNL model was simulated with 200 random Halton draws and the resulting experimental design had a D-error of 0.06 with the estimated required minimum sample size (S-error) of 35 participants per survey version (105 total).

Table 2-1 Priors from pilot test used for Bayesian D-efficient experimental design

<table>
<thead>
<tr>
<th></th>
<th>[90% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1c level decrease</strong></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>[0.182, 0.646]</td>
</tr>
<tr>
<td>0.5%</td>
<td>Reference</td>
</tr>
<tr>
<td>0%</td>
<td>[-0.814, -0.352]</td>
</tr>
<tr>
<td><strong>Stable blood glucose</strong></td>
<td></td>
</tr>
<tr>
<td>6 days/wk</td>
<td>[0.091, 0.574]</td>
</tr>
<tr>
<td>4 days/wk</td>
<td>Reference</td>
</tr>
<tr>
<td>2 days/wk</td>
<td>[-0.662, -0.187]</td>
</tr>
<tr>
<td><strong>Low blood glucose</strong></td>
<td></td>
</tr>
<tr>
<td>None* BWS</td>
<td>[0.0483, 0.5099]</td>
</tr>
<tr>
<td>During the day only</td>
<td>Reference</td>
</tr>
<tr>
<td>During the day and at night* (Lloyd et al)</td>
<td>[-0.6097, -.47088]</td>
</tr>
<tr>
<td><strong>Nausea</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>[0.1853, .6639]</td>
</tr>
<tr>
<td>30 min/day</td>
<td>Reference</td>
</tr>
<tr>
<td>90 min/day</td>
<td>[-0.6326, -.1727]</td>
</tr>
<tr>
<td><strong>Treatment burden</strong></td>
<td></td>
</tr>
<tr>
<td>1 pill/day</td>
<td>[0.188, 0.659]</td>
</tr>
<tr>
<td>2 pills/day</td>
<td>Reference</td>
</tr>
<tr>
<td>1 pill and 1 injection/day</td>
<td>[-0.930, -0.432]</td>
</tr>
<tr>
<td><strong>Out-of-pocket cost</strong></td>
<td>-0.0352</td>
</tr>
</tbody>
</table>
The experimental design consisted of 48 forced-choice, paired-comparison choice tasks divided into three survey versions of 16 choice tasks each. Two holdout tasks were added to the experimental design. The first holdout task tested for choice consistency by repeating an earlier choice task. The second holdout task was the same across all three survey versions and was used to test for survey differences and was used to test the predictive capabilities of the choice model. Respondents were randomly assigned to one of three survey versions. The order of the choice tasks was randomized for each respondent.

Data from the pilot survey was analyzed using the conditional logit model. Data from the national survey was analyzed using the mixed logit model. A conditional logit model was used for the stratification by education analysis.
Appendices

Appendix A – Stata code with example tests for the equivalence of stratified models

*This do file outlines several methods to conduct tests of equivalence between stratified models. It covers the Wald test, Chow test, and Likelihood ratio test. It also covers how to conduct a Swait-Louviere test to test for scale differences between groups. This is discrete choice experiment data has 6 attributes each at three levels each (A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3). All model are estimated using effects coding.

*Aggregate model
global vars "A1e A2e B1e B2e C1e C2e D1e D2e E1e E2e F1e F2e"
clogit response $vars, group(resptasktype) vce(cluster caseid)
scalar la - e(ll) /*store for likelihood ratio test*/
*Run model for group 1
clogit response $vars if group == 1, group(resptasktype) vce(cluster caseid)
scalar l1 = e(ll) /*store for likelihood ratio test*/
*Run model for group 2
clogit response $vars if group == 2, group(resptasktype) vce(cluster caseid)
scalar l2 = e(ll) /*store for likelihood ratio test*/
*Log likelihood test - tests whether the aggregate model has the same fit as the stratified models
*Likelihood ratio test : lr = -2 ln(L(m1)/L(m2)) = -2(ln(l1)-ln(l2))
ml is the more restrictive model, m2 is the less restrictive model. When we compare aggregate models with stratified models, m2 is the sum of the likelihood ratios of the stratified models, it is less restrictive because we do not force the parameters for the groups to be the same*/
*get chi-squared value for log likelihood ratio
di "chi2(2) = -2 * (la - (l1 + l2))"
*chi-squared test for log likelihood ratio given the degrees of freedom (number of attribute levels in the model)
di "Prob > chi2 = chi2tail(12, -2*la - (l1 + l2))"

*****SCALE TEST
G1egroup2 G2egroup2
testparm A1egroup2 A2egroup2 B1egroup2 B2egroup2 C1egroup2 C2egroup2 D1egroup2 D2egroup2 E1egroup2 E2egroup2 F1egroup2 F2egroup2 G1egroup2 G2egroup2
*test whether the interaction terms are significantly different from zero
estimates store model1
*Run interaction model
clogit response $vars $varsint if group == 2, group(resptasktype) vce(cluster caseid)
scalar lr = e(ll) /*store for log likelihood ratio*/
estimates store model2
qui: clogit response $vars int if group == 2, group(resptasktype) vce(cluster caseid)
scalar li = e(ll) /*store for log likelihood test*/
clogit response $vars, group(resptasktype) vce(cluster caseid)
scalar lhet = e(ll) /*store for log likelihood test*/
clogithet response $vars, group(resptasktype) vce(cluster caseid)
*****Log likelihood test
clogit response $vars if group == 2, group(resptasktype) vce(cluster caseid)
scalar la = e(ll) /*store for likelihood ratio test*/
clogit response $vars if group == 1, group(resptasktype) vce(cluster caseid)
scalar l1 = e(ll) /*store for likelihood ratio test*/
clogit response $vars, group(resptasktype) vce(cluster caseid)
scalar lhet = e(ll) /*store for likelihood ratio test*/
*Log likelihood test: if we failed to reject H1a, the two groups have the same underlying preferences and we need to move on to test H1b to see if there is a scale effect the degrees of freedom are equal
di "ch2(2) = -2*(lhet-la)
*Second test H1b: if we failed to reject H1a, H1b tests whether the scale factors (mu) are the same for the two groups If we fail to reject H1b, the scale factors are the same, there is no scale effect, and we fail to reject H1 (the preference parameters are the same for both groups). If reject H1b, one of the groups has a larger error variance */
di "Prob > chi2 = chi2tail(1, -2*(la-lhet))"
Appendix B - GfK KnowledgePanel

GfK’s KnowledgePanel® is a nationally representative online panel that includes over 4,000 patients with type 2 diabetes. It has detailed information on panel members’ age, gender, income, education, racial and location and can demonstrate. The GfK KnowledgePanel provides sampling coverage of 97% of the US adult population. Panel members are randomly recruited by random direct-dial telephone (until 2009) or by address (since 2009), and households are provided with access to the Internet and hardware if needed. GfK surveys are based on a sampling frame that includes both listed and unlisted numbers, are not limited to current Internet users or computer owners, and does not accept self-selected volunteers. GfK provided sampling weights so that US national average estimates can be derived.
References


CHAPTER THREE

A FRAMEWORK FOR INSTRUMENT DEVELOPMENT OF A CHOICE EXPERIMENT: AN APPLICATION TO TYPE 2 DIABETES

Ellen M. Janssen, Jodi B. Segal MD, John F.P. Bridges PhD

---

Abstract

**Objective:** Choice experiments are increasingly used to obtain patient-preference information for regulatory benefit-risk assessments. Despite the importance of instrument design, there remains a paucity of literature applying good research principles. We applied a novel framework for instrument development of a choice experiment to measure type 2 diabetes treatment preferences.

**Methods:** Applying the framework, we used evidence synthesis, expert consultation, stakeholder engagement, pretest interviews, and pilot testing to develop a best-worst scaling (BWS) and discrete choice experiment (DCE). We synthesized attributes from published DCEs for type 2 diabetes, consulted clinical experts, engaged a national advisory board, conducted local cognitive interviews, and pilot tested a national survey.

**Results:** From published DCEs (n=17), 10 attribute categories were extracted with cost (n=11) having the highest relative attribute importance (RAI) (range: 6-10). Clinical consultation and stakeholder engagement identified six attributes for inclusion. Cognitive pretesting with local diabetes patients (n=25) ensured comprehension of the choice experiment. Pilot testing with patients from a national sample (n=50) identified nausea as most important (RAI for DCE: 10, 95% CI: 8.5, 11.5; RAI for BWS: 10, 95% CI: 8.9, 11.1). The developed choice experiment contained 6 attributes (A1c decrease, blood glucose stability, low blood glucose, nausea, additional medicine, and cost).

**Conclusion:** The framework for instrument development of a choice experiment included five stages of development and incorporated multiple stakeholder perspectives. Further comparisons of instrument development approaches are needed to identify best practices. To facilitate comparisons, researchers need to be encouraged to publish or discuss their instrument development strategies and findings.
**Introduction**

One key aspect in health economics involves the study of tradeoffs between scarce resources and the benefits of healthcare services.\(^1\) While traditionally health economists have focused on the tradeoff between cost and effectiveness,\(^2\) more recently the tradeoff between benefits and risks of treatments have gained attention.\(^3,4\) One of the challenges in these assessments is weighing the relative importance of benefits and risks to find a balance between acceptable safety and meaningful benefit.\(^5\)

As part of the trend towards patient-centered care\(^6\) and research,\(^7\) regulatory agencies have emphasized that patient preferences are an important factor in regulatory benefit-risk decisions\(^3,8\) and have released draft guidance\(^9\) on incorporating patient preference information in benefit-risk assessment. Stated-preference methods can be used to quantify the tradeoffs people are willing to make, making them particularly suitable for benefit-risk analyses.\(^10\)

While guidelines exist for the use of or stated-preference methods for good research practices,\(^11\) experimental design,\(^12\) and data analysis,\(^13\) few studies follow good practices for instrument development.\(^14\) The use of qualitative methods to develop choice experiments\(^15,16\) is well recognized and researchers continue to develop novel methods for instrument development.\(^17,19\) However, the steps in developing stated-preference instruments have not been described comprehensively and few studies provide enough detail to replicate choice tasks.\(^14\)

This study utilized a novel framework for instrument development of a choice experiment to measure patient treatment preferences in type 2 diabetes. Type 2 diabetes presents an important case study given that it is a chronic condition with a growing literature on treatment preferences.\(^20-23\) This study examines the current findings on treatment preferences in type 2
diabetes and adds to this existing literature with a high quality preference study. In addition, this study aims to fill a gap in the stated-preference literature regarding instrument development.

**Methods**

This study applied the framework for instrument development of a choice experiment consisting of five stages: 1) Evidence synthesis, 2) Expert consultation, 3) Stakeholder engagement, 4) Pretest interviews, and 5) Pilot Testing. This approach combined the current evidence base with expert and stakeholder engagement, in addition to patient-input and quantitative preference assessment (Table 3-1). We utilized this framework to develop and test two choice experiments, a BWS and a DCE, that measured treatment preferences of patients with type 2 diabetes.

**Table 3-1 Recommended stages of instrument development process of a choice experiment and application to type 2 diabetes experiment example**

<table>
<thead>
<tr>
<th>Step</th>
<th>Theoretical framework</th>
<th>Application to type 2 diabetes experiment example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence synthesis</td>
<td>To gather existing evidence on preferences and utilize the existing literature to develop the instrument</td>
<td>Compiled articles from literature reviews on treatment preferences of adults with T2DM. • Ten attribute categories were extracted from 17 studies</td>
</tr>
<tr>
<td>Expert consultation</td>
<td>To ensure clinical accuracy and relevance of the attributes and decision framework</td>
<td>Convenience sample of clinical experts. Ensured clinical relevance of choice experiment • Experts selected 8 attributes • Experts developed a decision framework</td>
</tr>
<tr>
<td>Stakeholder engagement</td>
<td>To engage stakeholders and gather input on patient-centeredness, and salience of the instrument.</td>
<td>Advisory board including patients, community members, methods experts • The advisory board selected 6 attributes at 3 levels</td>
</tr>
<tr>
<td>Pretest interviews</td>
<td>To ensure acceptability of the instrument to patients and to update the instrument based on participants’ feedback.</td>
<td>Evaluated the choice experiment through interviews with adults with T2DM in Baltimore. • 25 interviews • The choice experiment was continuously updated based on participant feedback</td>
</tr>
<tr>
<td>Pilot testing</td>
<td>To quantitatively test the instrument and to estimate attribute priors.</td>
<td>Tested choice experiment with online, national panel of adults with T2DM. • 27 participants completed the DCE (12 tasks), 23 completed the BWS (18 tasks) • Nausea was valued highest, hypoglycemia lowest</td>
</tr>
</tbody>
</table>
Figure 3-1 details different activities performed during the instrument development process, which ranged from March 2013 until February 2016. This study was reviewed and deemed exempt from human subjects review by the Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health (IRB # 6001).

Figure 3-1 Timeline of instrument development process
Evidence Synthesis

An evidence review\textsuperscript{24-27} was conducted to gather evidence on preferences in type 2 diabetes. Evidence on previously used attributes, study populations, instrument development, and study quality were used to inform attribute selection. The evidence review also placed this study in context of other preference studies in type 2 diabetes.

All articles included in four previous literature reviews\textsuperscript{20-23} on preferences in type 2 diabetes were assessed for inclusion criteria. In addition, we conducted a search in PubMed to identify studies published after the last review had been conducted, between January 2015 and 2016. Keywords included: DCE, discrete choice experiment, and diabetes. Studies were included if they conducted a DCE or Conjoint Analysis (CA), if they assessed preferences for diabetes medication in people with type 2 diabetes, and if they were published in English. Studies were excluded if they assessed preferences for medications not used to treat diabetes, or if they assessed life-style interventions.

Quality of each study was evaluated using the PREFS quality checklist, which assesses the purpose, responses, explanation, findings, and significance of preference studies.\textsuperscript{21} We extracted treatment attributes, levels, and preference estimates for each DCE. Attributes were categorized based on underlying clinical factors. Attributes that measured clinical factors not common to diabetes medications were grouped into a “side effects” category. We calculated RAI for each attribute in each DCE to estimate the importance of an attribute in the choice process relative to the other attributes in the study. If a study directly reported preference weight or attribute importance, RAI was calculated by subtracting the preference weight of the least preferred level of the attribute from its most preferred level.
If preference weights were not reported, rankings of the attributes or willingness-to-pay values (WTP) were adapted to reflect RAI. When attributes were ranked, the least preferred attribute was assigned a RAI of 1, the second least preferred attributed a RAI of 2, etc. This approach assumed that the RAIs were equally spaced between the most important and least important attribute in a study.

If a study only reported WTP values, WTP for each attribute level was calculated. The difference between the highest and the lowest WTP within each attribute was taken to reflect the RAI. WTP presents the preference weight of an attribute level relative to the preference weight of cost. Taking the difference between the level with the largest and smallest WTP of each attribute presents the RAI of an attribute with respect to the preference weight for cost. Unfortunately, it is not possible to calculate the WTP (RAI relative to cost) for the cost attribute itself.

The RAIs within each study were standardized on a ten-point scale, where ten represented the most important attribute in the study. RAIs directly extracted from the studies, and RAIs calculated from ranking or WTP were standardized in the same way. RAI is dependent on framing of the attributes, the attributes included in the choice experiment, the levels of the attribute, and the study population. Aggregating scores for RAI across studies therefore has limitations, but the range of RAI across studies provides an idea of how much patients value an attribute.

**Expert Consultation**

Clinical experts were consulted to ensure accuracy of the survey instrument. Using a convenience sample, several local clinical experts, including two internal medicine practitioners, an RN with type 2 diabetes, and diabetes researchers, were involved in instrument development. The choice experiment was repeatedly discussed with clinical experts through in-person meetings...
and email correspondence.

Clinical experts helped categorize attributes for the evidence review. They also helped select attributes based on treatment guidelines\textsuperscript{29} and clinical relevance,\textsuperscript{30,31} and ensured the descriptions of attributes were accurate. Their expertise was used to create a clinically accurate decision framework to ensure that participant’s hypothetical choices approximated real-life choices. Results were presented to clinical experts after the pretest interviews and pilot test were completed and their input was used to revise the choice experiments.

**Stakeholder Engagement**

We formed a community advisory board,\textsuperscript{28} the Diabetes Action Board (DAB), composed of patients with type 2 diabetes, local community representatives who served on another community research advisory board, stated-preference methods experts, and regulatory experts.\textsuperscript{32} Following the Patient-Centered Outcomes Research Institute’s (PCORI) engagement rubric,\textsuperscript{33} we engaged members of the DAB throughout the study. Engagement with the DAB took place via in-person meetings at the Johns Hopkins School of Public Health.

DAB meetings educated members on the study purpose and stated-preference methods, and helped identify attributes. Stakeholders were engaged to gather input on patient-centeredness, and salience of the choice experiment. DAB members commented on and helped identify attributes. The DAB discussed, and edited the choice experiment for content and formatting. Separate meetings were held to educate the community board and to present results of pretest interviews and pilot testing to the community board and regulatory experts.
Pretest interviews

The attributes selected through expert consultation and stakeholder engagement were combined into a DCE and BWS choice experiment for pretesting. The choice experiments were evaluated through in-person semi-structured cognitive interviews among patients with type 2 diabetes in Baltimore, MD. Interviews were conducted by trained study personnel, recorded and transcribed. Participants were eligible if they were 18 or older, reported type 2 diabetes during the recruitment phone call, and spoke English.

Recruitment occurred through the distribution of flyers throughout the community in locations such as apartment buildings, markets, and churches. Local members of the DAB also reached out to personal networks. Interested individuals called a member of the research team to set up an interview. Recruitment was conducted until the study team determined that satiation had been reached. This determination was based on the interviewers’ assessment that no new information to improve the survey was gathered from additional interviews.

Participants were presented with an anonymous questionnaire to assess demographic and clinical characteristics. They were then presented with a paper version of the DCE and BWS. Participants were asked to read through the descriptions of the attributes and levels, complete the choice experiment, and verbalize responses, thoughts, and concerns about the content and formatting. Each pretest participant saw both DCE and BWS choice tasks and completed between 2 to 12 choice tasks. No particular experimental design was utilized.

Transcripts were analyzed using a five-item pretesting checklist. This checklist focused on i) comprehension of the attributes, ii) omitted attributes, iii) trade-offs, iv) elicitation format, and v) study purpose. These items were assessed to ensure that participants would be able to successfully complete the choice experiment, to limit respondent burden, and to ensure patient-
centeredness. The descriptions of the attributes were continuously updated based on feedback.

**Pilot testing**

After pretesting, the choice experiments were pilot tested to quantitatively test the instrument and to estimate attribute priors. The developed choice experiments were tested with members of an online, nationally representative panel. Participants were eligible to complete the survey if they were 18 or older, spoke English, and had self-reported type 2 diabetes. African Americans and Hispanics were oversampled. Due to limited sample size we did not analyze results separately for different subpopulations.

Detailed descriptions of the attribute and levels (in particular A1c decrease) were included in the instrument and long term effects of poor glycemic control or glucose stability were explained. It was explained that a change in one attribute did not mean that another attribute changed in response.

Participants were randomly assigned to complete either the DCE or the BWS. For the BWS, participants completed 16 choice tasks that followed a main-effects orthogonal design. For the DCE, participants completed 12 choice tasks; a main effects orthogonal design with 81 choice tasks was developed, then 12 tasks were randomly selected for each individual participant. Results were analyzed using effects coding and the conditional logit model with Stata 13 (College Station, TX). Concordance of the DCE and BWS results were compared using Spearman's rho.

