2008

Development and validation of a diabetes-specific health state classification system and valuation function based on the multi-attribute theory

Murali Sundaram
West Virginia University

Development and Validation of a Diabetes-specific Health State Classification System and Valuation Function Based on the Multi-attribute Theory

Murali Sundaram

Dissertation submitted to the
School of Pharmacy
at West Virginia University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in
the Pharmaceutical Sciences

submitted to

Michael J. Smith Ph.D., Committee Chair
Gerald Hobbs, Ph.D., Committee Member
S. Suresh Madhavan, M.B.A., Ph.D., Committee Member
Lesley-Ann Miller, Ph.D., Committee Member
Dennis A. Revicki, Ph.D., Committee Member

Department of Pharmaceutical Systems and Policy

Morgantown, West Virginia
2008

Keywords: Diabetes, Preference-Based Measure of Health, Utility, Health Utility, Quality-Adjusted Life Year, Multi-Attribute Utility Theory

ABSTRACT

Development and Validation of a Diabetes-specific Health State Classification System and Valuation Function Based on the Multi-attribute Theory

Murali Sundaram

Preference-Based Measures of Health (PBMH) provide ‘preference’ or ‘utility’ weights that enable the calculation of quality-adjusted life years for the economic evaluations of interventions. The Diabetes Utility Index (DUI) was developed as a two-page, self-administered diabetes-specific PBMH that can replace expensive time-consuming interviews with patients to estimate their health state utilities. Inputs from theory, an existing diabetes-specific measure of quality of life, and statistical analyses were submitted to a clinical expert panel. After three rounds of pilot surveys (n1=52, n2=65, n3=111) at primary care clinics in Morgantown, WV, five attributes and severity categories for each attribute were finalized on the basis of the results of Rasch Analysis and consultations with the panel. The final attributes were: ‘Physical Ability & Energy’, ‘Relationships’, ‘Mood & Feelings’, ‘Enjoyment of Diet’, and ‘Satisfaction with Management of diabetes’. The next step involved obtaining preferences for health states based on combinations of DUI attributes and severity levels from 100 individuals with diabetes, recruited from primary care and community settings in and around Morgantown, WV, in hour-long one-on-one interviews. These health states were anchor states, single-attribute level states including corner states, and marker states. The interviews provided data to calculate a Multi-Attribute Utility Function (MAUF) that calculates utilities for any of the 768 health states that can be defined by the DUI, on a scale where 1.00=Perfect Health and 0.00=the all worse ‘Pits’ state, from respondents’ answers to its five questions. In addition to an overall index score, attribute-level preference scores were also calculable by the function. Finally, a validation survey was conducted in collaboration with the West Virginia University (WVU) Diabetes Institute. For concurrent and construct validation purposes, the DUI was mailed to individuals with diabetes along with generic PBMH like the EuroQol EQ-5D, the SF-6D and other patient-reported outcomes measures like the Diabetes Symptoms Checklist-Revised, the Short Form 12 (SF-12) and the Well-Being Questionnaire (W-BQ12), and their surveys responses (n=396) were merged with a clinical database consisting of ICD-9 diagnosis codes. The DUI utilities were found to be largely free of socio-demographic effects and its scores were well distributed between 0.00 and 1.00. The DUI moderately correlated with generic PBMH and distinguished between severity groups based on diabetes symptoms and complications. The scoring function of the DUI calculated utilities favorably compared against cardinal Standard Gamble utilities obtained directly from patients for three DUI health states. These results show evidence of the feasibility and validity of the DUI. Further research is suggested to demonstrate the generalizability of these findings, to study the responsiveness of the DUI, and to examine the clinical meaningfulness of the DUI change scores.

DEDICATION

This research is dedicated to

Amma (Thangamani Sundaram)

In recognition of my mother who has provided me my earliest educational experiences

and

Appa (Prof. S. Sundaram)

In recognition of my father and his support to my life and education

INSPIRATION

கற்க கசடறக் கற்பவை கற்றபின்
நிற்க அதற்குத் தக.

Karka kacatarak karpavai karrapin
Nirka atarkut taka.

Whatever is to be learnt should be learnt flawlessly
so that the learning imbibed shapes one’s conduct.

From the ‘Thirukkural’,
a classical collection of couplets expounding various aspects of life, written in Tamizh,
my mother tongue and one of the few surviving classical languages in the world.

ACKNOWLEDGMENTS

Many of my friends and family members have been telling me that my successful defense of the Ph.D. degree will lend legitimacy to the emails IDs I've created with unimaginative usernames like '*doctormurali*', ever since my first account with Hotmail in the late nineties. The degree feels like a very natural suffix to my name since I've planned to do a Ph.D. since I was very young, thanks to individuals in my family who pursued similar endeavors.