**Results**

The choice experiment went through three iterations. Version 1 was developed through evidence synthesis, expert consultation, and stakeholder engagement and was pretested. Version 2 was updated after the pretest and was pilot tested. Version 3 was updated after the pilot (Table 3-2).
### Table 3-2 Adjustments in attributes and levels included in the BWS and DCE experiments through two rounds of revisions

<table>
<thead>
<tr>
<th></th>
<th>Version 1</th>
<th>Version 2 - Pilot survey</th>
<th>Version 3 - Final Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glycemic control</strong></td>
<td>A 1.1% reduction in HbA1c level</td>
<td>A1c levels decrease 1% – this is a good decrease</td>
<td>A1c level decrease 1% – this is a large decrease</td>
</tr>
<tr>
<td></td>
<td>A 0.7% reduction in HbA1c level</td>
<td>0.5% – this is a moderate decrease</td>
<td>0.5% – this is a moderate decrease</td>
</tr>
<tr>
<td></td>
<td>A 0.3% reduction in HbA1c level</td>
<td>0% – this is no decrease</td>
<td>0% – this is no decrease</td>
</tr>
<tr>
<td><strong>Glucose stability</strong></td>
<td>Stable glucose levels 6 days out of the week</td>
<td>Stable blood sugar 6 days per week</td>
<td>Stable blood glucose levels 6 days per week</td>
</tr>
<tr>
<td></td>
<td>Stable glucose levels 4 days out of the week</td>
<td>4 days per week</td>
<td>4 days per week</td>
</tr>
<tr>
<td></td>
<td>Stable glucose levels 2 days out of the week</td>
<td>2 days per week</td>
<td>2 days per week</td>
</tr>
<tr>
<td><strong>Hypoglycemic events</strong></td>
<td>No additional lows in a month</td>
<td>Low blood sugar events 0 per month</td>
<td>Low blood glucose events None</td>
</tr>
<tr>
<td></td>
<td>1 additional low in a month</td>
<td>1 per month</td>
<td>During the day only</td>
</tr>
<tr>
<td></td>
<td>2 additional lows in a month</td>
<td>2 per month</td>
<td>During the day and/or at night</td>
</tr>
<tr>
<td><strong>Nausea</strong></td>
<td>No nausea</td>
<td>Nausea None</td>
<td>Nausea None</td>
</tr>
<tr>
<td></td>
<td>Moderate nausea that lasts for 1-2 hours after you take a dose</td>
<td>2 hours per day</td>
<td>30 minutes per day</td>
</tr>
<tr>
<td></td>
<td>Moderate nausea that lasts most of the day after you take a dose</td>
<td>4 hours per day</td>
<td>90 minutes per day</td>
</tr>
<tr>
<td><strong>Dosage and administration</strong></td>
<td>1 pill daily</td>
<td>Treatment burden 1 pill per day</td>
<td>Treatment burden 1 pill per day</td>
</tr>
<tr>
<td></td>
<td>2 pills daily</td>
<td>2 pills per day</td>
<td>2 pills per day</td>
</tr>
<tr>
<td></td>
<td>1 pill and 1 injection daily</td>
<td>1 pill and 1 injection per day</td>
<td>1 pill and 1 injection per day</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>$10 in additional out-of-pocket costs a month</td>
<td>Medication costs $10 per month</td>
<td>Medication costs $10 per month</td>
</tr>
<tr>
<td></td>
<td>$30 in additional out-of-pocket costs a month</td>
<td>$30 per month</td>
<td>$30 per month</td>
</tr>
<tr>
<td></td>
<td>$50 in additional out-of-pocket costs a month</td>
<td>$50 per month</td>
<td>$50 per month</td>
</tr>
</tbody>
</table>
Evidence Synthesis

The four previous literature reviews combined reviewed 100 articles. Our PubMed search resulted in an additional 18 articles that were published in 2015 and 2016. 86 of 118 of articles were non-duplicate titles. 17 articles\textsuperscript{38-54} were included in the evidence review and had attributes extracted (Figure 3-2). Standardized RAIs could be calculated in 16 studies. Characteristics of included studies are presented in Table 3-3.

Figure 3-2 Evidence synthesis study selection results
<table>
<thead>
<tr>
<th>First Author</th>
<th>Country</th>
<th>Sample size</th>
<th>Attribute Number</th>
<th>Population included in review</th>
<th>Instrument Development</th>
<th>Funding Source</th>
<th>PREFS² quality measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aristides, M. (2004)[38]</td>
<td>France, UK, Germany, Italy, Spain</td>
<td>290</td>
<td>5</td>
<td>Type 2 diabetics receiving insulin mix injections</td>
<td>systematic review, pilot test</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Bogelund, M. (2011)[39]</td>
<td>Denmark</td>
<td>270</td>
<td>7</td>
<td>Type 2 diabetics</td>
<td>previous DCE</td>
<td>Industry</td>
<td>4 x x x x</td>
</tr>
<tr>
<td>Casciano, R. (2011)[42]</td>
<td>Africa/Middle East, Asia, Eastern Europe, Latin America</td>
<td>14,033</td>
<td>5</td>
<td>Type 1 and type 2 diabetics</td>
<td>NR²</td>
<td>Industry</td>
<td>4 x x x x</td>
</tr>
<tr>
<td>Gelhorn, H.L. (2013)[43]</td>
<td>UK</td>
<td>100</td>
<td>7</td>
<td>Type 2 diabetics using oral anti-hyperglycemic medication</td>
<td>literature review, clinical expert consultation, pilot test</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Gelhorn, H.L. (2015)[50]</td>
<td>UK</td>
<td>243</td>
<td>6</td>
<td>Type 2 diabetics using oral anti-hyperglycemic medication</td>
<td>literature review, clinical trail results, pilot study</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Guimaraes, C. (2010)[44]</td>
<td>Canada</td>
<td>274</td>
<td>6</td>
<td>Type 2 diabetics using oral anti-hyperglycemic medication and/or insulin</td>
<td>interviews, focus groups</td>
<td>Gov't, academic</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Hauber, A.B. (2005)[46]</td>
<td>Canada</td>
<td>936</td>
<td>4</td>
<td>Type 2 diabetics</td>
<td>NR²</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Hauber, A.B. (2009)[47]</td>
<td>UK, USA</td>
<td>407</td>
<td>6</td>
<td>Type 2 diabetics using oral anti-hyperglycemic medication</td>
<td>product label review, pretest interviews</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
<tr>
<td>Hauber, A.B. (2013)bc[45]</td>
<td>USA</td>
<td>1,114</td>
<td>7</td>
<td>Type 2 diabetics using oral anti-hyperglycemic medication, not using injectable T2DM treatment</td>
<td>product label review, pretest interviews</td>
<td>Industry</td>
<td>3 x x x x</td>
</tr>
</tbody>
</table>

Table 3.3: Summary characteristics of DCE studies included in the evidence review

<table>
<thead>
<tr>
<th>Score</th>
<th>P</th>
<th>R</th>
<th>E</th>
<th>F</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Sample Size</td>
<td>Sample</td>
<td>Description</td>
<td>Methodology</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>--------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Hauber, A.B. (2015) [51]</td>
<td>USA</td>
<td>923</td>
<td>7</td>
<td>Type 2 diabetes taking no or only 1 oral anti-hyperglycemic and no injectable</td>
<td>previous DCE</td>
</tr>
<tr>
<td>Hauber, A.B. (2016) [52]</td>
<td>USA</td>
<td>643</td>
<td>5</td>
<td>Type 2 diabetics</td>
<td>qualitative interviews, and pretesting</td>
</tr>
<tr>
<td>Jendle, J. (2010) [40]</td>
<td>USA</td>
<td>461</td>
<td>7</td>
<td>Type 2 diabetics prescribed anti-hyperglycemic medication</td>
<td>patient and clinical expert consultation</td>
</tr>
<tr>
<td>Lloyd, A. (2011) [41]</td>
<td>UK</td>
<td>485</td>
<td>8</td>
<td>Type 1 and type 2 diabetics using insulin</td>
<td>literature review, focus groups, pilot test</td>
</tr>
<tr>
<td>Morillas, C. (2015) [53]</td>
<td>Spain and Portugal</td>
<td>330</td>
<td>8</td>
<td>Type 2 diabetics receiving medical treatment for diabetes</td>
<td>literature review, expert consultation</td>
</tr>
<tr>
<td>Muhlbacher, A. (2015) [54]</td>
<td>Germany</td>
<td>626</td>
<td>8</td>
<td>Type 2 diabetics taking oral anti-hyperglycemic medication</td>
<td>literature review, pretest interviews, pilot testing</td>
</tr>
<tr>
<td>Polster, M. (2010) [48]</td>
<td>USA</td>
<td>1,061</td>
<td>4</td>
<td>Type 2 diabetics</td>
<td>clinical trial, pretest interviews</td>
</tr>
<tr>
<td>Porzsolt, F. (2010) [49]</td>
<td>Germany</td>
<td>827</td>
<td>7</td>
<td>Type 1 and type 2 diabetics</td>
<td>focus group</td>
</tr>
</tbody>
</table>

*WTP estimates; †Attributes ranked by importance; ‡Attribute rankings stratified by different subgroups; §Not able to extract attribute importance/rankings; ¶Two experiments in study; fNR defined as Not Reported; ‖PREFS (Joy et al 2013) is a quality assessment tool for preference studies that evaluates Purpose, Respondents, Explanation, Findings, and Significance.
Ten attribute categories were extracted including cardiovascular (CVD) risk, glucose monitoring, quality of life, side effects, treatment burden, hypoglycemia, gastrointestinal effects, weight changes, glucose measures, and cost. Treatment burden (23 times), glucose measures (19 times), and hypoglycemia (17 times) were used most frequently. On average, cost was most important (median RAI: 10), although RAI could not be inferred for the two studies that only reported WTP values. Treatment burden and glucose levels had the most variable RAIs ranging from 1-10. Extracted attributes with RAIs are presented in Figure 3-3.

Figure 3-3 Frequency and standardized relative attribute importance of diabetes medication attributes found in the DCE literature on preferences in type 2 diabetes

**Fig. 3-3** – Ten attributes from 12 Discrete Choice Experiment (DCE) studies are presented in order of standardized relative attribute importance (RAI). Number of times the attribute appeared across studies is indicated in parentheses behind the attribute name. An attribute could appear more than once in one study; it is therefore possible that an attribute appeared more than 12 times. The box around the median represents the interquartile range of RAI. Bars around the box indicate the minimum and maximum RAI of the attribute.
Expert consultation

Clinical experts defined a decision context that asked participants to consider adding a hypothetical novel diabetes treatment to their current diabetes medications: “Suppose that your doctor says that your current diabetes medicine is not working to keep your blood sugar controlled. Your doctor recommends that you add another diabetes medicine to lower your A1c.” This decision framework provided a realistic decision for patients living with type 2 diabetes and ensured that participants made decisions from the perspective of having uncontrolled A1c levels.

Clinical experts reviewed the extracted attributes from the evidence synthesis. After discussions about the clinical relevance of the attributes, they selected 8 attributes relevant to the decision framework, including glycemic control, hypoglycemia, weight changes, CVD risk, gastrointestinal side effects, blood glucose monitoring, administration frequency and mode of administration, and cost. Clinical experts helped identify 3 levels for each attribute and wrote descriptions for attributes and levels.

Stakeholder engagement

The DAB met three times to discuss instrument development: in March 2014, November 2014, and May 2015. Prior to the first DAB meeting, members from an existing community advisory board were invited to join the DAB. Between 13-20 DAB members participated in each DAB meeting. The first DAB meeting educated DAB members on stated-preference methods. Education efforts continued through September 2014 until the stakeholders had a full understanding of the project, and stated-preference methods.

At the second meeting, the eight attributes selected by clinical experts were presented to the DAB. The DAB did not consider blood glucose monitoring to be important and this was corroborated by the evidence review (Figure 3-3). They asked that gastrointestinal side effects
was changed to nausea in line with more recently developed medications and to simplify the attribute. The DAB asked to exclude weight changes from the choice experiment because they did not think it was important to patients with type 2 diabetes. After considering labeling issues, in which weight gain or weight loss could be associated with specific medications, and based on the opinion of the clinical experts, weight was excluded. The DAB discussed difficulties with interpreting risk in relation to the CVD attributes and suggested to remove CVD risk from the attributes. The clinical experts agreed because of the limited number of diabetes medications associated with CVD risk. The DAB asked the study team to add glucose stability as a patient-relevant attribute.

The third DAB meeting was used to discuss pretest results and version 2 of the choice experiment. DAB members approved the choice experiment, attribute descriptions, and language use. DAB members were given the opportunity to individually review the developed choice experiment. Based on their review, edits to formatting and descriptions of A1c levels were made. At a separate meeting, regulatory experts recommended adding questions to identify subgroups for heterogeneity analyses.

**Pretest interviews**

After expert consultation and stakeholder engagement, we selected six attributes at three levels each for pretesting. Attributes included: A1c decrease, blood glucose stability, additional low blood glucose, nausea, additional medicine, and additional out-of-pocket cost (Table 3-2). Twenty-five interviews were conducted in February and March 2015 in Baltimore. Interviews lasted between 20-60 minutes. Demographic characteristics of the study participants are presented in Table 3-4. Qualitative findings of the pretesting checklist are presented in Table 3-5.
Table 3-4 Demographic and clinical characteristics of pretest and pilot participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Pretest Subjects N=25</th>
<th>Pilot Subjects N=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender – no. (proportion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (0.36)</td>
<td>27 (0.54)</td>
</tr>
<tr>
<td>Age (years) – mean (range)</td>
<td>57 (31-89)</td>
<td>64 (32-82)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>2 (0.08)</td>
<td>22 (0.44)</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>2 (0.08)</td>
<td>13 (0.24)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>20 (0.80)</td>
<td>12 (0.26)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.04)</td>
<td>3 (0.06)</td>
</tr>
<tr>
<td>Education – no. (proportion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No High School diploma</td>
<td>2 (0.08)</td>
<td>3 (0.06)</td>
</tr>
<tr>
<td>High School degree/GED</td>
<td>7 (0.28)</td>
<td>18 (0.36)</td>
</tr>
<tr>
<td>Some college</td>
<td>11 (0.44)</td>
<td>21 (0.42)</td>
</tr>
<tr>
<td>Bachelor degree or higher</td>
<td>5 (0.20)</td>
<td>8 (0.16)</td>
</tr>
<tr>
<td>Diabetes diagnosis – no. (proportion)*</td>
<td>23 (0.92)</td>
<td>50 (1.00)</td>
</tr>
<tr>
<td>Years with diabetes – mean (range)</td>
<td>10 (0-37)</td>
<td>13 (0-62)</td>
</tr>
<tr>
<td>Self-reported A1C– no. (proportion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater than 8.0%</td>
<td>4 (0.17)</td>
<td>8 (0.16)</td>
</tr>
<tr>
<td>Between 7.1-7.9%</td>
<td>5 (0.21)</td>
<td>12 (0.24)</td>
</tr>
<tr>
<td>Smaller than 7.0%</td>
<td>8 (0.33)</td>
<td>21 (0.42)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>7 (0.29)</td>
<td>9 (0.18)</td>
</tr>
<tr>
<td>Medication – no. (proportion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A pill</td>
<td>14 (0.48)</td>
<td>32 (0.64)</td>
</tr>
<tr>
<td>An injection</td>
<td>11 (0.38)</td>
<td>14 (0.28)</td>
</tr>
<tr>
<td>None</td>
<td>4 (0.14)</td>
<td>4 (0.08)</td>
</tr>
</tbody>
</table>

*During the pretest interviews some participants reported that they did not have diabetes during a self-administered survey that was not part of eligibility requirements.

Pretest participants understood the attributes presented, but some participants had difficulty interpreting A1c decrease. Participants liked the presentation of additional low blood glucose because it contained detailed descriptions of symptoms. Names and descriptions of the attributes were adapted throughout the pretesting process to maximize participants’ understanding. Descriptions were made more comprehensive and listed short and long-term side effects. A1c levels were described in more detail. We added an explanation that none of the attributes were correlated; specifically, higher A1c decrease did not mean better glucose stability.

Participants thought that the attributes were relevant and were willing to make tradeoffs. Five participants (20%) mentioned omitted attributes such as severe hypoglycemia, amputation, blood
glucose monitoring, and diarrhea. While cost was important to participants, they were willing to pay more if the medicine offered sufficient benefits and small risks. To ensure proper trading, the levels for nausea were adapted to eliminate interaction effects with additional medicine. In addition, the levels for A1c decrease received qualitative descriptors. No attributes were added or omitted based on participants’ feedback.

Participants easily engaged with the elicitation method. Participants did not express a preference for the DCE or BWS; some participants preferred the BWS because it presented less information, others preferred the DCE because they thought it related to making choices in their everyday life. When asked how many DCE or BWS tasks participants thought they could complete in one sitting estimations varied from four to twenty. Based on participants’ feedback, we put attributes on the left and treatment profiles on the right in the choice tasks. Participants understood the purpose of the study and hoped that it would lead to the development of better medications. One participant believed that completing the survey could help people learn about their diabetes.
<table>
<thead>
<tr>
<th>Checklist Item</th>
<th>Findings</th>
<th>Quotes</th>
</tr>
</thead>
</table>
| Understood attributes and levels     | • Confusion on HbA1c levels.  
• Expressed desire for more comprehensive explanation of attributes.                                                                                                                                 | • “It comes to your glucose levels and maybe they could elaborate a little bit more so that people that are not fully understanding about the glucose levels will better understand it.” [Participant 12]  
• “[The description of the risks are] even better than the benefits, because here it give me all the symptoms of low sugar.” [Participant 5] |
| No omitted attributes                | • Omitted attributes included severe hypoglycemia, availability of medicine, diarrhea, amputation, and glucose monitoring                                                                                                                                         | • “The risk of passing out [severe hypoglycemia].” [Participant 7]  
• “Only thing is the nausea’s not so bad, it’s the diarrhea from all this.” [Participant 19]                                                                                                          |
| Made trade-offs                      | • Participants considered all attributes  
• No one attribute dominated.                                                                                                                                                                               | • “They all [attributes] had their own place of importance.” [Participant 11]  
• “It’s really a question of trading one off against the other and seeing which one you prefer over the other.” [Participant 24]                                                                 |
| Accepted the elicitation format      | • Participants were able to make choices  
• There was no clear preference for the BWS or DCE. About 25% preferred the DCE and about 25% preferred the BWS, about 50% of participants did not have a preference.                                                                 | • On the BWS: “It was kind of complicated ‘cause you don’t know what’s really going to be good for you and what’s not— some things are good for you and some things are not.” [Participant 4]  
• On the DCE: “It was a bit difficult to make the tradeoffs because there were some good things about medicine A that were just as good about medicine B and there were some things that were not so good about both.” [Participant 15] |
| Understood study purpose             | • Participants generally understood purpose of the instrument.                                                                                                                                                                                                     | • “Well, if I rate more questions, the better it might be for the people…. Because you do the survey, they might make a better medicine.” [Participant 1]  
• “Well, you explain what lows would be… So after taking… this they would have a clearer understanding of when they are.” [Participant 5] |
Pilot testing

The choice experiment for the pilot test was developed based on the pretest interviews and reviewed by clinical experts and DAB members. Figure 3-4 and Figure 3-5 present an example task of the DCE and the BWS employed in the pilot test. Fifty participants from the online panel completed the pilot survey in July 2015. Participants lived across the US (data not shown). Twenty-three participants completed the BWS and twenty-seven completed the DCE. Characteristics of pilot participants are presented in Table 3-4. Our participants were more likely to take medication than a 2011 national sample (92% v. 81% respectively) and were living with diabetes slightly longer than the 2011 sample (12 years v. 11.4 years). All 50 participants fully completed either the BWS or DCE and were included in the analysis.

Figure 3-4 Example of a Discrete Choice Experiment Choice Task

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c levels go down by</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Stable blood sugar</td>
<td>2 days per week</td>
<td>4 days per week</td>
</tr>
<tr>
<td>Additional low blood sugar events</td>
<td>2 per month</td>
<td>None</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>4 hours per day</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>1 pill per day</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$50 per month</td>
<td>$30 per month</td>
</tr>
<tr>
<td>Which medication would you choose?</td>
<td>I would choose Medicine A</td>
<td>I would choose Medicine B</td>
</tr>
</tbody>
</table>
Figure 3-5 Example of a Best-Worst Scaling Choice Task

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Level</th>
<th>Best</th>
<th>Worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c levels go down by</td>
<td>1%</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Stable blood sugar</td>
<td>2 days per week</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Additional low blood sugar events</td>
<td>2 per month</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$50 per month</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Figure 3-6 represents the preference weights for each attribute level (DCE or BWS). With the exception of hypoglycemia in the DCE, higher preference weights were associated with better clinical outcomes or lower harm and lower preference weights were associated with worse outcomes or higher harm. Nausea had the highest RAI for both the DCE (RAI: 10, 95% CI: 8.5, 11.5) and BWS (RAI: 10, 95% CI: 8.9, 11.1). Correlation between the results of the DCE and BWS was high (rho>0.9).

Based on the pilot test, the choice experiment was adapted. The duration of nausea was lowered from 4 and 2 hours to 90 and 30 minutes. In addition, hypoglycemia was adapted to measure preferences for daytime and nighttime hypoglycemic events (Table 3-2). These changes were made in consultation with the clinical experts in an attempt to increase utility balance and allow for a more efficient experimental design.59,60
Figure 3-6 Preference weights and relative attribute importance for diabetes medication attributes by type of preference experiment conducted

Fig. 6 – Preference weights and relative attribute importance (RAI) from the pilot test for each attribute (level) are indicated both for the BWS and DCE experiments. Brackets indicate the 95% confidence interval for each preference weight and RAI. Within each attribute, more preferred outcomes are indicated by higher preference weights and less preferred outcomes with lower preference weights. RAI indicates the difference between the most preferred and least preferred level of each attribute.
Discussion

Qualitative\textsuperscript{61-63} and mixed methods\textsuperscript{64} have been previously used to elicit patient preferences and are recognized to provide useful information for the development of stated-preference instruments.\textsuperscript{65} The steps applied in this framework for the development of a choice experiment have been considered in previous work,\textsuperscript{66} but have not yet been combined in this way. Coast et al. (2012)\textsuperscript{16} recognize two phases in choice experiment development: conceptual development and refinement of language. The first three steps in the framework apply to conceptual development, while the fourth pretesting step concerns refinement of language. The framework also includes pilot testing as a third phase. The insights gained from the pilot test resulted in significant changes to the choice experiment, suggesting that qualitative pretesting and quantitative pilot testing both play a crucial role in instrument development.

This is the first study to synthesize the DCE literature on patient preferences in type 2 diabetes. Previous literature reviews have compiled the preference literature for diabetes,\textsuperscript{21,22} ranked attributes within studies,\textsuperscript{20} and created standardized importance scores,\textsuperscript{23} but none aggregated importance scores across studies.

With the advent of patient-centered benefit-risk assessment and the inclusion of patient preference information, the demand for patient-preference studies may increase dramatically. As the evidence base on patient preferences grows, comprehensive evidence reviews will be needed to place the results of preference studies in context of the existing literature and help researchers build on existing research. In addition, comparisons stated-preference studies can help establish the reliability of preference studies.

In this study, we showed that is possible to identify a range of attributes from previously conducted DCEs if a significant literature exists. Attributes can also be extracted from other
literature sources such as clinical trials\textsuperscript{39} or previous qualitative studies.\textsuperscript{18} Since it is likely that more attributes will be identified in the literature than can be included in a choice experiment, expert consultation and stakeholder engagement can help narrow down attributes for inclusion.

The framework applied in this study is adaptable to a variety of needs. The approach to instrument development used in this study focused on evidence-synthesis, which is appropriate for benefit-risk assessments in the pre-market approval space. The framework can be easily adapted to place more emphasis on stakeholder engagement, making it more suitable to patient-focused drug development initiatives.\textsuperscript{67}

This study had several limitations. First, expert consultation was informal and made use of a convenience sample of clinical experts. Instrument development could be improved through a more formal expert consultation process, for example by creating an expert advisory board. The number of experts and the formality of the engagement should depend on the research question and clinical area of interest. A particularly complicated or controversial clinical area might need more clinical experts engaged in a more formal manner than a relatively well-known clinical area.

While many current studies make use of one or two thought experts in the field, this study attempted to incorporate a larger variety of perspectives. The engagement of a variety of experts is critical to ensure that different perspectives are represented.

Second, the stakeholder engagement process was long and spanned from March 2013, when education activities began, to October 2015, when the national survey launched. Engagement continues to the present day through engagement/dissemination activities with the DAB. DAB meetings were generally attended by variety of people and not every DAB members was able to attend every meeting.

Third, pretest participants were not representative of the US population and were recruited from
local DAB members’ networks. Issues identified by urban residents in Baltimore may differ from those in other populations and this may have influenced instrument development. More targeted recruitment may have resulted in a smaller but more diverse sample.

Fourth, patients might not be aware of certain treatment characteristics and might therefore not identify these characteristics during semi-structured interviews. To minimize this limitation, we could have asked pretest participants about certain attributes, such as CVD risk, directly.

Fifth, changes made to the survey after pilot testing were not further tested before they were incorporated into the final survey. It is important to test major changes to the choice experiments, but most researchers work with limited resources and time. When to finish testing the choice experiment is largely a value judgment made by the research team. Furthermore, the changes made after the pilot test were approved by clinical experts and stakeholders.

Additional research on the instrument development of choice experiments should be conducted to move towards a set of best practices. Studies should be conducted to determine how engaging different stakeholders from different backgrounds influences instrument development. The choice experiment developed in this study will be administered to a national sample of patients with type 2 diabetes. The results obtained in the pilot test will inform a Bayesian, D-efficient design in NGene. The results from the national study will be used to estimate preferences of patients with type 2 diabetes as well as identify subgroups of patients with different preference profiles based on stratification and segmentation techniques.
Conclusion

This paper applies a framework for instrument development of a choice experiment using five stages. As choice experiments and other stated-preference methods are incorporated in regulatory decision making such as benefit-risk analysis, greater attention must be placed on the development of these instruments. While our approach is relatively novel, additional comparisons of approaches are needed to identify best practices for instrument development. To facilitate these comparisons, preference researchers should be encouraged to publish or discuss their instrument development strategies and findings.
References


CHAPTER FOUR

CONDUCTING A DISCRETE-CHOICE EXPERIMENT STUDY
FOLLOWING RECOMMENDATIONS FOR GOOD RESEARCH
PRACTICES: AN APPLICATION TO ELICITING PATIENT
PREFERENCES FOR DIABETES TREATMENTS

Ellen M. Janssen, A. Brett Hauber PhD, and John F.P. Bridges PhD
Abstract

Objective: Given recent methodological advances in stated-preference methods, we sought to consolidate and illustrate good research practices in healthcare in a study measuring patient preferences for type 2 diabetes medications.

Methods: ISPOR good research practices and other recommendations were used to conduct a discrete-choice experiment. Members of a United States online panel with type 2 diabetes completed a web-enabled, self-administered survey that elicited choices between treatment profile pairs with six attributes at three possible levels each. A D-efficient experimental design blocked 48 choice tasks into three 16-task surveys. Preference estimates were obtained using mixed-logit estimation and used to calculate treatment choice probabilities.

Results: 552 participants (51% male) completed the survey. Participants were more likely to take medication (93%) than a national sample (81%). Avoiding 90 minutes of nausea was valued highest (Mean: -10.00, 95% confidence interval (CI): -10.53, -9.47). Participants wanted to avoid hypoglycemia during the day and/or night (Mean: -3.87, CI: -4.32, -3.42) or 1 pill and 1 injection/day (Mean: -7.04, CI: -7.63, -6.45). Participants preferred stable blood glucose 6 days/week (Mean: 4.63, CI: 4.15, 5.12) and a 1% decrease in A1c (Mean: 5.74, CI: 5.22, 6.25). If cost increased by one dollar, the probability that a treatment profile would be chosen decreased by 1%.

Conclusion: Despite numerous efforts to produce recommendations for stated-preference methods, ambiguity surrounding good practices remains. Partially due to lack of evidence comparing study designs, various judgments need to be made when conducting stated-preference studies. To ensure transparency these judgments should be described and justified.
**Introduction**

Patient-centered outcomes research (PCOR) aims to elicit patients’ viewpoints to inform healthcare decision making.¹ The value of patients’ experiential knowledge about living with their health condition to healthcare decision making is increasingly recognized.²⁻⁴ Several countries have initiated patient-centered approaches to regulatory decision making.⁵⁻⁷

There are several approaches to evaluating patient preferences.²⁻⁸,⁹ Qualitative information may be sufficient for relatively straightforward decisions, but stated-preference methods help to quantify preferences in support of more difficult assessments.¹⁰ Many stated-preference methods have the advantage of measuring preferences in a controlled experimental setting.¹¹ They can also be used to estimate specific tradeoffs people are willing to make in treatment choices.¹²

The increased role of patient-preference information in patient-centered decision making requires preference studies that meet standards of transparency consistent with clinical-evidence.¹³ Due to a lack of thoroughness and transparency in reporting, it is often not clear how good research practices are implemented.¹⁴ Therefore, we sought to quantify patient preferences following good research practices, and to demonstrate the application of good research practices to the development, implementation, analysis, and dissemination of this study.

We applied recommendations on stated-preference studies in a healthcare setting as part of a study focused on measuring treatment preferences of patients with type 2 diabetes mellitus.¹⁵ By choosing a disease area for which a preference research base exists,¹⁶⁻¹⁹ we could place our study in context of the existing diabetes-preference literature. While type 2 diabetes is an important case study, this research will advance methods that should have broad generalizability across diseases and stakeholder groups.
Methods

This discrete-choice experiment (DCE) survey was developed by consolidating various recommendations on conducting stated-preference studies in healthcare.\(^8,12,14,17,20-31\) We followed reporting recommendations of the ISPOR Conjoint Analysis Task Force checklist for good research practices\(^32\) because it provides consensus-based recommendations for the different phases of a stated-preference study. We also used the Task Force’s recommendations on experimental design,\(^11\) and statistical analysis.\(^33\) Table 4-1 presents how each item in this checklist was addressed and which recommendations can be referenced for additional information.

I. Research question

Defining a specific research question is not only the first step in a stated-preference study, but it guides all subsequent decisions.\(^8\) This study measured the treatment preferences of patients with type 2 diabetes mellitus. Like previous contributions,\(^34-44\) we estimated a set of preference weights for medication attributes. We then examined how each attribute level affected the probability that a treatment profile would be chosen. Based on stakeholder input,\(^28\) patients with type 2 diabetes were asked to consider the perspective of a patient that needed to start using an additional diabetes medicine. This is a common clinical occurrence in diabetes treatment\(^45,46\) and it helped standardize the choice scenario for participants with different disease histories. We chose to use a choice-based conjoint analysis, or a DCE, to allow for the examination of trade-offs across treatment attributes.\(^10\)
<table>
<thead>
<tr>
<th>Checklist item</th>
<th>How it was addressed</th>
<th>Additional guidance</th>
</tr>
</thead>
</table>
| 1. Research question | - This study measured whether benefits and harms of diabetes medication played a role in patients’ treatment decisions.  
- The perspective of a patient choosing a treatment was taken.  
- A DCE was used to measure tradeoffs. | Bridges (2003), Clark et al. (2014), Coast et al. (2012), dosReis et al. (2016)*, FDA (2016)+, Ho et al. (2016)*, Hollin et al. (2016)*, Janssen et al. (2016)*, Joy et al. (2013), MDIC (2015), Muhlbrucher et al. (2016)* |
| 2. Attribute and levels | - Attributes and levels were identified using mixed methods.  
- Qualitative methods included evidence synthesis, expert consultation, stakeholder engagement, and pretest interviews.  
| 3. Construction of tasks | - Per convention, full profile choice tasks (6 attributes per profile).  
- Per convention, paired-comparison choice tasks (2 profiles/task).  
- No cut-off was included to maximize information on tradeoffs. | Bridges (2003), dosReis et al. (2016), Janssen et al. (2016), Hollin et al. (2016), MDIC (2015), Muhlbrucher et al. (2016), Ryan et al. (2000) |
| 4. Experimental design | - This study used a blocked D-efficient Bayesian design.  
- Balance was achieved. Cost levels overlapped for 14/48 tasks.  
- 16 choice tasks was found to be appropriate during pilot testing. | Clark et al. (2014), Johnson et al. (2013), Marshall et al. (2010), Muhlbrucher et al. (2016), Ryan et al. (2003), MDIC (2015) |
| 5. Preference elicitation | - Explanations about the study purpose and how to complete the task, and an example choice task were provided. Cheap talk was included.  
- Method enables measurements of tradeoffs. We did not allow indifference.  
| 6. Instrument design | - Demographic and clinical history was collected.  
- Descriptions on attributes, levels and decision content were included.  
- Median survey completion time: 21 minutes. Survey included an explanation of purpose. Participants were paid a $10 cash equivalent for participating. | Bridges (2003), dos Reis et al. (2016), FDA (2016), Janssen et al. (2016), Joy et al. (2013), Hollin et al. (2016), MDIC (2015), Muhlbrucher et al. (2016) |
| 7. Data collection | - Participants with self-reported diagnosed type 2 diabetes were recruited from a nationally representative online panel. Minorities were oversampled.  
- Online self-administered survey to reach national population.  
| 8. Statistical analysis | - Clinical/demographic characteristics were compared to national population.  
- Non-response and simplifying heuristics were evaluated.  
| 9. Results and conclusions | - Results indicated that people with diabetes valued both treatment benefit and harm. CIs accounted for uncertainty.  
- Results agreed with published studies.  
| 10. Study presentation | - This study demonstrated good research practices.  
- An example choice task with decision scenario was included.  

*Some recommendations were published after this study commenced and were not used in this study design. We included these recommendations as reference. +FDA final guidance referenced, however the draft guidance was used throughout this study design as the final guidance had not been published. #This recommendation was developed and applied as part of the development process of this study; Recommendations only marked first time appears in table.
II. Attributes and levels

Identifying relevant preference attributes and levels is key to designing any stated-preference study. All relevant attributes and levels for this study were identified from the diabetes preference literature and supplemented with qualitative and quantitative data obtained by engaging diverse stakeholders (clinicians and diabetes researchers, stated-preference and regulatory experts, and people with type 2 diabetes). We conducted qualitative pretest interviews with people with type 2 diabetes living in the local community ($n = 25$) to refine the survey and to assess the salience of the attributes to the treatment decision. We conducted quantitative pilot testing with a national sample of people living with type 2 diabetes ($n = 27$) to obtain priors for the attribute level. We selected 6 attributes, which consisted of treatment benefits, harms, and burdens, at 3 possible levels each (the range most reported in the literature). Other recent development processes have placed more focus on community engagement. Further details regarding survey development were previously reported.

III. Construction of tasks

Construction of the choice tasks determines if the generated data can be used to answer the research question. We tested multiple choice-elicitation formats, and chose a full-profile, forced-choice choice tasks between two treatment profiles in which participants indicated which treatment they would prefer to take. This set-up allowed for the elicitation of acceptable tradeoffs people between different treatment attributes. In healthcare, full-profile, forced-choice, and paired treatment profiles are common and considered good research practice. If the number of attributes is low enough that participants can reasonably complete a full-profile task, these maximize information about trade-offs. We did not allow participants to select an opt-out to maximize information obtained about tradeoffs and to reduce biases in how participants evaluate an opt-out option. Given that forced-choice scenarios might not be realistic, we could
have included two steps in which the forced-choice was followed by an opt-out option. 47 An example choice task with decision scenario is shown Figure 4-1.

**Figure 4-1 Example of a discrete-choice experiment choice task**

Suppose that your doctor says that your current diabetes medicine is not working to keep your blood sugar controlled. Your doctor recommends that you add another diabetes medicine to lower your A1c. Which medication would you choose?

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c levels go down by</td>
<td>0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Stable blood sugar</td>
<td>2 days per week</td>
<td>6 days per week</td>
</tr>
<tr>
<td>Additional low blood sugar events</td>
<td>During the day</td>
<td>During the day and/or at night</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>90 minutes per day</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>1 pill per day</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$10 per month</td>
<td>$30 per month</td>
</tr>
<tr>
<td>Which medication would you choose?</td>
<td>I would choose Medicine A</td>
<td>I would choose Medicine B</td>
</tr>
</tbody>
</table>

**IV. Experimental design**

Experimental design affects both statistical and response efficiency. 11 NGene 48 software used row-based swapping to create a Bayesian D-efficient design. 49 D-efficient designs maximize the precision of the estimated parameters given a set number of choice tasks. 50 Priors for the Bayesian design were estimated from pilot results. Cost was assumed to be continuous and fixed 51 and the other attribute levels were assumed to be categorical and uniformly distributed. This design was sufficient to estimate main preference effects without interactions between attributes and was sufficient to answer our research question. A more advanced design approach could have been to adjust the levels of the cost attribute until a predicted choice probability of 75% for one of the treatment profiles and 25% of the other profile was achieved.
The design contained 48 choice tasks. Attribute balance was achieved but cost levels overlapped for 14 tasks. The design was blocked into three 16-task survey versions. Versions were selected to minimize average correlation between the versions and attribute levels. While, blocking reduces response burden, other desirable properties of the experimental design may not hold for individual blocks. Two additional tasks were added to each survey version. The first was a repeat-task that tested choice consistency. The second was a holdout-task that was the same across all three of the survey versions and tested for equivalence in the choices across survey versions. Participants completed 18 choice tasks, which is slightly more than average, but was deemed appropriate based on pilot results.

V. Preference elicitation

Various preference elicitation methods with different design and analytical approaches can be used to answer different research questions among different study populations. We conducted a DCE because it is the most commonly used stated-preference method in healthcare with good research practices recommendations available (Table 4-1). In addition, it allowed us to place preference results in the context of other diabetes DCEs.

To encourage participants to pay attention to the choice tasks, they were told that we were interested in measuring their preferences to help improve medical practice. We also included cheap talk and a $10 cash equivalence to motivate participants. At the start of the choice tasks participants were asked to keep in mind the choice scenario. This perspective could have been reinforced throughout, but we did not do this to reduce reading time. Three questions asked participants to evaluate the choice tasks for ease of understanding, answering, and choice consistency. To limit response burden, we did not include warm-up questions or choice tasks that tested choice task understanding. We felt participants understood how to complete the choice tasks due to extensive pretesting, but learning effects might have been present.
VI. Instrument design

As a DCE is part of a larger survey, general good research practices for survey design apply. Demographic and clinical background information was collected so characteristics of respondents could be examined and subgroup analyses could be performed Table 4-2. The survey included detailed descriptions on attributes and levels and an explanation of the decision scenario to ensure that participants understood each attribute, the choice they were asked to make, and how to complete the choice tasks (Appendix C).

Respondents were randomly assigned to one of three survey versions and the order of the choice tasks was randomized for each respondent to minimize order effect. Response burden for the overall survey was assessed during pretesting and pilot testing and it was found that participants could complete all parts of the survey. The finalized survey was translated into Spanish and reviewed by an independent, native Spanish speaker. Due to time constraints, the Spanish version of the survey was not pretested or pilot tested.

VII. Data collection

An appropriate data collection plan ensures that the research question can be answered for the population of interest. Participants were recruited via email from a national online panel that includes over 4,000 patients with type 2 diabetes. Participants were eligible if they were at least 18 years old, had self-reported physician-diagnosed type 2 diabetes, and spoke English or Spanish. African Americans and Latinos were oversampled to ensure a large enough sample size given the high prevalence of diabetes in these populations. We did not exclude participants based on medication use because we were not interested in evaluating preferences for a specific medication. Representativeness of the study sample is a desirable property, but tradeoffs with response rate, time in the field, and cost of survey administration exist.
We targeted a sample size of 500 to ensure sufficient power. Sample size calculations in stated-preference methods often rely on rules of thumb\textsuperscript{23,61,62} or on assumptions regarding preference estimates.\textsuperscript{30} Participants self-administered the survey through an online platform October 9-24, 2015. An online survey was chosen to reach a large sample of participants across the US. Paper based surveys might be more successful in reaching older populations and interviewer-led surveys may improve the quality of the data, but they take longer, are more expensive, and can be more difficult to administer across a wide geographic area. The Johns Hopkins School of Public Health IRB determined this study exempt from human subjects review (IRB 6001) and all data collection, such as de-identifying data and allowing participants to stop the survey at any time, was done in accordance with ethical standards.

VIII. Statistical analysis

Given their different characteristics, different analytical methods are appropriate to use in different settings.\textsuperscript{33} In this study, respondent characteristics for demographic and clinical characteristics were examined and compared to a national population. No formal tests of equivalence were conducted. We examined rates of non-response, task non-attendance, attribute dominance, and self-reported evaluations of the choice tasks. The two additional tasks were excluded from the main analysis and were used to test for quality of the responses in terms of choice consistency and survey version equivalence.