In acknowledging the support provided by my Ph.D. Committee Chair, Dr. Michael Smith, I'd like to borrow from a fellow graduate student colleague's admission - he has been patient^n, something that provided me with the cushion needed to devour the many hurdles that came my way towards the completion of this dissertation. I am thankful to Dr. Madhavan for the opportunity to be a part of this dynamic graduate program, for his stewardship of the program, and his support and encouragement through the years I have spent here. With Dr. Lesley-Ann Miller regularly urging me to aim higher with this project, I found it feasible to execute components of the project that I am now glad that I did. It has been an honor to have Dr. Dennis Revicki on the committee and I look forward to co-authoring papers resulting from this work along with him. Special thanks go out to Mike Linacre and Bill Furlong, who I'd like to refer to as e-committee members even though they did not formally serve on my committee, who rescued me on many an occasion with their lightning-fast answers to the questions that I would come up with at an even greater pace.

The members of the clinical expert panel, the individuals who provided selfless assistance in recruiting people with diabetes, and those who helped acquire the data I needed all played important supportive roles during different stages of the project. Several names come to my mind, and I thought it deserving to list each of them at the end of this note as a humble gesture of thanks. Credit is also due to Eli Lilly & Company for their financial support to this project. Of course, the people who participated as research subjects in the various project components deserve special mention and recognition. Some of them declined to accept the participation rewards, preferring to simply take part in the larger interest of promoting diabetes-related research.

Being an academic for most of his life, my father has been keenly observing my progress in graduate school and will be very glad to see me at this stage in my life. My mother has played many roles with selfless zeal from providing me my earliest educational experiences, to enduring my daily tantrums about finishing up breakfast that went on right up to the day I prepared to leave India to attend Graduate school in the U.S. I would also like to recognize my dear wife of six months, Preeti Krishnan, for giving up her career to be with her ABD husband, as she adjusted to a new life with me in a new country, far away from almost everybody she has known.

A few days ago, I stumbled upon a copy of a 'Statement of Purpose' that I had submitted to the graduate program at WVU as part of my application material in the year 2002. Back then, I had expressed a desire for my graduate education to 'provide me with an insight into the intangibles that make up Quality of Life'. As I leave WVU in 2008, I hope that this training would foster my abilities in the development of tools that will assist in the assessment of strategies aimed at enhancing the Quality of Life of people.

Thank You!

Committee Members
Gerry Hobbs
Suresh Madhavan
Lesley-Ann Miller
Dennis Revicki
Mike Smith

Unofficial Committee Members
Dr. Bill Furlong
Dr. Mike Linacre

Clinical Expert Panel
Betsy Elswick
Andrea Hasley
Risa Hayes
April Lombardo
Charlotte Nath
Liz Quintana
Tara Whetsel

Data Support
John Brazier
J.L. Brown
Michele Garlick
Preeti Krishnan
Tina Podolinski
The U.S. Post Office at the Health Sciences Center

Logistics Support
Howard Franks
Mickey Howell
Deborah Kuis
Adrienne Tucker

Navigating the Red Tape
Lilo Ast
Debbie Anderson
M.J. Buckland

Being there
Preeti Krishnan
Mark Moffett
Ginger Scott
Cindy Tworek
My graduate student colleagues
Friends & Family

Recruitment Support
James Arbogast
Rose Bakos
Emily Devault
Betsy Elswick
Kathryn Godley
Tim Jackson
Laura Jones
Kevin Halbritter
Andrea Hasley
April Lombardo
Liz Quintana
Beth Semmens
Deborah Shriver
Tara Whetsel
Team Health Right
Team WVU Diabetes Ed. Ctr.

PRO Instrument Licensors
Clare Bradley
Mapi Research Trust
The Michigan Diabetes Research and Training Center
QualityMetric Health Outcomes Solutions Inc.
The EuroQol Group

Financial Support
Eli Lilly & Company

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

1.1 The Economic, Clinical and Humanistic Burden of Diabetes

The National Institute of Diabetes and Digestive and Kidney Diseases (2005) has estimated that about 21 million people in the U.S., seven percent of the population, had diabetes in 2005. Diabetes is a chronic disease associated with complications and other co-morbid conditions that add to the burden of diabetes. Some of the more significant complications resulting from the inefficient management of diabetes are eye disease (blindness, retinopathy), kidney disease (nephropathy, end stage renal disease), and nervous system damage (neuropathy, foot ulcers), among others. The Centers for Disease Control and Prevention (CDC) estimates that the risks of heart disease and stroke are both two to four times higher in persons with diabetes, and an estimated 73% of persons with diabetes have hypertension (CDC, 2003).

Diabetes is also the leading cause of adult blindness and end stage renal disease accounting for 44% of new cases of kidney failure (CDC, 2003). Approximately 60-70% of persons with diabetes have neuropathies; severe forms of diabetic nerve disease are a major contributing cause of lower-extremity amputations (CDC, 2003). Other conditions associated with diabetes include acute life-threatening events such as diabetic ketoacidosis and hyperosmolar coma, both short-term complications of the disease. People with diabetes are also more susceptible to many other illnesses and once they acquire these illnesses, their prognoses worsen (CDC, 2003). It is not surprising that diabetes is the sixth leading cause of death by disease in the U.S. (Kochanek & Smith, 2004).