The choice model was estimated using the mixed logit model with panel data\textsuperscript{33} to adjust for within-subject correlation\textsuperscript{63,64} and account for unobserved preference heterogeneity.\textsuperscript{33} The model was estimated using the \textit{mixlogit} command with 1000 random Halton draws in Stata 13 (College Station, TX). No interaction terms were included. Cost was included as a continuous and fixed variable.\textsuperscript{51} The other attributes were included as effects-coded categorical variables that we assumed to be normally distributed. This assumption was based on convenience, as appropriate
assumptions for these distributions remain ambiguous.\textsuperscript{23,24,33,51} We chose to use effects coding to account for non-linearities.\textsuperscript{24} The level of each attribute that we expected to be most neutral was used as the omitted or reference attribute parameter. Effects coding was chosen over dummy coding as it allows for the calculation of standard errors for the omitted category.\textsuperscript{33}

Preference estimates from the choice model can be used to calculate the probability that a treatment profile will be chosen given its attribute levels.\textsuperscript{33} We used the \textit{mixlpred} command in Stata to calculate the change in choice probability for a treatment profile given that one attribute level changed compared to a reference treatment profile that contained the most neutral level for each attribute.\textsuperscript{31}

**IX. Results and conclusions**

Implications of the study results include a careful consideration of the uncertainty in the data and limitations of the study.\textsuperscript{32} The coefficients reported in this study represent relative-importance or preference estimates, the amount of weight people place on the attribute level. 95\% confidence intervals are reported to account for measurement uncertainty and to examine whether attribute levels were statistically different from each other. We expect the preference estimates to be most positive/least negative for the levels with highest benefit/lowest harm and most negative/least positive for the levels with lowest benefit/highest harm.\textsuperscript{21}

We examined how participants value the benefits of diabetes medication as opposed to the harms or burden/cost of diabetes medication to determine what medication aspects they place more weight on. We chose not to calculate willingness-to-pay, marginal rates of substitution,\textsuperscript{31} or maximum acceptable risk\textsuperscript{47} as this was not needed to address our research question. We compared results from this study to published DCEs in diabetes\textsuperscript{28} to evaluate external validity. We placed the study within existing good research practices (Table 4-1) and discussed limitations in the discussion.
X. Study presentation

Transparent reporting and clear explanations of the objective, methods, results, and conclusions are crucial in conducting a stated-preference study, but can be hampered by publication word limits. This study was targeted towards researchers interested in conducting stated-preference research and attempts to explain issues and terminology unique to stated-preference studies. To inform readers of the survey instrument, this paper includes a sample choice-task and a shortened version of the survey instrument in Appendix C. We briefly discuss, but do not elaborate on implications for diabetes clinical practice or regulatory decision making as this study focused on applying good research practices.

We present preference results in Table 4-3 and Figure 4-2. For ease of interpretation we shifted the coefficients of each attribute so that the coefficient of the lowest benefit/lowest harm was equal to zero and we rescaled the coefficients so that the absolute value of the most important outcome equaled ten. Furthermore, we provide results in terms of the predicted change in treatment profile choice probability. Figure 4-2 also provides a visual representation of the distributions of the preference estimates to illustrate individual preference heterogeneity.

**Results**

552 people, 51% male, completed the DCE. Sample characteristics are presented in Table 4-2. Mean age of the participants was 61.3 years old. Participants were more educated than the national population between 45-64 years of age. Our participants were more likely to take medication than a 2011 national sample (93% v. 81% respectively) and were living with diabetes longer (12.6 years v. 11.4 years).
Table 4-2 Demographic and clinical characteristics of survey participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=552</td>
</tr>
<tr>
<td>Age - years</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>61.30</td>
</tr>
<tr>
<td>Range</td>
<td>(24, 91)</td>
</tr>
<tr>
<td>Gender – N (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>279 (0.51)</td>
</tr>
<tr>
<td>Female</td>
<td>273 (0.49)</td>
</tr>
<tr>
<td>Race – N (%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>289 (0.52)</td>
</tr>
<tr>
<td>Black</td>
<td>126 (0.23)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>119 (0.22)</td>
</tr>
<tr>
<td>Other</td>
<td>18 (0.03)</td>
</tr>
<tr>
<td>Education – N (%)</td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>43 (0.08)</td>
</tr>
<tr>
<td>High school</td>
<td>188 (0.34)</td>
</tr>
<tr>
<td>Some college</td>
<td>156 (0.28)</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>165 (0.30)</td>
</tr>
<tr>
<td>Income – N (%)</td>
<td></td>
</tr>
<tr>
<td>&lt; 25,000</td>
<td>132 (0.24)</td>
</tr>
<tr>
<td>25,000 – 50,000</td>
<td>148 (0.27)</td>
</tr>
<tr>
<td>50,000 – 74,999</td>
<td>111 (0.20)</td>
</tr>
<tr>
<td>≥ 75,000</td>
<td>161 (0.29)</td>
</tr>
<tr>
<td>Time since diabetes diagnosis - years</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>12.6</td>
</tr>
<tr>
<td>Range</td>
<td>(11.4, 13.7)</td>
</tr>
<tr>
<td>Self-Reported Health – N (%)</td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>33 (.06)</td>
</tr>
<tr>
<td>Very good</td>
<td>158 (.29)</td>
</tr>
<tr>
<td>Good</td>
<td>232 (0.42)</td>
</tr>
<tr>
<td>Fair</td>
<td>106 (0.19)</td>
</tr>
<tr>
<td>Poor</td>
<td>23 (0.04)</td>
</tr>
<tr>
<td>No. Hypoglycemic events in last 6 months – N (%)</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>293 (0.53)</td>
</tr>
<tr>
<td>1 time</td>
<td>80 (0.14)</td>
</tr>
<tr>
<td>2-5 times</td>
<td>134 (0.24)</td>
</tr>
<tr>
<td>&gt; 5 times</td>
<td>45 (0.08)</td>
</tr>
<tr>
<td>Most recent A1c level – N (%)</td>
<td></td>
</tr>
<tr>
<td>&gt; 9.0%</td>
<td>21 (0.04)</td>
</tr>
<tr>
<td>≥ 8.0%, but &lt; 9.0%</td>
<td>59 (0.11)</td>
</tr>
<tr>
<td>≥ 7.0%, but &lt; 8.0%</td>
<td>153 (0.28)</td>
</tr>
<tr>
<td>&lt; 7.0%</td>
<td>228 (0.41)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>89 (0.16)</td>
</tr>
<tr>
<td>Type of diabetes medicine used – N (%)</td>
<td></td>
</tr>
<tr>
<td>No prescription medicine</td>
<td>37 (0.07)</td>
</tr>
<tr>
<td>Only pills</td>
<td>345 (0.63)</td>
</tr>
<tr>
<td>Only injections/shots</td>
<td>42 (0.08)</td>
</tr>
<tr>
<td>Pills and injections/shots</td>
<td>127 (0.23)</td>
</tr>
</tbody>
</table>
Participants took 21 minutes (median, range: 4-18,939 min) to complete the entire survey with 10 minutes (median, range: 0.9-192 min) for the DCE. 71% of participants agreed or strongly agreed that the choice tasks were easy to understand, 64% (strongly) agreed that the choice tasks were easy to answer, and 80% (strongly) agreed that they answered in a way consistent with their preference. 88 choice tasks were not completed among 9 participants (1.6%). Three participants (0.6%) always selected “Medicine A” or “Medicine B” across all choice tasks indicating task non-attendance. 81 participants (15.7%) displayed attribute dominance by always choosing the alternative with the better level of one attribute. 76% of participants answered the repeat-task consistently. Probability of choosing Treatment A for the holdout task was not statistically different across version 1 (0.59), version 2 (0.62) and version 3 (0.68) (p=0.15). Analyses include all participants who completed at least one choice task. Excluding certain participants from analysis might increase internal validity but decrease external validity. 23

Preference results

The results from the mixed logit model are presented in Figure 4-2 and Table 4-3. Given the significant standard deviation (SD) for all but two attribute levels (6 days of stable blood glucose (SD: 0.00, p=0.99) and hypoglycemia during the day and/or night (SD: 0.98, p=0.07)), the assumption of random parameters was appropriate and indicated that there was significant preference heterogeneity.

Preference estimates moved in the direction hypothesized. Avoiding 90 minutes of nausea was valued highest by participants (rescaled mean preference estimate (Mean): -10.00, 95% Confidence interval (CI): -10.53, -9.47). Participants wanted to avoid hypoglycemia during the day and/or night (Mean: -3.87, CI: -4.32, -3.42) or 1 pill and 1 injection/day (Mean: -7.04, CI: -7.63, -6.45). Participants showed a preference for stable blood glucose 6 days/week (Mean: 4.63, CI: 4.15, 5.12). Participants did not show a difference in preference for a 0.5% decrease in A1c.
(5.61, CI: 4.92, 6.29) and a 1% decrease (Mean: 5.74, CI: 5.22, 6.25). Avoiding hypoglycemia during the day only or avoiding 2 pills only was not valued significantly.

**Figure 4-2 Preference estimates for diabetes treatment related attributes**

Preference estimates for different attribute levels (n=552). Relative-importance estimates for attributes were rescaled relative to 10, corresponding to the absolute value of the most important outcome—90 minutes of nausea. The bar chart indicates the mean preference estimates. The short, bracketed, black vertical bars around the mean preference estimate indicate the 95% CI for the estimates (using SE). The box and whiskers plots indicate the 95% CI of the distribution of the preference estimates (using SD) around the mean. It indicates the spread in preference estimates (the level of preference heterogeneity) in the population. Within each attribute the vertical distance between mean estimates indicates the relative change in preference estimate for one level of the attribute compared to an adjacent level of the same attribute.
The probability that a treatment profile was chosen from a profile pair was 33% higher if the treatment decreased A1c by 1% compared to a treatment that did not decrease A1c. Treatment choice probability increased by 28% if the treatment offered stable blood glucose 6 days/week as opposed to 2 days/week. If a treatment profile caused 90 minutes of nausea a day instead of no nausea, the profile was 54% less likely to be chosen. A one-dollar increase in out-of-pocket cost was associated with a 1% decrease in the probability the treatment was chosen (Table 4-3).

Discussion

Our results are consistent with the idea that patients have strong preferences for immediate consequences of medication and discount future consequences of uncontrolled conditions. To accept 90 minutes of nausea a day, A1c levels had to decrease by 1% and blood glucose needed to be stable 4 days a week. Adherence for medications with significant side effects can be expected to be low, even if they are clinically effective.

We conducted preliminary conditional logit analyses to examine preference heterogeneity among people with different clinical backgrounds. No significant differences in preference estimates were shown between persons that had or had not experienced hypoglycemic events in the past 6 months, except for taking 1 pill and/or 1 injection a day. Participants that were taking an injection as part of their treatment regimen had less pronounced differences for treatment burden than participants that only took oral medication (Appendix C).

There is a rich and growing literature exploring the treatment preferences of patients with type 2 diabetes using stated-preference methods. However, there is a paucity of studies that have been able to transparently describe the full research process given common publication word limits. In a systematic review on treatment preferences for diabetes only 5 out of 12 studies met all 10 Taskforce checklist items.
### Table 4-3 Results of mixed logit analysis

<table>
<thead>
<tr>
<th></th>
<th>Preference estimates</th>
<th>Change in choice probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1c level decrease</strong></td>
<td>Coefficient</td>
<td>95% CI</td>
</tr>
<tr>
<td>1% Mean</td>
<td>5.74</td>
<td>(5.22, 6.25)</td>
</tr>
<tr>
<td>SD</td>
<td>2.88</td>
<td>(2.30, 3.46)</td>
</tr>
<tr>
<td>0.50% Mean</td>
<td>5.61</td>
<td>(4.92, 6.29)</td>
</tr>
<tr>
<td>SD</td>
<td>5.45</td>
<td>(4.73, 6.17)</td>
</tr>
<tr>
<td>0% Mean</td>
<td>0.00</td>
<td>(-0.51, 0.51)</td>
</tr>
<tr>
<td>SD</td>
<td>2.57</td>
<td>(1.91, 3.22)</td>
</tr>
<tr>
<td><strong>Stable blood glucose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 days/wk Mean</td>
<td>4.63</td>
<td>(4.15, 5.12)</td>
</tr>
<tr>
<td>SD</td>
<td>0.00</td>
<td>(-1.62, 1.63)</td>
</tr>
<tr>
<td>4 days/wk Mean</td>
<td>3.52</td>
<td>(2.94, 4.10)</td>
</tr>
<tr>
<td>SD</td>
<td>2.47</td>
<td>(0.48, 4.46)</td>
</tr>
<tr>
<td>2 days/wk Mean</td>
<td>0.00</td>
<td>(-0.46, 0.46)</td>
</tr>
<tr>
<td>SD</td>
<td>2.47</td>
<td>(1.91, 3.02)</td>
</tr>
<tr>
<td><strong>Hypoglycemia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None Mean</td>
<td>0.00</td>
<td>(-0.47, 0.47)</td>
</tr>
<tr>
<td>SD</td>
<td>1.27</td>
<td>(0.42, 2.13)</td>
</tr>
<tr>
<td>During the day Mean</td>
<td>-0.48</td>
<td>(-1.07, 0.11)</td>
</tr>
<tr>
<td>SD</td>
<td>2.26</td>
<td>(0.28, 4.23)</td>
</tr>
<tr>
<td>During the day and/or at night Mean</td>
<td>-3.87</td>
<td>(-4.32, -3.42)</td>
</tr>
<tr>
<td>SD</td>
<td>0.98</td>
<td>(-0.08, 2.04)</td>
</tr>
<tr>
<td><strong>Nausea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None Mean</td>
<td>0.00</td>
<td>(-0.58, 0.58)</td>
</tr>
<tr>
<td>SD</td>
<td>3.05</td>
<td>(2.44, 3.66)</td>
</tr>
<tr>
<td>30 minutes/day Mean</td>
<td>-2.78</td>
<td>(-3.34, -2.23)</td>
</tr>
<tr>
<td>SD</td>
<td>5.88</td>
<td>(5.12, 6.65)</td>
</tr>
<tr>
<td>90 minutes/day Mean</td>
<td>-10.00</td>
<td>(-10.53, -9.47)</td>
</tr>
<tr>
<td>SD</td>
<td>2.83</td>
<td>(2.21, 3.45)</td>
</tr>
<tr>
<td><strong>Treatment burden</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 pill/day Mean</td>
<td>0.00</td>
<td>(-0.51, 0.51)</td>
</tr>
<tr>
<td>SD</td>
<td>2.39</td>
<td>(1.75, 3.03)</td>
</tr>
<tr>
<td>2 pills/day Mean</td>
<td>0.21</td>
<td>(-0.54, 0.97)</td>
</tr>
<tr>
<td>SD</td>
<td>6.68</td>
<td>(5.28, 8.08)</td>
</tr>
<tr>
<td>1 pill and 1 injection/day Mean</td>
<td>-7.04</td>
<td>(-7.63, -6.45)</td>
</tr>
<tr>
<td>SD</td>
<td>4.29</td>
<td>(3.68, 4.90)</td>
</tr>
<tr>
<td><strong>Out-of-pocket cost</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1 increase Mean</td>
<td>-0.14</td>
<td>(-0.17, -0.12)</td>
</tr>
</tbody>
</table>

Relative-importance estimates for attributes were rescaled relative to 10, corresponding to the absolute value of the most important outcome—90 minutes of nausea. Mean refers to the estimated mean preference estimate, SD refers to the standard deviation of the distribution around the mean preference estimate and is a measure of preference heterogeneity, the confidence intervals (CI) around the Mean and SD give a measure of the precision of these estimates.
Despite differences in reporting transparency, the results of 17 diabetes DCEs are similar to the results from this study.\textsuperscript{28} It remains important to transparently discuss each step of the research process to ensure reproducibility. While results across studies might be similar, if methods are not clearly described, implications for clinical practice or for benefit-risk assessment might be limited. To remedy limited publication space, supplemental online appendices might provide an avenue for increasing reporting transparency.

While this study attempted to follow good research practices, limitations exist. First, the Spanish version of the survey was not pretested although a native Spanish speaker checked the translation and initial responses to the Spanish survey were monitored. Generally survey translations should be pretested and pilot tested with native speakers. 56 out of 552 participants (10\%) completed the Spanish version of the survey. Those who took the Spanish survey showed some differences in preference estimates (Appendix D), which could not be attributed to scale.\textsuperscript{72}

Second, this study might suffer from common issues in stated-preference methods such as ordering effect,\textsuperscript{73,74} hypothetical bias,\textsuperscript{53,75,76} and framing effect.\textsuperscript{21,77} Randomizing the order of choice tasks for each participant minimized ordering effects but we could also have randomized the order of the attributes. Hypothetical bias was limited due to the inclusion of cheap talk and the use of a clinically relevant decision scenario. However, the lack of an opt-out option might have decreased the realistic nature of the choice tasks. While framing effects were limited because of extensive pretesting, by framing hypoglycemia as low blood glucose\textsuperscript{78} it might not have been perceived as harmful. A1c decrease and stable blood glucose might not have been understood clearly causing it to have an artificially low preference estimate.

We attribute several positive outcomes in this study to the application of good research practices such as the low rates of attribute dominance, minimal participant burden, and the high level of choice stability and version equivalence. We recommend that researchers closely follow good
research practices and carefully consider limitations of stated-preference methods. While existing recommendations do not always detail which approach to take, by making deliberate decisions during study design choices can be justified.

The Taskforce checklist

The Taskforce checklist provides a useful tool for researchers to carefully consider each aspect of a stated-preference study. However, the checklist has limitations for use as a reporting template. First, some of the items do not have a clear place in study reporting. The statistical analysis item and results and conclusions item span the methods, results, and discussion section of a traditional manuscript, which makes it difficult to clearly report on them. The study presentation item cannot be summarized in one section but needs to be considered throughout the manuscript. Second, the requirement for “clear, concise, and complete” reporting is a paradox. We attempted to be comprehensive by including preference estimate distributions in Figure 4-2, but this might make it less clear to interpret. To be concise we did not conduct elaborate stratification analyses, which might reduce completeness of the study reporting. Third, the checklist is lengthy and certain aspects overlap. For example, questions around response burden appear both in experimental design and data collection. While these aspects should be considered at different points during the study, this can make following the checklist as a reporting template difficult. Other more concise reporting measures such as the PREFS checklist could be explored.

Implications for good practices

Despite numerous efforts to produce recommendations for the application of stated-preference methods, gaps on good research practices remain. Some of these gaps are presented in Table 4-4. Various methods, ranging from direct-assessment, threshold techniques, best-worst scaling, and DCE, require different approaches that are often not well differentiated. For example, it remains unclear which method is most appropriate for which research question or study population.
### Table 4-4 Gaps in current guidance on stated-preference methods using the general framework

<table>
<thead>
<tr>
<th>Checklist item</th>
<th>Current gaps:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Research question</td>
<td>• What research questions can be answered by a stated-preference study? In what circumstances are other preference elicitation methods more appropriate?</td>
</tr>
<tr>
<td></td>
<td>• What steps can be taken to ensure that the perspectives of multiple stakeholders are adequately represented in each phase of the study process?</td>
</tr>
<tr>
<td>2. Attribute and levels</td>
<td>• How can the negative effects of omitting attributes and positive effects of decreasing response burden be balanced?</td>
</tr>
<tr>
<td></td>
<td>• What is the effect of including levels outside of the current clinical range in terms of respondents’ acceptance and in terms of extrapolating results?</td>
</tr>
<tr>
<td>3. Construction of tasks</td>
<td>• What is the effect of showing partial, instead of full profiles? How can results from partial profiles be appropriately combined?</td>
</tr>
<tr>
<td></td>
<td>• Are there specific types of choice tasks that should be included in the instrument (e.g. a task with a dominated profile to test for rationality)?</td>
</tr>
<tr>
<td>4. Experimental design</td>
<td>• What are the effects of using an experimental design developed under different modeling assumptions than used for statistical analysis?</td>
</tr>
<tr>
<td></td>
<td>• How can the tradeoffs between efficiency, orthogonality, level balance, and level overlap be balanced?</td>
</tr>
<tr>
<td>5. Preference elicitation</td>
<td>• What type of stated-preference method is most appropriate given the research question, study perspective, and proposed participants?</td>
</tr>
<tr>
<td></td>
<td>• What types of evaluation questions on the preference-elicitation tasks should be asked? How do these evaluation questions affect confidence in preference results?</td>
</tr>
<tr>
<td>6. Instrument design</td>
<td>• What specific steps, such as pretest interviews or pilot testing, should be conducted for high quality instrument design?</td>
</tr>
<tr>
<td></td>
<td>• What specific steps can be taken to lower response burden, hypothetical bias, framing effects, and ordering effects?</td>
</tr>
<tr>
<td>7. Data collection</td>
<td>• What are sample size requirements for more complicated statistical models (such as mixed logit or latent class)?</td>
</tr>
<tr>
<td></td>
<td>• How representative should the study sample be? How can representativeness be assessed (by demographic or disease history profiles, or preference profiles?)</td>
</tr>
<tr>
<td>8. Statistical analysis</td>
<td>• Which methods to analyze preference heterogeneity are most appropriate in which circumstance? What are limitations of each method?</td>
</tr>
<tr>
<td></td>
<td>• What tests can be done to assess validity and reliability of stated-preference study? How many tests should a study meet for preference results be considered valid?</td>
</tr>
<tr>
<td>9. Results and conclusions</td>
<td>• What type of preference results (e.g. preference estimates, predicted probabilities, odds ratios) and uncertainty measures (e.g. p-values, standard errors) should be reported?</td>
</tr>
<tr>
<td></td>
<td>• What are preferred terms to describe preference research and results? What are appropriate ways to test for and report preference heterogeneity?</td>
</tr>
<tr>
<td>10. Study presentation</td>
<td>• What steps should be taken to ensure uniform reporting standards of stated-preference studies? How can transparency in stated-preference research be encouraged?</td>
</tr>
<tr>
<td></td>
<td>• What steps can be taken to increase interpretability of preference results for stakeholders that do not specialize in stated-preference methods?</td>
</tr>
</tbody>
</table>
With the increased use of complex experimental designs such as D-efficient or Bayesian designs, updated recommendations on experimental design would be valuable. Johnson et al. (2013) provide a comprehensive overview of the importance and the theoretical properties of experimental design. However, they give limited advice around practical considerations in trading off experimental design properties. Step-by-step recommendations on experimental design decisions would need to be supported by research such as the simulation approaches that have been used in transportation research.79

Transparently reporting on study design, data analysis, and potential limitations will provide stakeholders with information to replicate and evaluate the study. Additional evidence on the tests that can be used to evaluate stated-preference data quality, such as rationality tests, or scope tests, needs to be collected to formally evaluate preference study quality.2 Eventually, standardization of tests to assess validity and reliability should be considered.

**Conclusion**

This study is explicit in the decisions that were made while conducting a stated-preference study and can be used as a case study on applying and reporting on good research practices. Until more research is done comparing different research practices, recommendations on conducting stated-preference studies will remain opinion based. As various judgments need to be made when conducting stated-preference studies, these decisions should be carefully considered and transparently reported.
Appendices

Appendix C – selections of the diabetes treatment preferences survey

SURVEY ELIGIBILITY QUESTION

1. Have you been diagnosed by a doctor or other qualified medical professional with Type 2 diabetes?
   o Yes
   o No
   o Not sure

Understanding your medication preferences

The purpose of this section is to assess your preferences for diabetes medicines. Specifically, we ask that you consider the following situation.

Suppose that your doctor says that your current diabetes medicine is not working to keep your blood sugar controlled. Your doctor recommends that you add another diabetes medicine to change your A1c.

You will be shown a number of potential medicines that vary across several characteristics. These include the effect on your A1c levels, blood sugar stability, low blood sugar events, nausea, how it is taken, and its cost.

The medicines will be exactly the same, except for the characteristics we show you. For example, you will not need to change how often you check your blood sugar levels (with for example a finger stick).

These medicines do not actually exist, but answer as if they are real. First, let's have a look at the characteristics of the medicine.

A1c levels
Doctors prescribe diabetes medicines to help lower your A1c, or average blood sugar level during the past three months. Keeping your A1c at the recommended level may decrease your risk for serious health problems such as heart attack, blindness, amputation, and kidney failure. When taking the new medicine your A1c level might go down by:

   • 1% – this is a large decrease
   • 0.5% – this is a moderate decrease
   • 0% – this is no decrease

2. How important is lowering your A1c?
   o Not important at all
   o Slightly important
   o Somewhat important
   o Important
   o Very important

Stable blood glucose levels
The new medicine might help keep your blood glucose levels stable on a daily basis. Your blood glucose levels are stable for the day if they stay in a range of 70-180 mg/dl. When taking this new medicine your blood glucose levels might be stable for:

   • 6 days per week
   • 4 days per week
   • 2 days per week

3. How important is having stable blood glucose levels?
   o Not important at all
   o Slightly important
   o Somewhat important
   o Important
   o Very important

Low blood glucose
You might experience low blood glucose, also known as hypoglycemia. This may make you feel shaky/drowsy and have blurred vision or difficulty walking/talking. You might pass out (if you don’t eat or drink). Low blood glucose can also happen at night while you sleep. Then you won’t know about it and you might be more likely to pass out. You might experience low blood glucose:

- None
- During the day only
- During the day and/or at night

4. How important is avoiding low blood glucose events?
   - Not important at all
   - Slightly important
   - Somewhat important
   - Important
   - Very important

**Nausea**
The new medicine may cause moderate nausea. This means you feel sick to your stomach and like you need to vomit. When taking this new medicine you might experience nausea for a total of:

- None
- 30 minutes per day
- 90 minutes per day

5. How important is avoiding nausea?
   - Not important at all
   - Slightly important
   - Somewhat important
   - Important
   - Very important

**Treatment burden**
You will have to take the new medicine daily. You need to take this medicine in addition to the medicines you already take. We will consider three different ways of taking the medicine. You might have to take an additional:

- 1 pill per day
- 2 pills per day
- 1 pill and 1 injection per day

6. How important is minimizing treatment burden?
   - Not important at all
   - Slightly important
   - Somewhat important
   - Important
   - Very important

**Medication costs**
The medicine will require out-of-pocket costs in addition to what you already pay for other medicines. The money you spend on this medicine cannot be spent on other things. Your additional costs might be:

- $10 per month
- $30 per month
- $50 per month

1. How important is cost to you?
   - Not important at all
   - Slightly important
   - Somewhat important
   - Important
   - Very important
Discrete Choice Experiment

Consider the following two diabetes medicines. Which medicine would you prefer?

Select one choice – either Medicine A or Medicine B at the bottom of the list, below.

Example:

Tom took the survey and was shown the following two medicines. He prefers Medicine B over Medicine A.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c levels go down by</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Stable blood glucose</td>
<td>2 days per week</td>
<td>4 days per week</td>
</tr>
<tr>
<td>Low blood glucose</td>
<td>During the day</td>
<td>None</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>90 minutes per day</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>1 pill per day</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$50 per month</td>
<td>$10 per month</td>
</tr>
<tr>
<td>Which medication would you choose?</td>
<td>I would choose Medicine A</td>
<td>I would choose Medicine B</td>
</tr>
<tr>
<td>(pick one)</td>
<td>☐</td>
<td>☑</td>
</tr>
</tbody>
</table>

Task 8 – 25 Which medicine would you prefer?

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c levels go down by</td>
<td>0.5%</td>
<td>0%</td>
</tr>
<tr>
<td>Stable blood glucose</td>
<td>4 days per week</td>
<td>6 days per week</td>
</tr>
<tr>
<td>Low blood glucose</td>
<td>During the day and/or at night</td>
<td>During the day</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>30 minutes per day</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>1 pill and 1 injection per day</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$50 per month</td>
<td>$30 per month</td>
</tr>
<tr>
<td>Which medication would you choose?</td>
<td>I would choose Medicine A</td>
<td>I would choose Medicine B</td>
</tr>
<tr>
<td>(pick one)</td>
<td>☐</td>
<td>☑</td>
</tr>
</tbody>
</table>

(18 questions of a similar format but with different levels. Participant were randomly assigned to one of three version blocks)

[Displayed after question 9]

Make sure that you consider all six characteristics of both medicines in your answers. You are halfway done with this part of the survey! Please keep going! Your answers are important to our research.

Evaluation questions

The following statements refer to the questions that asked about the diabetes medicines. Please indicate if you agree or disagree with the statements in the grid below.

<table>
<thead>
<tr>
<th>Things impacting your own diabetes self-management</th>
<th>Strongly disagree 1</th>
<th>Disagree 2</th>
<th>Neither agree nor disagree 3</th>
<th>Agree 4</th>
<th>Strongly agree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found it easy to understand the questions</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I found it easy to answer all the questions</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I answered all questions in a way consistent with my preferences</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
### Appendix D – Stratified conditional logit analyses

#### Preference estimates for stratified sample based on having experience a hypoglycemic event in the past 6 months

<table>
<thead>
<tr>
<th></th>
<th>No hypoglycemic event</th>
<th>At least one hypoglycemic event</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c decrease</td>
<td>1.00%</td>
<td>0.94%</td>
</tr>
<tr>
<td></td>
<td>-0.20</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>Stable blood</td>
<td>6 dy/wk</td>
<td>4 dy/wk</td>
</tr>
<tr>
<td>glucose</td>
<td>-4.85</td>
<td>-3.51</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Low blood</td>
<td>None</td>
<td>2 dy/wk</td>
</tr>
<tr>
<td>glucose</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Day</td>
<td>0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>Day/night</td>
<td>3.94</td>
<td>4.00</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>30 min</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>10.00</td>
</tr>
<tr>
<td>Treatment</td>
<td>1 pill</td>
<td>2 pills</td>
</tr>
<tr>
<td>burden</td>
<td>0.00</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>Cost</td>
<td>$1</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Wald test chi2(11) = 16.74
Prob > chi2 = 0.1158

#### Preference estimates for stratified sample based on medication use

<table>
<thead>
<tr>
<th></th>
<th>No injection</th>
<th>Injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c decrease</td>
<td>1.00%</td>
<td>0.94%</td>
</tr>
<tr>
<td></td>
<td>-0.22</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>-5.51</td>
<td>-5.51</td>
</tr>
<tr>
<td>Stable blood</td>
<td>6 dy/wk</td>
<td>4 dy/wk</td>
</tr>
<tr>
<td>glucose</td>
<td>-5.40</td>
<td>-3.88</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Low blood</td>
<td>None</td>
<td>2 dy/wk</td>
</tr>
<tr>
<td>glucose</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Day</td>
<td>0.38</td>
<td>0.95</td>
</tr>
<tr>
<td>Day/night</td>
<td>3.81</td>
<td>4.31</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>30 min</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>10.00</td>
</tr>
<tr>
<td>Treatment</td>
<td>1 pill</td>
<td>2 pills</td>
</tr>
<tr>
<td>burden</td>
<td>0.00</td>
<td>-0.35</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.06</td>
</tr>
<tr>
<td>Cost</td>
<td>$1</td>
<td>1 pill</td>
</tr>
<tr>
<td></td>
<td>0.13</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Wald test chi2(11) = 16.74
Prob > chi2 = 0.1158

#### Preference estimates for stratified sample based on survey language

<table>
<thead>
<tr>
<th></th>
<th>English survey</th>
<th>Spanish Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c decrease</td>
<td>1.00%</td>
<td>0.94%</td>
</tr>
<tr>
<td></td>
<td>-0.75</td>
<td>-5.30</td>
</tr>
<tr>
<td></td>
<td>-5.64</td>
<td>-6.58</td>
</tr>
<tr>
<td>Stable blood</td>
<td>6 dy/wk</td>
<td>4 dy/wk</td>
</tr>
<tr>
<td>glucose</td>
<td>-4.27</td>
<td>-3.13</td>
</tr>
<tr>
<td></td>
<td>2.75</td>
<td>0.00</td>
</tr>
<tr>
<td>Low blood</td>
<td>None</td>
<td>Day</td>
</tr>
<tr>
<td>glucose</td>
<td>0.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Day</td>
<td>0.40</td>
<td>3.80</td>
</tr>
<tr>
<td>Day/night</td>
<td>3.80</td>
<td>6.90</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>30 min</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>3.60</td>
<td>10.00</td>
</tr>
<tr>
<td>Treatment</td>
<td>1 pill</td>
<td>2 pills</td>
</tr>
<tr>
<td>burden</td>
<td>0.00</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>Cost</td>
<td>$1</td>
<td>1 pill</td>
</tr>
<tr>
<td></td>
<td>0.12</td>
<td>6.52</td>
</tr>
</tbody>
</table>

Wald test chi2(11) = 16.74
Prob > chi2 = 0.1158
References


74. Craig BM, Runge SK, Rand-Hendriksen K, Ramos-Goni JM, Oppe M. Learning and satisficing: an analysis of sequence effects in health valuation. (1524-4733 (Electronic)).


CHAPTER FIVE

EDUCATION AND PATIENT PREFERENCES FOR TREATING TYPE 2 DIABETES: A STRATIFIED DISCRETE-CHOICE EXPERIMENT

Ellen M. Janssen, Daniel R. Longo ScD, Joan Bardsley MBA, RN CDE, FAADE, and John F.P. Bridges PhD
Abstract

**Purpose:** Diabetes is a chronic condition that is more prevalent among people with lower educational attainment. We sought to assess the impact of educational attainment on treatment preferences of patients with type 2 diabetes.

**Methods:** A diverse sample of patients with type 2 diabetes was recruited from a national online panel in the United States. Treatment preferences were assessed using a discrete-choice experiment (DCE). Participants completed 16 choice tasks in which they compared pairs of treatment profiles composed of 6 attributes: A1C decrease, stable blood glucose, low blood glucose, nausea, treatment burden, and out-of-pocket cost. We estimated choice models using conditional logit analysis and calculated willingness-to-pay (WTP) estimates stratified by educational status.

**Results:** 231 participants with high school or less, 156 participants with some college, and 165 participants with a college degree or more completed the survey. Participants with college or more were willing to pay more for A1C decreases ($58.84, SE: 10.6) than participants who had completed some college ($28.47, SE: 5.53) or high school or less ($17.56, SE: 3.55) (p<=.01). In addition, people with a college education were willing to pay more than people with high school or less to avoid nausea, low blood glucose events during the day/night, or 2 pills per day.

**Conclusion:** WTP for aspects of diabetes medication differed for people with a college education or more and a high school education or less. A thorough understanding of preference heterogeneity across different subgroups can be used in patient-centered benefit-risk assessments and personalized care approaches.
Introduction

Diabetes is a chronic disease that affects 29.1 million people (9.3%) and is the seventh leading cause of death in the United States (US).\(^1\) In 2012 in the US, 1.7 million people 20 years or older (or 7.8 per 1000 people) were newly diagnosed with diabetes and thirty-three percent of people in the US will develop type 2 diabetes during their life time.\(^1\) Diabetes related complications result in significant morbidity and costs\(^2\) and are associated with a reduced quality of life compared to the general population.\(^3\) While the prevalence of diabetes has gone up over time among all educational groups, diabetes remains more prevalent among people with lower educational attainment.\(^4\)

Treatment of type 2 diabetes most often requires a combination of lifestyle adaptation and medication. There exist multiple types of glucose lowering medication: oral medications, non-insulin injections, and insulin.\(^5-8\) These treatments are associated with a range of side effects, including weight and blood pressure changes, gastrointestinal side effects, and cardiovascular risk.\(^5,6,8\) Treatment guidelines emphasize the need to achieve treatment goals, including achieving targeted blood glucose levels, blood pressure control, lipid optimization, and prophylactic aspirin therapy.\(^9-11\)

Treatment outcomes for diabetes have been shown to improve when patients are involved in the treatment decision.\(^12\) Despite the high need for self-management in type 2 diabetes, much clinical research and care continues to neglect the patient perspective\(^13\) resulting in low rates of medication adherence\(^14\) and low rates of achieving treatment goals.\(^15\) When incorporating the patient perspective, the effects of patient characteristics, such as educational attainment, on their treatment preferences should be considered. Not only are people with lower educational attainment more likely to develop diabetes,\(^16-18\) they are also more likely to suffer from complications and may benefit less from medication, self-care interventions, and diabetes
Efforts to deliver diabetes treatments may be more effective if tailored to individuals' preferences and educational background.

This study evaluates treatment preferences of patients with different educational attainment by estimating their willingness-to-pay (WTP) for type 2 diabetes medications. We hypothesize that people with more education are willing to pay more for diabetes medication. Understanding treatment preferences and WTP of patients with different levels of educational attainment might aid in informing and targeting patient-centered policies, treatment, and education efforts. The study illustrates an increasingly important research method for the measurement of patient preference and quantifies the preferences of patients with type 2 diabetes across the benefits, risks, and treatment burden of anti-glycemic medications.

**Methods**

Discrete Choice Experiments (DCEs) are one of the most common form of stated-preference methods, and guidelines have emerged for utilizing them. Our approach was consistent with existing guidelines and is summarized under four headings: conceptual model, creation of the choice tasks, survey respondents, and statistical analysis.

In a DCE, it is assumed that the treatment profile (in this case diabetes medication) is defined by a variety of characteristics, or attributes that can exist at different levels. A respondent is presented with two or more distinct profiles and is asked to select the profile they prefer. Based on participants’ repeated choices the relative preferences for the different attributes and levels can be estimated.

**Conceptual model**

The association between educational attainment and health outcomes is well known. In addition to life expectancy, education is associated with numerous mental and physical health
outcomes, including diabetes. Individuals with lower income and less education are 2 to 4 times more likely to develop diabetes and more likely to be affected by diabetes complications. In addition, treatments, such as intensive lifestyle interventions and metformin, can be more effective in people with higher educational attainment resulting in improved health outcomes.

The association between education and health outcomes can be explained in four ways. First, schooling increases the efficiency with which health is produced. Second, schooling may affect how quickly people can obtain and process information. Third, schooling might teach people how to following instructions. Fourth, schooling may broaden social networks, including access to physicians.

In the reverse association, health outcomes might influence educational attainment; people with poor health outcomes might have lower levels of education. This association is less likely to hold in the case of type 2 diabetes because type 2 diabetes is generally diagnosed later in adulthood when education has been completed.

The education-health association might be mediated by treatment preferences (Figure 5-1). Educational attainment can affect treatment preferences and preference stability. Furthermore, patient’s preferences might affect health outcomes. When patients’ preferences are accommodated, greater adherence to therapy or feelings of control and greater health-related quality of life might results. In diabetes, time preference for the future are significantly related to diabetes complications. A link between patient preferences for treatment burden and adherence to diabetes medications has been shown. In addition, when physicians met their patients’ preferences for information, patients with type 2 diabetes attained better metabolic control, self-reported adherence, and treatment satisfaction.
The link between education and preferences in diabetes needs to be further explored. In a DCE on diabetes treatment, Guimaraes et al. (2009) found that income was related to treatment preferences, but they did not find this link between educational attainment and treatment preferences. However, a DCE published by Hauber et al. (2009) did find a difference in treatment preferences between people with some college education and no college education. In addition, educational attainment has been shown to influence people’s preferences for different types of diabetes self-management programs. This study will examine the link between educational attainment and treatment preferences for type 2 diabetes medication by estimating separate WTP models for people that completed high school or less education, that completed some college, and that completed college or more education.

**Creation of the choice tasks**

This DCE was developed using a rigorous engagement process that included synthesis of the existing evidence, expert consultations, stakeholder engagement, qualitative pretest interviews
(n=25), and quantitative pilot testing (n=27). The final instrument contained six attributes: A1C decrease (0%, 0.5%, 1% decrease), stable blood glucose (2 days/week, 4 days/week, 6 days/week), low blood glucose/hypoglycemia (none, during the day, during the day and/or at night), nausea (none, 30 minutes/day 90 minutes/day), treatment burden (1 pill/day, 2 pills/day, 1 pill and 1 injection/day), out-of-pocket costs ($10/month, $30/month, $50/month). The survey included detailed descriptions on all attributes and levels as well as an explanation and example on how to complete a choice task.

With six attributes at three levels each, we could generate 729 distinct treatment profiles and more than 500,000 paired profile choice tasks. We selected a subset of these possible choice task using specialized software (Ngene, Choice Metrics) to create a D-efficient experimental design that took into account prior preference results from a pilot test. Respondents were randomly assigned to one of three survey versions that each contained 16 choice tasks. A sample choice task is presented in Figure 5-2.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Medicine A</th>
<th>Medicine B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1C levels go down by</td>
<td>0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Stable blood sugar</td>
<td>2 days per week</td>
<td>6 days per week</td>
</tr>
<tr>
<td>Additional low blood sugar events</td>
<td>During the day</td>
<td>None</td>
</tr>
<tr>
<td>Nausea</td>
<td>None</td>
<td>90 minutes per day</td>
</tr>
<tr>
<td>Additional medicine</td>
<td>2 pills per day</td>
<td>1 pill per day</td>
</tr>
<tr>
<td>Additional out-of-pocket costs</td>
<td>$10 per month</td>
<td>$30 per month</td>
</tr>
<tr>
<td>Which medication would you choose?</td>
<td>I would choose Medicine A</td>
<td>I would choose Medicine B</td>
</tr>
</tbody>
</table>
Statistical analysis

We stratified the model by educational status; we estimated separate choice models for participants that had completed high school or less education, participants that had completed some college, and participants that had a college degree or more education. We analyzed the data in Stata 13 (College Station, TX) using a conditional logit model.\(^57\) WTP estimates were calculated by taking the coefficient of each attribute level and dividing by the coefficient for cost:

\[
WTP_{attribute\ A} = \frac{\beta_{attribute\ A}}{\beta_{cost}}
\]

Standard errors were estimated using the delta method. Paired t-tests were used to test for the equivalence of individual coefficients between educational groups. We conducted a Swait-Louviere scale test\(^58\) to test for differences in error variance between the model (Appendix).

Survey respondents

Members of a nationally representative online panel with type 2 diabetes were invited by email to participate in the survey (GfK KnowledgePanel). All participants were required to be 18 years or older with a self-reported physician-diagnosed type 2 diabetes and able to read English or Spanish. African Americans and Latinos were oversampled to account for the high prevalence of diabetes in these population.\(^59,60\) Survey participants received compensation from the online panel equivalent to about $10. The Johns Hopkins School of Public Health IRB determined this study to be exempt from human subjects review (IRB 6001).
Results

552 people (66% response rate). Our participants were more likely to take medication and were living with diabetes longer than a 2011 national sample.61,62 231 participants had completed High School or less education, 156 participants had completed at least some college, and 165 people had completed college or more education (Table 5-1). Participants with more education were more likely to be male, were more likely to report better health status, and had higher incomes. There were no differences in years since diabetes diagnosis, having had a hypoglycemic events in the past 6 months, A1C levels, or medication use.

WTP Results

Table 5-2 presents WTP results stratified by educational attainment. Participants with high school education or less and participants with some college were willing to pay the most to avoid having to take 1 pill and 1 injection a day (WTP for high school or less: $-30.54, SE: 4.11; WTP for some college: $-38.25, SE: 6.10). They were willing to pay the least to avoid having to take 2 pills a day (WTP for high school or less: $-5.22, SE: 2.89; WTP for some college: $-5.50, SE: 3.43). Participants with a college degree or more were willing to pay the most for a 1% decrease in A1C (WTP: $58.84, SE: 10.60). They were willing to pay least to avoid low blood glucose events during the day (WTP: $-9.41, SE: 4.73).

Participants with college or more were willing to pay more for A1C decrease ($58.84, SE: 10.6) than participants who had completed some college ($28.47, SE: 5.53) or participants who had completed high school or less ($17.56, SE: 3.55) (p<=.01). People with a college degree or more were also willing to pay more to avoid 30 minutes of nausea, low blood glucose events during the day and/or at night, and to avoid 2 pills a day. Other differences in WTP between educational groups were not statistically significant.
Table 5-1 Characteristics of sample by educational attainment

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>High School or Less (N = 231)</th>
<th>Some college (N = 156)</th>
<th>College or more (N=165)</th>
<th>P high school = college</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years - Mean (SD)</td>
<td>61.58 (11.27)</td>
<td>61.03 (12.57)</td>
<td>61.18 (11.54)</td>
<td>0.89</td>
</tr>
<tr>
<td>Male – N (%)</td>
<td>100 (43.3%)</td>
<td>86 (55.1%)</td>
<td>93 (56.4%)</td>
<td>0.015*</td>
</tr>
<tr>
<td>Race – N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>136 (58.9%)</td>
<td>80 (51.3%)</td>
<td>73 (44.2%)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Black</td>
<td>34 (14.7%)</td>
<td>45 (28.8%)</td>
<td>47 (28.5%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>56 (24.2%)</td>
<td>28 (17.9%)</td>
<td>35 (21.2%)</td>
<td></td>
</tr>
<tr>
<td>Income equal or greater to $50,000 – N (%)</td>
<td>78 (33.8%)</td>
<td>77 (49.4%)</td>
<td>117 (70.9%)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Years since diabetes diagnosis – Mean (SD)</td>
<td>11.02 (7.73)</td>
<td>10.90 (7.85)</td>
<td>11.88 (7.89)</td>
<td>0.46</td>
</tr>
<tr>
<td>Self-reported health as good or better – N (%)</td>
<td>165 (71.4%)</td>
<td>121 (77.6%)</td>
<td>137 (83.0%)</td>
<td>0.025*</td>
</tr>
<tr>
<td>At least 1 hypoglycemic event in the last 6 months – N (%)</td>
<td>109 (47.2%)</td>
<td>76 (48.7%)</td>
<td>74 (44.8%)</td>
<td>0.78</td>
</tr>
<tr>
<td>Do not know their A1C level – N (%)</td>
<td>176 (76.9%)</td>
<td>138 (88.5%)</td>
<td>147 (89.1%)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Most recent A1C level &gt; 7.0% – N (%)</td>
<td>86 (48.9%)</td>
<td>75 (54.3%)</td>
<td>72 (49.0%)</td>
<td>0.57</td>
</tr>
<tr>
<td>Type of diabetes medicine used – N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No prescription medicine</td>
<td>20 (8.7%)</td>
<td>9 (5.8%)</td>
<td>8 (4.8%)</td>
<td>0.32</td>
</tr>
<tr>
<td>Only pills</td>
<td>135 (58.4%)</td>
<td>98 (63.2%)</td>
<td>112 (67.9%)</td>
<td></td>
</tr>
<tr>
<td>Pills and/or injections/shots</td>
<td>76 (32.9%)</td>
<td>48 (31.0%)</td>
<td>45 (27.3%)</td>
<td></td>
</tr>
<tr>
<td>Minutes to complete DCE – Mean (SD)</td>
<td>13.39 (11.15)</td>
<td>14.08 (18.50)</td>
<td>11.72 (7.69)</td>
<td>0.23</td>
</tr>
<tr>
<td>Agree or strongly agree with the following statement – N (%)</td>
<td>161 (69.7%)</td>
<td>106 (67.9%)</td>
<td>124 (75.6%)</td>
<td>0.27</td>
</tr>
<tr>
<td>Choice tasks were easy to understand</td>
<td>152 (66.1%)</td>
<td>99 (63.5%)</td>
<td>100 (61.0%)</td>
<td>0.58</td>
</tr>
<tr>
<td>Answer were consistent with my preferences</td>
<td>185 (80.1%)</td>
<td>126 (80.8%)</td>
<td>133 (80.6%)</td>
<td>0.98</td>
</tr>
<tr>
<td>Displayed lexicographic preferences – N (%)</td>
<td>42 (18.42%)</td>
<td>19 (12.26%)</td>
<td>15 (9.2%)</td>
<td>0.027*</td>
</tr>
</tbody>
</table>

*significant at the 0.05 level
Table 5-2 Willingness to pay for diabetes medication

<table>
<thead>
<tr>
<th></th>
<th>HS diploma or less (N= 231) WTP a SE b</th>
<th>Some College (N= 156) WTP a SE b</th>
<th>College or more (N=165) WTP a SE b</th>
<th>( p_{HS-SC} )</th>
<th>( p_{SC-CO} )</th>
<th>( p_{HS-CO} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1C decrease</td>
<td>1%</td>
<td>17.56 3.55</td>
<td>28.47 5.53</td>
<td>58.84 10.60</td>
<td>0.08</td>
<td>0.01*</td>
</tr>
<tr>
<td>Stable glucose</td>
<td>4 dy/wk</td>
<td>23.12 3.88</td>
<td>29.82 5.06</td>
<td>35.00 8.17</td>
<td>0.29</td>
<td>0.59</td>
</tr>
<tr>
<td>Nausea</td>
<td>30 min/dy</td>
<td>-14.86 1.92</td>
<td>-19.25 2.83</td>
<td>-24.80 4.74</td>
<td>0.18</td>
<td>0.32</td>
</tr>
<tr>
<td>Low glucose events</td>
<td>None</td>
<td>-10.27 2.47</td>
<td>-9.46 3.49</td>
<td>-9.41 4.73</td>
<td>0.85</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>-16.62 2.88</td>
<td>-19.95 4.00</td>
<td>-31.65 7.51</td>
<td>0.49</td>
<td>0.18</td>
</tr>
<tr>
<td>Treatment burden</td>
<td>1 pill/dy</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
<td>0.95</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>2 pills/dy</td>
<td>-5.22 2.89</td>
<td>-5.50 3.43</td>
<td>-16.22 4.47</td>
<td>0.28</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>1 pill + 1 inj/dy</td>
<td>-30.54 4.11</td>
<td>-38.25 6.10</td>
<td>-47.49 9.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*WTP = Willingness to pay; SE = standard error, \( t \)-test for equivalence between WTP of educational groups; *significant at the .05 level

Figure 5-3 presents how people in different educational groups would allocate $100 dollars between the different aspects of diabetes medication. This budget allocation was done to examine the priorities of the different educational groups under equal budgets. It only considered the most extreme levels for the categorical attributes (nausea, low blood glucose, and 1 pill + 1 injection) to avoid doubling up between attributes.

The biggest differences between educational groups could be seen in allocation for A1C decrease, followed by 1 pill and 1 injection a day. People with a college education or more would allocate $30 for a 1% decrease in A1C, while people with some college would pay $21, and people with high school or less would pay $17. In addition, people with some college or less education were willing to pay more for 4 days of stable blood glucose a week than for a 1% decrease in A1c. People with a college education or more would allocate $24 to avoid 1 pill and 1 injection a day, while people with some college would pay $28, and people with high school or less would pay $30. All groups would be willing to pay the least ($13-14) to avoid 30 minutes of nausea a day.
Figure 5-3 Budget allocation for attributes of diabetes medication by educational attainment

<table>
<thead>
<tr>
<th>Attribute</th>
<th>HS diploma or less (N= 231)</th>
<th>Some College (N= 156)</th>
<th>College or more (N=165)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1C decrease 1%</td>
<td>$17</td>
<td>$23</td>
<td>$30</td>
</tr>
<tr>
<td>Stable glucose 4 dy/wk</td>
<td>$21</td>
<td>$22</td>
<td>$30</td>
</tr>
<tr>
<td>Nausea 30 min/dy</td>
<td>$14</td>
<td>$14</td>
<td>$13</td>
</tr>
<tr>
<td>Low blood glucose day/night</td>
<td>$16</td>
<td>$15</td>
<td>$16</td>
</tr>
<tr>
<td>1 pill + 1 inj/dy</td>
<td>$30</td>
<td>$28</td>
<td>$24</td>
</tr>
</tbody>
</table>

Discussion

DCE provides researchers with a theoretically grounded method for the study of preferences. Stated-preference methods are particularly valuable in the healthcare realm because we often cannot observe people’s real world choices when medications are in development or are not yet approved. We conducted a DCE to assess WTP in patients with type 2 diabetes for diabetes medication. We showed that patients with type 2 diabetes do not only value reductions in their A1C levels, but also had preferences for medications that stabilized their daily glucose levels and that reduced burden/harms of medication.

We found that patients with different educational attainment displayed different preferences for diabetes treatment. In particular we found that patients with a higher education were willing to pay more for all medication attributes than patients with lower educational attainment. For example, patients with a high school diploma or less education were willing to pay less than half for a 1% reduction in A1C than participants with college or more. In addition, patients with a college degree or more were willing to allocate the most of their budget to A1C decrease, while
patients with some college or less education allocate most of their budget to avoiding 1 pill and 1 injection per day. We also found that people with some college or less education were willing to allocate more for 4 days of stable blood glucose a week than for a 1% decrease in A1c.

This study has several limitations. First, weight was not included as an attribute based on our instrument development process. This decision was partly made because weight gain could be interpreted as either positive or negative depending on the person and their current weight. While it is generally assumed that patients with type 2 diabetes would benefit from weight loss, this does not apply to some elderly, frail patients. To avoid estimation ambiguities that could arise due to this issue, and after consulting with diabetes experts and patients with diabetes we decided to not include weight changes as an attribute. Unless weight changes were dependent on one of the included attributes, this should not have affected the estimated importance of each attribute.

Second, a previous study found that people with lower educational attainment were less consistent in the choices they made in a DCE. If participants do not understand the choice tasks or attributes, they might be less consistent in the choices they make and their estimated preference weights are biased towards the null. In our results, we observed that people with a lower education are willing to pay less for every medication attribute. However, participants with lower educational attainment did not report that they had more difficulty with answering the choice tasks consistently. In addition, participants with lower educational attainment did not report more difficulty with understanding and/or answering the choice tasks. This suggests that their lower willingness to pay might not be due to lower choice consistency or lack of understanding.

Finally, to maintain sufficient power, the analyses could not be controlled for other variables such as race, or income. Therefore it is not clear whether observed differences in preferences were due to educational differences or other differences in the groups. We did conduct a stratified analysis by income, which showed that between two income groups (below $50,000 and above $50,000 a
year) WTP differed only on A1C decrease and avoiding 1 pill + 1 injection per day (Appendix).

Segmentation techniques using finite mixture models\(^{64,65}\) might be more appropriate to detect subgroup heterogeneity. These techniques group individuals based on the preferences they displayed. The characteristics of people that make up these preference-based groups can then be observed to detect multiple factors that are predictive of preferences.

Stratification approaches present a way to analyze preference heterogeneity based on predefined subgroups, but are limited by the number of subgroups that can be examined. Other analytical techniques might be able to predict factors that influence preferences. A thorough understanding of preference heterogeneity can be used in patient-centered benefit-risk assessments and personalized care approaches.

**Conclusion**

The findings of this study help clarify what aspects of diabetes medication are important to patients and describe differences in WTP of patients with different educational attainment. Our results suggest that patients with lower levels of education might be more bothered by treatment burden. In addition, we found that people with lower educational attainment might place more value on keeping their glucose levels within a daily target range rather than maintaining controlled A1c levels. Being aware of patient preferences and how these preferences are distributed by demographic characteristics can help clinicians tailor treatment approaches to particular patients.
References


20. Kim SH. Educational attainment moderates the associations of diabetes education with health outcomes. LID - 10.1111/ijn.12454 [doi]. (1440-172X (Electronic)).


CHAPTER SIX

- 

ASSESSING THE VALIDITY AND RELIABILITY OF APPLICATIONS

OF DISCRETE-CHOICE EXPERIMENTS IN HEALTHCARE

Ellen M. Janssen and John F.P. Bridges PhD
Abstract

While discrete-choice experiments (DCE) and other stated-preference methods are now routinely applied in healthcare, their recent acceptance by regulatory and health technology assessment (HTA) agencies have placed a greater focus on demonstrating validity and reliability of preference results. Based on a targeted literature review, we identify a variety of methods to test for validity and/or reliability of DCEs. We identify four domains for the validity of a DCE: measurement validity, preference reliability, decision processes, and choice rationality. These domains consist of 14 components that can be identified using 24 possible tests of validity and reliability. We discuss each test and direct readers to applications of these tests in the healthcare DCE literature. We discuss study design considerations that need to be made to test for the validity and reliability of a DCE and consider limitations to the current application of validity tests for DCEs in healthcare.
Introduction.

In recent years there has been a general trend towards patient involvement in treatment,\(^1,2\) research\(^3\) and regulatory decision making.\(^4,5\) As regulatory and health technology assessment decision making\(^6,7\) is becoming more patient-centered, ways to measure preferences of health care stakeholders are being explored.\(^8\) Patient preferences play an important role in many quantitative approaches to benefit-risk assessment\(^9\) and stated-preference methods, such as discrete-choice experiments (DCEs) are increasingly used to study benefit-risk.\(^10\) A growing number of HTA agencies are now formally incorporating patient preferences into their decision making.\(^6,11,12\) The United States (US) Food and Drug Administration (FDA) has already provided guidance on quantitative preference assessment for medical devices.\(^13\)

To be able to fully incorporate patient-preference information in healthcare decision making processes, preference studies will need meet evidence standards consistent with standards for clinical evidence.\(^8\) Guidance from the United States Food and Drug Administration on incorporating patient preference information in benefit-risk assessments calls for checks on the quality of stated-preference studies,\(^13\) including the logical soundness and validity of a patient preference results. However, validating stated-preference assessments is difficult due to the hypothetical nature of the choices participants make.\(^14\) Therefore, it is unknown whether patients would display the same preferences if they would experience the consequences of their choices.

Validity of stated-preference methods when compared to revealed preferences has been explored in transport, marketing, and environmental economics.\(^15,16\) Opportunities to observe preferences through real-life choices in healthcare are be limited\(^10\) because healthcare options might not be available on the market yet,\(^13\) or choices are made on behalf of the patient by a physician\(^17\) or a third-party payer.\(^18,19\) A recent systematic review in the application of DCEs in environmental economics found that environmental DCEs generally provide limited evidence on the internal
reliability and validity of DCEs. The Medical Device Innovation Consortium (MDIC) has called for studies that evaluate specific aspects of the validity of DCEs that are applied to healthcare. While literature reviews on the application of DCEs in healthcare have discussed tests of internal validity, the tests discussed in these reviews has not been comprehensive.

Furthermore, these discussions often do not provide details on how to conduct tests for validity and reliability.

The MDIC report defines validity of a stated-preference study as “the extent to which quantitative measures of relative importance or tradeoffs reflect the true preferences of patients.” We used this definition to identify specific tests in the DCE literature in healthcare that might assess the validity and reliability of a discrete-choice study. We conceptualize the validity and reliability of a DCE into four domains. Then we compile a comprehensive list of tests of validity and reliability applied in the healthcare DCE literature. Furthermore, we explain how these tests can be conducted and refer readers to applications in the healthcare literature. We then discuss considerations that need to be taken into account when designing a DCE to allow for tests of validity and reliability. This review can help researchers identify and apply tests into their studies that can be used to examine the validity of their results.

**Methods**

We conducted a targeted literature review to identify articles that discussed validity of DCEs. We were only interested in tests of validity and reliability that were mentioned in the DCE literature in healthcare. We searched PubMed Central for stated-preference studies that used the word validity and/or reliability and were published between January 2000 and January 2017. While we were only interested in DCEs, we included more general search terms related to stated-preference studies including “patient preference”, "stated preference”, "conjoint analysis”, and “choice experiment” to avoid excluding relevant studies. We also conducted a hand search to identify
studies that reviewed DCEs or that assessed the reliability or validity of a DCE but were omitted in the targeted search.

We included papers that gave an overview of DCEs as a stated-preference method and discussed the theory and methods of DCEs and that mentioned ways to ensure validity or reliability of results. Select empirical papers were included to illustrate how tests of validity and reliability can be applied with examples from the literature. We excluded studies that did not discuss DCEs, or that did not discuss tests that could be done to evaluate the validity or reliability of a DCE. We also excluded studies that were not related to healthcare or that only employed qualitative methods. From the included literature we extracted tests of validity and reliability. We grouped tests into domains and components based on the concept of validity they addressed.

Results

Targeted Literature Search

Using the PubMed search, we identified 139 studies. From the hand search, we identified an additional four articles. After title and abstract review, we excluded 109 studies. 42 studies went through full text review. Thirteen of these studies were reviews or made a methodological contribution in describing and applying a test (Table 6-1). Four were literature reviews,\textsuperscript{21-23,25} three provided overviews on how to conduct stated-preference studies,\textsuperscript{26-28} and seven provided a methodological explanation of a particular test with an empirical illustration.\textsuperscript{28-34}
### Table 1: Articles included in targeted literature review with discussed tests of validity and reliability

<table>
<thead>
<tr>
<th>Validity/reliability test</th>
<th>Measurement validity</th>
<th>Preference reliability</th>
<th>Decision process</th>
<th>Rationality of choice</th>
<th>Type of paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Content validity</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convergent validity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External validity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Preference reliability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-retest stability</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Version consistency</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Decision process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compensatory preferences</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attribute dominance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attribute non-attendance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level recoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task non-attendance</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rationality of choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotonicity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Within-set dominated</strong></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Across-set dominated</strong></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitivity</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Sen's consistency</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sen's contraction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sen's expansion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type of paper</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review paper</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guide</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Types of tests

We identified four domains that define the validity and reliability of a DCE: Measurement validity, preference reliability, decision processes, and choice rationality (Figure 6-1). Measurement validity and preference reliability capture traditional measures of validity and reliability including how accurately the instrument measures preferences, how consistently it measures preferences, and the generalizability of these preference measurements. Decision processes and rationality of choice examine whether the results adhere to assumptions specific to random utility theory and consumer choice behavior. The four domains consist of 14 components that can be examined through 24 identified tests. Below we discuss each domain and its components. The 24 identified tests are discussed as part of the component of validity that they assess. In Table 6-2 we present a summary of each test and provide a reference for an application of each test.

Figure 6-1 Domains and tests to assess the validity and reliability of a DCE

- Measurement validity
  - Face validity
  - Content validity
  - Convergent validity
  - External validity

- Preference reliability
  - Test-retest reliability
  - Test-retest stability
  - Version consistency
  - Prediction of choices

- Choice rationality
  - Monotonicity
  - Transitivity
  - Sen’s consistency

- Decision process
  - Compensatory preferences
  - Level recoding
  - Task non-attendance
<table>
<thead>
<tr>
<th>Measurement validity</th>
<th>Definition</th>
<th>How to test</th>
<th>Study design</th>
<th>Application example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>Whether results are consistent with a priori preference expectations.</td>
<td>Compare study results to previously formulated hypothesis.</td>
<td></td>
<td>Ryan et al. (2001a) De Bekker et al. (2010)</td>
</tr>
<tr>
<td>Content validity</td>
<td>Whether the choice task accounts for important preference attributes given the research question.</td>
<td>Evaluate instrument development process, ask clinical experts/patients about relevance of included and/or omitted attributes.</td>
<td></td>
<td>Kenny et al. (2003) Muhlbaier et al. (2009)</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Whether the results are consistent with other measures that measure the same construct.</td>
<td>Conduct multiple preference experiments using different preference elicitation methods. Add a second preference experiment (different method) to an existing experiment.</td>
<td>x</td>
<td>Bijlenga et al. (2009) Hollin et al. (2016)</td>
</tr>
<tr>
<td>External validity</td>
<td>Whether the results can accurately predict preferences and choices outside of the study context.</td>
<td>Compare study results to revealed preference results (adjust for scaling effects).</td>
<td></td>
<td>Mark and Swait (2004) Teuter and Zweifel (2007)</td>
</tr>
<tr>
<td>Preference reliability</td>
<td>Test-retest reliability</td>
<td>Whether the instrument measures preference consistently over time.</td>
<td>Repeat the same DCE with the same responders and compare responses.</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Test-retest stability</td>
<td>Whether the instrument measures preferences consistently for the duration of the survey.</td>
<td>Include the same choice task in the experiment twice and compare responses within individuals.</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Version consistency</td>
<td>Whether different versions of the instrument result in consistent preference estimates</td>
<td>Include one or more hold-out tasks that are the same across survey versions and compare responses across survey versions. Make small changes to survey versions and compare results between versions.</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Predictive accuracy</td>
<td>Whether the instrument can predict choices outside the choice model (within the instrument) accurately</td>
<td>Include one or more hold-out tasks in the instrument that are not used in the choice model estimation (either additional choice tasks, or different survey versions). Use the choice model to predict choices for these hold-out tasks and compare to observed choices.</td>
<td>x</td>
</tr>
<tr>
<td>Decision processes</td>
<td>Compensatory preferences</td>
<td>Whether participants trade between all attributes of treatment profiles.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute dominance</td>
<td>Whether participants focus on one attribute only when making decisions.</td>
<td>Determine whether certain participants always choose the treatment profile that has the better level of a particular attribute.</td>
<td>Ho et al. (2015)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Attribute non-attendance</td>
<td>Whether participants ignore attributes when making decisions.</td>
<td>Use a multi-step latent class analysis approach to determine whether participants ignore certain attributes in their decisions.</td>
<td>Lagarde et al. (2015)</td>
<td></td>
</tr>
<tr>
<td>Level recoding</td>
<td>Whether participants make choices using the absolute values of numeric attributes.</td>
<td>Conduct a scope test using multiple survey versions with some overlapping numeric attribute levels but with levels that span different ranges. Conduct a simple scope test in which difference between numeric attribute levels are not linear.</td>
<td>Ho et al. (2015)</td>
<td>Johnson et al. (2011)</td>
</tr>
<tr>
<td>Task non-attendance</td>
<td>Whether participants pay attention to the choice tasks.</td>
<td>Examine whether certain participant always choose the treatment profile that occurs in a certain position in the choice task.</td>
<td>Bridges et al. (2010)</td>
<td>Ho et al. (2015)</td>
</tr>
</tbody>
</table>

**Rational choices**

<table>
<thead>
<tr>
<th>Monotonicity</th>
<th>Whether participants prefer treatment profiles that provide more utility.</th>
<th>x</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-set dominated pairs</td>
<td>Whether participants choose the treatment profile from a choice task that has better attribute levels than the other choice task</td>
<td>Include a dominated choice task in the design.</td>
<td>x</td>
</tr>
<tr>
<td>Across-set dominated pairs</td>
<td>Whether participants choose the preferred treatment profile if one profile dominates another treatment profile across choice tasks</td>
<td>Include multiple choice tasks that are related. Include a choice task between Treatment A and treatment B, include a choice task between Treatment A and C. Ensure that Treatment C has better attributes than treatment B</td>
<td>x*</td>
</tr>
<tr>
<td>Transitivity</td>
<td>Whether participants make choices that meet the transitive property (if a participant prefers A over B and B over C, they should prefer A over C)</td>
<td>Include three choice tasks that ask participants to choose between treatment profile A and treatment profile B, between treatment profile B and C, and treatment profile A between C.</td>
<td>x***</td>
</tr>
<tr>
<td>Sen's consistency</td>
<td>Whether participants make choices that follow Sen's contraction and expansion principle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contraction principle</td>
<td>If a choice set is narrowed, then no unchosen alternatives should be chosen now and no chosen alternatives should be unchosen now.</td>
<td>Present respondents an initial choice task and a follow-on choice task. The follow-on contraction choice task should present fewer choice tasks then the initial choice task.</td>
<td>x</td>
</tr>
<tr>
<td>Expansion principle</td>
<td>If a choice set is expanded, then no unchosen alternatives from the original set should be chosen now.</td>
<td>Present respondents an initial choice task and a follow-on choice task. The follow-on expansion choice task should present more choice tasks then the initial choice task.</td>
<td>x</td>
</tr>
</tbody>
</table>

*Can be tested by including an additional choice task or by including instrument versions.
*Necessary choice tasks might occur naturally in the experimental design, but the researchers needs to check the choice task properties.
**Simplified scope test can be conducted with a particular attribute level setup; this simplified test does not require multiple survey versions.
***To test for transitivity, at least two additional choice tasks need to be included.
Measurement validity

Measurement validity captures whether the results from a DCE meet face validity, content validity, convergent validity, and external validity. These concepts identify how accurately the instrument measures preferences and how generalizable these preferences are to other circumstance.

Face validity

Face validity, also referred to as theoretical validity, \textsuperscript{21,22,25,27} assesses the extent to which the results are consistent with \textit{a priori} expectations. We prefer the term face validity as theoretical validity can also refer to the axioms of consumer welfare theorem. \textsuperscript{28,33} Face validity of results can be examined by setting \textit{a priori} hypotheses for the preference relation between the attribute levels. For example, it can be expected that a high clinical benefit will be seen as more positive than a low benefit. This hypothesis can then be checked by looking at the direction of the preference estimates. \textsuperscript{28,37} Assessing face validity is a relatively simple test of validity; about 60\% of DCEs published between 2009–2012 reported it. \textsuperscript{21}

Ryan et al (2003) suggest that considering how results match “intuition” can assess face validity. For example, it might be expected that patients that have experience with an illness value treatment differently than the general public. \textsuperscript{23} We suggest that researchers use caution using this intuitive approach because there are many ways to establish preference heterogeneity.

Content validity

Content validity refers to the extent to which the DCE takes account of all things deemed important in the construct’s domain. \textsuperscript{25} In a DCE, this refers to the attributes and levels included in a study. While not every attribute that is important to every possible respondent needs to be
included, it is crucial to capture attributes important to a majority of participants.\textsuperscript{26} Content validity cannot easily be assessed through quantitative means. Rather it can be assessed by examining the instrument development process.\textsuperscript{27,38} for example whether experts were consulted and whether pretest or pilot tests were conducted.\textsuperscript{34} In addition, clinical experts and study participants can shed light on whether important attributes were omitted and whether the instrument was deemed relevant.\textsuperscript{23,39}

\textit{Convergent validity}

Convergent validity measures the extent to which the results from the DCE are consistent with other measures that are believed to measure the same construct.\textsuperscript{25} Preferences elicited for the same treatment from the same population are expected to be similar regardless the method used to elicit these preferences. Comparisons can be drawn between DCE, visual analog scale, and time-trade off\textsuperscript{40} or direct elicitation of WTP.\textsuperscript{41} Participants should either complete both types of instruments, or should be randomized to each instrument to minimize sample biases.

Comparisons can also be drawn between different stated-preference studies in the same preference elicitation instrument. For example, Hollin et al. (2016) conducted a best-worst scaling experiment to measure treatment preferences. They included a choice based discrete choice question after each BWS task that asked whether participants would accept the treatment.\textsuperscript{42}

If resources are not available to conduct two types of experiments, researchers can compare preference estimates they obtain to those reported in the literature.\textsuperscript{43,44} In this case, care should be taken that preference estimates are appropriately compared: different studies might use different estimation methods and slightly different attributes. In these cases it can help to standardize these preferences by, for example, calculating willingness-to-pay estimates.\textsuperscript{34}
**External validity**

External validity, sometimes referred to as criterion validity, refers to the extent to which the preference results obtained from the instrument can be generalized to situations and people outside of the study. In stated-preference studies it is often described as the capability of a model to accurately predict preferences and choices outside of the immediate model context.\textsuperscript{31} For DCEs this is most easily examined by studying the choices people make in the real world, their revealed preferences, and comparing those to the preference estimated in the DCE.\textsuperscript{34} This can be done by jointly estimating revealed and stated preference estimates.\textsuperscript{45} In the healthcare sector, there have been relatively few studies that have examined external validity of DCEs,\textsuperscript{21} partially because opportunities to observe preferences through real-life choices in healthcare might be limited.\textsuperscript{10}

**Preference reliability**

Preference reliability captures whether the DCE produces similar results under consistent conditions. Preference reliability includes concepts of test-retest reliability, test-retest stability, version consistency, and predictive accuracy. It measures whether results can be reproduced or repeated over a given time. However, random utility theory, on which stated-preference methods are based, includes a random component. Therefore, reliability cannot be expected to be perfect.

**Test-retest reliability**

Test-retest reliability refers to the extent to which the instrument results in consistent preference estimates over time. This is tested by asking a sample of respondents to complete the same DCE twice at different times. The results are then compared with those from the first time the DCE was administered.\textsuperscript{40} Test-retest reliability can be examined by calculating the correlation coefficient \((r)\) when using continuous variables and the Kappa coefficient \((\kappa)\) when using categorical variables.\textsuperscript{25} Test-retest reliability can be problematic if a long period is allowed to pass between
the exercises, as people’s circumstances and preferences can change. In that case, the study wouldn’t be testing for test-retest reliability, but rather for the temporal stability of preferences.46

**Test-retest stability**

Test–retest stability refers to the extent to which preferences are consistent for the duration of the survey.33 To test for test-retest stability, the same choice task needs to be presented to participants twice. The proportion of participants that answer this repeat-choice task the same way twice can then be established.33 A Kappa coefficient can also used to examine the difference in expected and observed consistency.

Test-retest stability is different from test-retest reliability in that it measures choice consistency for only one choice task. Test-retest stability has the advantage that the instrument only needs to be administered once and that preferences are less likely to change over the course of single instrument than over the period between consecutive survey administrations. However, test-retest stability only measures consistency for a few (at minimum one) choice tasks and might therefore give a less accurate view on reliability of preferences. For example, if the repeat choice task is first presented at the beginning and then at the end of the instrument, learning effects or fatigue might affect the test-retest stability.33 In addition, the options presented in the repeat task might affect test-retest stability. For example, a task might include options that many people are indifferent to. This increases the probability that they choose a treatment profile almost randomly. In this case test-retest stability can be expected to be low, even if test-retest reliability for the entire instrument could still be relatively high.

**Version consistency**

Version consistency refers to the extent that different versions of the same DCE result in consistent preference estimates. In experimental designs in which different respondents see
different combinations of choice tasks, such as blocked or individualized designs, one or more hold-out of fixed choice tasks that are the same for each respondent can be included. Responses to this hold-out task can then be compared across survey blocks or across random splits in the sample using a simple $t$-test. The more hold-out tasks are included across the survey versions, the more accurately version consistency can be established. However, adding extra tasks to the experimental design decreases statistical efficiency and increases response burden.

Another way to test version consistency is by making small changes to survey versions, by for example using labeled and unlabeled decision scenarios or in the framing of attributes. This approach has the disadvantage that by making changes to the survey a different preference construct might be measured and responses are not comparable.

**Predictive accuracy**

Predictive accuracy refers to the extent to which the choice model predicts choices outside the model but within the DCE. One or more choice tasks, holdout tasks or fixed tasks, which are not part of the experimental design of the DCE, can be included in the instrument. Then, preference estimates obtained from the choice tasks that are part of the experimental design can be used to predict the probability that each treatment profile in the additional choice task(s) is chosen. These predictions can be compared to the observed proportion of participants that chose each profile.

Another method to test predictive validity is to take use multiple surveys. Preference estimates for one of the survey versions can be estimated and used to predict the choices for the other survey version. Muhlbacher and Johnson (2016) note that it may be easier to predict choices for one task than for another choice task which could lead to misleading results of this test.
**Decision processes**

Decision processes examine whether participants engage with the choice task in the intended manner and whether they are willing to accept trade-offs within the ranges of levels in the study design. It includes concepts such as task non-attendance, level recoding, and compensatory preferences.

**Compensatory preferences**

Compensatory preferences refer to the extent that participants trade between all attributes of treatment profiles. DCEs assume that participants consider all attributes in their decision making and that there is always some improvement in one attribute that can make up for the reduction in another. Non-compensatory preferences can either be indicated by attribute dominance or attribute non-attendance.

Generally, compensatory preferences have been tested by looking at attribute dominance. In attribute dominance choices are solely based on one attribute. Attribute dominance is present if a participant chooses the treatment profile with the best level of a particular attribute for every choice task.

When attribute non-attendance is present, respondents make choices while ignoring one or more attributes. Attribute non-attendance can be examined through a multi-step latent class analysis approaches. Attribute non-attendance might be a more comprehensive approach to examining non-compensatory preferences in DCE, but is also more complicated to test. Lagarde (2013) argues that attribute dominance can also be conceptualized as attribute non-attendance because all attributes but the dominant one are ignored in the decision making process.
**Level recoding**

Level recoding refers to the extent that participants process the actual numbers presented in continuous or numeric attributes. Recoding may be a strategy for simplifying evaluations of a relatively unfamiliar but important attribute. Instead of interpreting the actual number and the numeric differences between attribute levels, participants might recode the levels to ‘low,’ ‘medium,’ and ‘high’ categories.

Recoding can be tested using a scope test, which involves creating two survey versions. For the numeric attribute, these two survey versions need to include some of the same levels (overlapping levels) and at least one different level so the level range is different for the two surveys. If preference estimates are the same for overlapping numeric levels between the two survey versions, we can infer that respondents did not recode the levels but rather reacted to the absolute levels. If the preference estimates for the overlapping levels are different, this suggest that participants recoded the numeric values as “high,” “medium,” or “low.”

An approximation of the scope test, a simple scope test, can be done using just one survey version with a large difference and small difference between the levels of the numeric attribute included. If the difference in preference estimates between the large and the small level difference are similar, this might indicate that participants recoded the numeric levels as ‘low,’ ‘medium,’ and ‘high.’ However, the simple scope test does not determine with certainty that recoding occurred because the preference function might not be linear. Figure 6-2 gives a graphical illustration of the full scope test and the simple scope test.
Panel a. Simple scope test. In this chart the cost attribute appeared at three levels $0, $3, and $10. The red line indicates no recoding took place. The blue line indicates recoding might have taken place – the difference in preference estimates between the small increase in cost ($0-$3) is the same as the difference in preference estimates between the large increase in cost ($3-$10). However, this could also indicate diminishing marginal utility.

Panel b and c. In these charts, two survey versions are represented. In version 1 the cost attribute appeared at $0, $3, $10. In version 2, the cost attribute appeared at $0, $6.

Panel b. Scope test – no recoding. In this chart, the preference estimates for $0, $3 are the same for version 1 and version 2. This indicates that participants traded based on the numerical values of the cost attribute.

Panel c. Scope test – recoding. In this chart, the preference estimates for $0, $3 are different for version 1 and version 2. This indicates that participants recoded attributes to “low”, “medium”, “high”.

Figure 6-2 Illustration of the Scope test and Simple Scope test to test for level recoding
**Task non-attendance**

Task non-attendance refers to the extent that participants pay attention to the choice tasks. It is assumed that participants actively engage in the choice task, but if the task is too cognitively burdensome or not realistic, they might not carefully consider their decisions.²⁷ Participants that always choose the treatment profile in a position in the choice task, for example the treatment profile on the right, are likely not paying attention to the choice task.⁵³,⁵⁴

**Rational choices**

Rational choices examine whether participants make choices in accordance with consumer choice theory. This theory assumes choices reflect completeness, reflexivity, transitivity, continuity, monotonicity or non-satiation, convexity, and Sen’s expansion and contraction principles.²²,³³ Using DCEs we can measure whether choices adhered to monotonicity, transitivity, and Sen’s expansion and contraction principles.

**Monotonicity**

Monotonicity, also referred to as non-satiation, means that people will prefer “better” levels of an attribute, or attributes that give them higher utility. Two types of monotonicity can be tested: within-set monotonicity and across-set monotonicity.²⁷,³³

Within set monotonicity tests for non-satiation within a choice task. It can be tested by adding a dominant choice task to the instrument design. This means that for one of the treatment profiles in a choice task all the attribute levels are just as good as the attributes levels of the other profile and at least one attribute level is better.²⁸,³²,³³ If uncertainty about how participants value certain attributes exists, these attributes should be held constant across the treatment profiles in the choice task. To satisfy the monotonicity requirement, a participant should choose the preferred
treatment profile. However, there in a two-profile case there is a 50% probability that someone that doesn’t pay attention to the choice tasks will pass the test for within-set monotonicity. 33

Across-set monotonicity tests for non-satiation across choice tasks; it checks whether people choose the preferred treatment profile if one profile dominates another across choice tasks. For example, participants might choose treatment B over treatment A in one choice task. If all of treatment C’s attribute levels are better than for treatment B, they should also choose treatment C over treatment A in another choice task. Treatment profiles that can be used to test for cross-set monotonicity can occur spontaneously in an experimental design, especially if a constant opt-out or status-quo option is included.33 However, researchers might need to add one or more choice tasks to test for cross-set monotonicity.

**Transitivity**

Transitivity refers to the preference relation between three or more treatment profiles. In particular, if Treatment A is preferred to Treatment B in one choice task, and Treatment B is preferred to Treatment C in another choice task, then the transitive property states that Treatment A should be preferred to Treatment C in a third choice task. To test for transitivity, at least three related choice tasks need to be included in the DCE instrument. The first choice-task, the choice between Treatment A and Treatment B, can be part of the regular experimental design. The other two choice tasks (Treatment B v. C and Treatment A v. C) will most likely need to be added to the instrument outside of the existing experimental design.

**Sen’s consistency**

Sen’s consistency principles provide a more stringent test of rationality than the traditional monotonicity tests.28 Sen’s consistency principles 55 consist of both the contraction principle and the expansion principle. Sen’s contraction principle states that if a choice set A is narrowed (to B)
and some of the chosen from A are still in B, then no unchosen alternatives should be chosen now
and no chosen alternatives should be unchosen now. Sen’s expansion principle states that if a
choice set A is expanded (to C) and some of the chosen from A are still in C, then no unchosen
alternatives from A should be chosen now. To test the contraction or expansion property,
respondents should be presented with an initial choice task and a follow-on choice task.

For the contraction principle, the initial choice task should contain at least three treatment choices
(e.g. Treatment A, Treatment B, no treatment). The follow-on contraction choice task should
present fewer choice tasks then the initial choice task (e.g. Treatment A, Treatment B). The
contraction property is satisfied if a person that chose Treatment A in the initial choice task, still
chooses Treatment A in the follow-on contraction choice task.32

For the expansion principle, the first choice task should contain at least two treatment options
(e.g., Treatment A, Treatment B). The follow-on expansion choice task should present more
choice tasks then the initial choice task (e.g. Treatment A, Treatment B, no treatment). The
expansion property is satisfied if a person that chose Treatment A in the initial choice task, does
not chose Treatment B in the follow-on expansion choice task.32

Study design considerations

To be able to conduct many of the discussed validity tests, particular adaptations to the study
design of a DCE need to be made (Table 6-2). These adaptations might increase study costs or
might decrease statistical efficiency or increase response burden. Face validity, content validity,
attribute dominance, and attribute non-attendance do not require special study design
considerations so can be relatively easily incorporated into almost any DCE if the entire study
process, including instrument development, is transparently reported. While attribute non-
attendance does not require special study design considerations, it does require the use of a
relatively complex analytical process. In addition, external validity does not require changes to the study design, but revealed preferences might not exist and/or might require complex statistical analyses. Therefore, these tests might be more difficult to conduct.

To be able to test for monotonicity, transitivity, Sen’s consistency principles and test-retest reliability at least one additional choice task needs to be added to the study design. To be able to test for level recoding, different instrument versions need to be administered to different participants. To be able to test for test-retest reliability, the DCE needs to be administered twice to the same group of participants. Convergent validity can be tested by conducting a second choice experiment either as part of the DCE instrument or outside of it. Version consistency and prediction accuracy can be tested by either adding a fixed-choice or hold-out task or by administering multiple survey versions to different participants.

In the cases where validity tests can be performed using multiple study design adaptation, the type of adaptation that requires more resources generally presents a more stringent test. For example, to test for level recoding, a scope test requires that different survey versions are administered to different participants. The simple scope test only requires the adaption of the numeric attribute levels, but this test is not conclusive whether recoding occurred or whether preferences are not linear. Generally, more stringent test for choice rationality require larger adaptations to the study design. Within-set monotonicity is regarded the weakest test of choice rationality because it can relatively easily be passed by chance. Testing for within-set monotonicity only requires one additional choice task. Transitivity and Sen’s consistency theory are more stringent tests of rationality but require larger changes to study design. Testing transitivity requires the addition of at least two choice tasks. Testing Sen’s consistency principle requires the addition of at least one choice task that doesn’t follow the format of the other choice tasks which might add complexity to the instrument.
Discussion

While a variety of tests exist to assess the validity of a DCE, important questions remain. It is unclear what to do with participants that fail tests for validity or reliability. Excluding participants that fail the test for compensatory preferences or choice rationality might increase internal validity but will decrease external validity of results. Deleting responses can increase sample selection bias and increase the statistical efficiency and power of the estimated choice models. In addition, participants that fail these tests might have made their choices according to their actual preferences. Follow up questions or debriefing interviews could be shed light on the way participants answer choice tasks.

Furthermore, it remains unclear exactly when a test of validity or reliability is met. This is especially difficult for tests that require some qualitative or subjective assessment such as face or content validity. However, ambiguous requirements for more quantitative tests exist as well. For example, the acceptable frequency of dominant preferences, or failure rates of rationality tests or test-retest stability in a study are not well-established. It is also not clear what tests need to be met for a DCE study to be considered valid. It is unlikely that one study can incorporate each test of validity or reliability, especially because many tests require an adaptation to the study design.

This study has several limitations. First, we conducted a targeted review of only one database that only spanned the past 10 years. Thus, certain articles discussing tests of validity and reliability might have been omitted. However, because most articles discussed similar tests of validity and reliability and because we supplemented our targeted literature search with a hand search, we are confident that we identified the most common tests.

Second this study does not discuss the results of validity and reliability tests; it does not discuss whether DCEs published in the literature meet tests of validity and reliability, but discusses the
types of tests that could be performed. The purpose of this manuscript was to provide researchers with various tests they could consider when conducting or evaluating stated-preference studies. We did not want to conduct a study on the validity of published DCEs.

Third, this study did not discuss Lanscar and Swait’s reconceptualization of external validity as a process rather than an outcome measure. Lanscar and Swait raise excellent points in how to ensure that the process of conducting a DCE supports external validity and suggest that this can be done by employing qualitative methods. Qualitative methods in the development and implementation of a DCE are also important to ensure content validity, face validity, and to examine rationality of choices. We chose not to discuss this conceptualization as it touches upon many other aspects of validity and is related to the implementation of good research practices for instrument development.

**Conclusion**

Due to the complexity of DCEs, most studies will neither include nor meet all tests of validity and reliability identified in this review. More studies should be conducted to investigate the determinants of DCE validity. A standard set of tests to be included in every DCE and cut-off points at which these tests are or are not met should be established. While limitations in establishing reliability and validity of a DCE remain, this study can help researchers gain insights into the available types of tests for validity and reliability and how to incorporate these tests into their studies.
References


CHAPTER SEVEN

- 

CONCLUSION AND POLICY IMPLICATIONS
Summary of findings

This dissertation illuminates the mechanisms behind advancing good research practices and reporting conventions for preference research. It illustrates the cyclical and iterative process and the need to engage multiple stakeholders in the instrument development process. Furthermore, we conducted a subgroup analysis of preferences based on a conceptual model on the link between education, preferences, and health outcomes instead of a post-hoc analysis. This dissertation illustrates the importance of reporting study steps and decisions when conducting a stated-preference study: all steps from instrument development to data analysis were carefully detailed and data driven to ensure transparency of research methods.

We attribute several positive outcomes in this study to the application of good research practices. First, during the pretest, patients demonstrated that they understood the purpose of the study and thought it was important to measure preferences. Second, the rate of attribute dominance in the final survey was relatively low (16% vs. 40% in the pilot study) and participant burden seems to have been minimized. Thirds, the high level of choice stability and version equivalence point to the validity of the preference results. Other advantages of applying good research practices lie within the appropriateness of the mixed logit model that identified preference heterogeneity, and the face validity of the results.

The dissertation has several limitations. First, it was based on consensus based good research practices and is therefore limited by existing methodological gaps as outlined in Chapter 4. Current recommendations are limited by the lack of evidence on good research practices. Until more research is done comparing different research practices, recommendations on conducting stated-preference studies will remain opinion based.

Second, the approach of this dissertation was largely based on synthesis of published evidence.
While we involved patients with type 2 diabetes in the pretesting of the survey instrument, patient engagement was relatively limited. Our stakeholder committee was made up of a variety of clinical, methodological, and regulatory experts. Engagement was mainly focused around clinical and methodological questions, not around gaining insight into patients and/or community experiences. Other stated-preference studies have placed much more focused on patient engagement and emphasized the patient experience, rather than the published literature, in their study conduct. ¹

Third, while this work considered preference heterogeneity through the use of a mixed logit model and stratification analyses by education, more advanced methods to explore preference heterogeneity were not employed. While stratification models are useful to detect differences in preferences based on observed characteristics they are limited in the number of subgroups that can be evaluated due to sample size issues. Segmentation techniques using finite mixture models²,³ might present more advanced statistical models to detect subgroup heterogeneity. However, the advantage of stratification models is that they are familiar to a wider audience and might therefore be more easily disseminated. In addition, if preference differences are based on an observed characteristic this knowledge can inform clinical practice and/or benefit risk assessment.

A fourth limitation was that our sample was more highly educated than the national population between age 45-64.⁴ In addition our participants were more likely to take medication than a 2011 national sample (93% v. 81% respectively)⁵ and were living with diabetes longer (12.6 years v. 11.4 years).⁶ This reduces representativeness of our sample and might affect the generalizability of our preference estimate results.
**Policy implications**

Both the Prescription Drug User Fee Act (PDUFA) and the 21st Century Cures Act call for continued efforts by FDA to incorporate the patient experience in drug development. To reduce regulators’ burden, the Paper Work Reduction Act requirements have been waived for studies conducted by FDA to receive patient input (21st Century Cures). In addition, FDA has been tasked with providing guidance on how patient experience data might be collected and incorporated into benefit-risk assessment.

The Center for Drug Evaluation and Research (CDER) and the Center for Devices and Radiological Health (CDRH) are collaborating on exploring strategies to incorporate patient perspectives in regulatory decision making. CDER is working on integrating qualitative data from a list of therapeutic areas into existing evidence-based review processes through the patient focused drug development program. This initiative consists of public workshops with disease communities to gather information about the experience of living with a particular condition.

To properly account for patient preferences, investigators must have reliable and accurate methods, tools, and approaches. To that extend, FDA is now working on developing an approach to bridge from patient-focused drug development meetings to systematic tools to collect meaningful patient input that can be incorporated into regulatory review. Stated-preference studies have been used before to present information on the patient experience and treatment preferences at public meetings at FDA\(^7\) and could continue to be a valuable tool.

FDA is committed to improving review transparency, including how they conduct their benefit-risk assessments and how they arrive at approval decisions.\(^8\) If patient preference information is to be included in patient-centered benefit risk assessment, these studies need to meet the same level of transparency as clinical evidence studies.
In addition, CDRH now requires that patient-preference data satisfies standards of valid scientific evidence. It has released guidance on voluntary submission of patient preference information. In this guidance they specifically mention the use of stated-preference methods to collect patient-preference information. This strategy could facilitate integrating patient perspectives directly with clinical data on benefits and harms.

The guidance disseminated by CDRH recognizes that these various approaches exist and therefore does not make recommendations regarding specific methods that should be employed to measure patient preference information. Good research practices can reasonably include a variety of processes through all stages of a stated-preference study. There may be study-specific factors that determine good research practices, such as the homogeneity of the population under study, study resources, and funder requirements. For example, while conducting qualitative work during instrument development constitutes good research practices the precise method can vary from interviews or focus groups, to informal consultation with patients and other stakeholders.

There are always trade-offs between efficiency, comprehensiveness, and resources required to conduct a study. While investigator discretion is required based on the study objective and resources available, FDA guidance can be highly influential in determining minimum requirements for studies used in benefit-risk assessment. To that extend the CDRH guidance includes 11 qualities that should be considered when conducting a patient preference study.

As quantitative patient preference information is intended to represent the patients’ voice in benefit-risk assessment, CDRH first study quality, patient centeredness, might be the most important. While FDA also recognizes relevance as a desirable study quality, the guidance focuses on clinical relevance, not relevance to the patient experience. As a recent public meeting at FDA around clinical endpoints in diabetes indicates, endpoints that are recognized as
clinically relevant might not be patient-relevant. A balance between these two interests needs to be reached.

Many of the qualities FDA describe on ensuring comprehension by study participants so the preference study can be successfully. Study comprehension might be aided by focusing on two other study qualities: the need to minimize cognitive bias, and the need for proper communication of benefit, harm, risk and uncertainty. To meet these three qualities, cognitive pretest interviews should be employed.

FDA urges researcher to follow proper study conduct and established good research practices. These qualities are not clear-cut as gaps regarding good-research practices exist. In addition, while FDA asks to ensure that preference results meet logical soundness, qualitative work has shown that preference results that seem to violate rationality on the surface are logical from the patient’s perspective. Therefore, ensuring logical soundness might not be as straightforward as expected.

Representativeness and generalizability of results is a quality that is advocated by FDA and is generally considered a positive, if not essential, aspect of high quality studies. Unfortunately, representativeness of the sample might come at a high financial and time cost. Considering that previously the patient’s voice was represented solely through anecdotes of select patient representatives, almost any quantitative sampling strategy presents an improvement. If much emphasis gets placed on ensuring generalizability of results, financial resources associated with conducting a stated-preference study might be prohibitive to patient-advocacy groups interested in conducting patient preference studies for consideration in benefit-risk assessment. This would severely limit the opportunity of patient advocacy groups to make their patients’ voices heard.
Ensuring *robustness of the statistical analyses* is important to evaluate preference results. The uncertainty in the results might be reduced by incorporating techniques that *capture heterogeneity of patients’ preferences*. While not always emphasized in traditional research, results should be presented and explained so that patients and others without statistical expertise can interpret the data. For that reason, complex statistical approaches (such as latent class analysis) could be supplemented by less complex approaches (such as stratification) to ensure that a lay audience can understand the results.

In future guidance, FDA will need to consider which study qualities to prioritize. Preference studies that do not focus on representing the patient experience will ultimately fail at incorporating the patients’ voice into benefit risk assessments. By setting standards and providing further guidance, FDA can ensure that the collection of patient preference information is not just a token activity but will truly advance the incorporation of the patient experience into regulatory decision making.

**Further considerations**

While the place of patient preferences in healthcare decision making is being recognized, there is still little systematic or evidence-based application of patient preferences in healthcare decision making. To ensure the use of preference information in healthcare decision making, quality of preference results needs to be able to be evaluated. Standards to evaluate the quality of a stated-preference study, through tests for data quality and through establishing standard procedures or reporting templates, can ensure the transparency, reproducibility, and validity of stated-preference studies.

Future studies need to consider what these desirable properties of stated-preference studies are and how these properties compare to regular study and survey design. Given the patient-centered nature of preference research, research standards might focus around patient engagement and
relevance and less around ensuring theoretical validity. Standards that would only focus on theoretical validity without representing patients’ voices would not be appropriate for incorporating the patient experience into healthcare decision making.

In this dissertation, we have applied good research practices in stated-preference methods by measuring treatment preferences of people with type 2 diabetes. The applications of our approach will help other researchers conduct high quality stated-preference research. Ultimately high quality patient preference information can be used in the benefit-risk assessment of medications to ensure that the patient experience is better reflected in the regulatory process.
References


CURRICULUM VITAE

Ellen Margreet Janssen, BA
624 N. Broadway, Rm. 690, Baltimore, MD 21205
email: ejanssen@jhu.edu tel: +1 571-332-8538

Date of birth: January 23, 1990
Place of birth: Cali, Colombia
Nationality: Netherlands

EDUCATION AND TRAINING

Ph.D.  
Johns Hopkins Bloomberg School of Public Health  
Baltimore, MD, September 2012 - May 2017  
Health Economics and Policy

B.A.  
University of Virginia  
Charlottesville, VA, September 2008 - May 2012  
Major: Economics awarded with Distinction, Echols Scholar, Deans List (8 semesters)

POSITIONS AND HONORS

HONORS
2016  
Best Student Podium Presentation, 19th Annual European Congress, International Society for Pharmacoeconomics and Outcomes Research.

2016  
Certificate of Award for Student Abstract Presentation, Mid-Year Symposium, International Society for Pharmacoepidemiology.

2015  
The Lee Lusted Student Prize in Patient and Stakeholder Preferences and Engagement, 37th Annual North American Meeting, Society of Medical Decision Making.

2014 - current  
Centers of Excellence in Regulatory Science and Innovation scholar, Johns Hopkins School of Public Health – CERSI.

2012 - 2015  
Arthur Newsholme scholarship, Johns Hopkins School of Public Health.

POSITIONS AND EMPLOYMENT

2012 - current  
Research associate, Johns Hopkins School of Public Health

2014  
Predictive modeling intern, CNA insurance

PROFESSIONAL ACTIVITIES

Professional memberships
2014 - current  
Member, International Society for Pharmacoeconomics and Outcomes Research

2012 - current  
Member of Phi Beta Kappa Honor Society

Society working groups
2015 - 2016  
International Society for Pharmacoeconomics and Outcomes Research, (Patient-Focused) Benefit-Risk Analysis Working Group

Conference Peer Review Activities
2016  
International Society of Pharmacoepidemiology (ISPE)

2013  
Southeastern Health Economics Study Group
RESEARCH EXPERIENCE

FUNDING PARTICIPATION

Stated Preferences in Lung Cancer: Phase 2
Johns Hopkins School of Public Health, Baltimore, MD 02/2017-01/2019
Funder: LUNGevity Foundation
PI: John Bridges
Role: Research associate
This study will use qualitative and quantitative methods to learn about the patient experience and to measure patient preferences for the treatment of lung cancer.

Upper Limb Prosthetic User Needs and Preferences Study
Johns Hopkins School of Public Health, Baltimore, MD 04/2016 – current
Funder: Johns Hopkins – FDA Center for Excellence in Regulatory Science and Innovation
PI: John Bridges
Role: Research associate
To gather qualitative and quantitative patient preference information to better define user needs and preferences for novel upper limb prosthetic devices. To incorporate patient preference information into regulatory benefit-risk assessment.

Patient and Provider Views on Clinical Endpoints: A Qualitative Preference Study of Minimally Invasive Glaucoma Surgery (MIGS) Devices
Johns Hopkins School of Public Health, Baltimore, MD 01/2016 – current
Funder: FDA Intramural
PI: Malvina Eydelman, Michelle Tarver
Role: Co-investigator
To gather qualitative patient preference information to better define a set of outcomes (or endpoints) best reflecting patient-valued aspects of MIGS devices.

Using life expectancy to inform individualized cancer screening
Johns Hopkins School of Medicine, Baltimore, MD
Funder:
PI: Nancy Schoenberg 01/2016 – current
Role: Research associate
This study will explore older adults’ perspective regarding how clinicians and patients should communicate and incorporate life expectancy in cancer screening. In addition it will determine and quantify older adults’ preferences for how clinicians and patients should communicate and incorporate life expectancy in cancer screening among a nationally representative populations.

Stated Preferences in Lung Cancer: a Pilot Study
Johns Hopkins School of Public Health, Baltimore, MD 04/2015 – 04/2016
Funder: LUNGevity
PI: John Bridges
Role: Research associate
The study will apply principles of community-centered research to develop and implement a national survey of people diagnosed with lung cancer and will apply stated-preference methods.

Advancing stated-preference methods for measuring the preferences of patients with type 2 diabetes
Johns Hopkins School of Public Health, Baltimore, MD  
Funder: Patient Centered Outcomes Institute  
PI: John Bridges  
09/2014 – 11/2016  
Role: Research associate  
To advance and disseminate methods for patient and stakeholder engagement in patient-centered outcomes research. Applying community-based participatory research, we address methodological gaps associated with stated-preferences methods in a project that it focused at understanding the priorities and preferences of patients with type 2 diabetes.

OTHER RESEARCH EXPERIENCE

Comparing the costs of an in-person and a phone-based behavioral weight loss intervention in people with Serious Mental Illness  
Johns Hopkins School of Medicine, Baltimore, MD  
PI: Gail Daumit  
09/2016 – current  
Role: Research associate

Cost estimation of Project Achieve weight loss trial for people with serious mental illness  
Johns Hopkins School of Medicine, Baltimore, MD  
PI: Gail Daumit  
09/2013 – 12/2016  
Role: Research associate

Review of the evidence - prevalence of medical conditions among the population with serious mental illness  
Johns Hopkins School of Medicine, Baltimore, MD  
PI: Gail Daumit  
12/2012 – 09/2013  
Role: Research associate

Occurrence of mental health related claims in worker’s compensation claims  
CNA insurance, Chicago IL  
06/2014 – 08/2014  
Supervisor: Wayne Wendling  
Role: Predictive Modeling Intern.
TEACHING EXPERIENCE

Guest lecturer – Johns Hopkins School of Public Health, Baltimore, MD
• 04/2016  Economic Evaluation 3 (40 students)
  • Lecture: Good research practices for the design of discrete choice experiments
  • Lecture: Stated-Preference Methods: Data Analysis

Teaching Assistant – Johns Hopkins School of Public Health, Baltimore, MD
• 10/2015-12/2015  Health Economics 1 (Instructor: Dr. Douglas Hough)
• 09/2014-10/2014, 09/2015-12/2015, 09/2016-12/2016  Mathematical Microeconomics (Instructor: Dr. John Bridges)
• 09/2013-10/2013, 09-2014-10/2014, 09/2015-10/2015  Introduction to Microeconomic Theory (Instructor: Dr. Alan Sorkin)
• 03/2014-05/2014  Introduction to Economic Evaluation (Instructor: Dr. Chuck Shih)
• 03/2015-05/2015, 03/2016-05/2016  Introduction to Economic Evaluation (Instructor: Dr. Eric Roberts)
• 01/2015-03/2015  Introduction to Health Economics (Instructor: Dr. Douglas Hough)
• 10/2014-12/2014  Assessing Health Status and Patient Outcomes (Instructor: Dr. Albert Wu)
• 01/2014-03/2014  Economic Evaluation II (Instructor: Dr. Dagna Constenla)
• 10/2013-12/2013  Economic Evaluation I (Instructor: Dr. Gregory de Lissovoy)
• 09/2013-05/2014  Public Health Economics Seminar (Instructor: Dr. John Bridges)
PUBLICATIONS

PUBLISHED OR ACCEPTED


UNDER REVIEW


WORKING PAPERS (COMPLETED)


PUBLISHED ABSTRACTS


ORAL AND POSTER PRESENTATIONS


BIOGRAPHY

Ellen M. Janssen holds a bachelor degree with distinction in Economics from the University of Virginia and is currently preparing a doctoral thesis at the Johns Hopkins Bloomberg School of Public Health. Ms. Janssen’s research is focused on measuring patient preference information for use in regulatory science across a wide spectrum of diseases. Her current work examines good research practices for the development, implementation, analysis, and evaluation of stated-preference methods to ensure that patient preference studies are transparent and meet high quality standards. Ms. Janssen is passionate about engaging patients as research partners in all levels of healthcare decision-making and has worked with a variety of patient and community groups in her research.

Key words