Traditional measures of clinical outcomes like morbidity and mortality continue to be measured and reported, but there is increasing appreciation for the patient's perspective on health, disease, and medical treatments – referred to as 'patient-reported outcomes' (PROs) or 'humanistic' outcomes. While healthcare professionals examine the impact of diseases using clinical parameters, patients can provide their own assessments of their health-related quality of life (HRQoL), treatment satisfaction and adherence, and symptom burden, and others. Among these, HRQoL (often considered synonymous with quality of life or QoL) is one of the widely assessed patient-reported outcomes in the evaluation of healthcare interventions. There is an ongoing debate on the need to differentiate between the constructs representing QoL and HRQoL that considers the former to be a broader term encompassing the latter (Bradley, 2001). For the sake of simplicity, this document will use the term HRQoL in reference to both. Patient-reported outcomes, including HRQoL, are assessed using instruments that measure their respective constructs in a descriptive manner. These instruments can be commonly referred to as profile-based measures; some of these measures provide a single score, while others provide a profile of scores.

Diseases like diabetes can affect people in different ways depending on the aspects of life that are compromised. The management of diabetes imposes considerable demands on patients and their families as well. Apart from the emotional and social burdens this may cause, they face the acute distresses of hypoglycemia or hyperglycemia and chronic distress of diabetes-related complications (Rubin, 2000). Hence, diabetes is said to impact the physical, psychological and social functioning of individuals. In order to measure this impact, it is necessary to take into account several factors: patient's perceived physical distress due to diabetes-specific symptoms, loss of physical function and independence, as well as the perceived emotional distress due to symptoms, self-care, and the interference with common activities and social situations (Polonsky, 2000). While this list is neither complete nor standard, it is clear that such measurements are not straight-forward. Hence profile-based instruments are used that provide a descriptive assessment of the impact of disease on health in general or in specific aspects of life.

The economic expense of diabetes was estimated at $132 billion in 2002, with direct medical expenditures totaling $92 billion and $40 billion due to the indirect costs of lost productivity resulting from lost workdays, restricted activity days, permanent disability, and mortality (Hogan, Dall, & Nikolov, 2003). Recognizing that the population with diabetes tends to be older, on average, than the population without diabetes, the total health care costs for people with diabetes in the U.S. have been found to be between four to five times the costs for people without the condition (American Diabetes Association, 1998; Rubin, Altman, & Mendelson, 1994; Hogan et al., 2003). This forms a tremendous national economic burden.

The dynamics of the U.S. health care system have influenced how diabetes treatments and interventions are evaluated. The scarcity of healthcare resources and the emergence of managed care as a dominant influence on health care delivery has lead to an increasing emphasis on assessment of the costs associated with new and existing treatments relative to their effectiveness. Such economic assessments, in the form of cost-effectiveness analysis (CEA) and cost-utility analysis (CUA), may be sought as an input to aid in decision-making or simply may be required to make submissions due to legislative mandate.

1.2 Statement of the Problem

One of the key inputs for the CEA and CUA models are utility values for health states of interest in the treatment or intervention being evaluated. A utility-gathering exercise is typically a separate study in which, people (a sample of similarly diseased subjects or the general population) are given comprehensive descriptions of specific health states under consideration. Respondents then answer questions, which involve the use techniques such as the Rating Scale, the Standard Gamble, or the Time Trade-Off, that will enable the estimation of a numeric value that represents their utility for that health state. The direct evaluation of health state utilities in this manner is a challenging task. While these techniques themselves help researchers in enabling participants to state their health-state preferences, they can be time-consuming and expensive. People may find it cognitively difficult to understand health state scenarios, to state their preferences for health states, and to undertake the gambles and trade-offs with health.

An alternative to this approach is to obtain utilities indirectly, by using preference-based measures of health (PBMH). All these measures contain health-state classification systems, composed of attributes of health that may be influenced, and severity levels for those attributes, which can together describe diverse health states. Respondents' answers to the classification are fed into a pre-coded valuation function that identifies the pattern of responses and provides a numerical value of utility. Examples of PBMH are the EuroQoL EQ-5D, the Health Utilities Index (HUI), the Quality of Well Being (QWB), and the Short Form 6D (SF-6D). These measures differ in the types and number of attributes included and the severity levels allowed to be assessed.

Most of the existing PBMH employ generic classification systems and, like generic measures of HRQoL, may be inappropriate or insensitive for specific disease conditions. There are several disease-specific profile-based measures available that provide descriptive assessments of the impact of disease and its treatments. Some of them provide more than one score, which when used in economic evaluations would require multiple cost-effectiveness ratios. However, these are largely profile-based measures that do not explicitly incorporate preferences for different possible health states into their scoring algorithms. Hence, such instruments are not suitable to be used in economic evaluations of healthcare interventions in their current formats.

One solution to this problem would be to link the disease-specificity of existing profile-based measures with the convenience and appropriateness of existing generic PBMH. This process would involve concurrently administering the two measures and using statistical techniques to predict utility weights from profile-based scores. The resultant conversion algorithms can be used to obtain utility values for health states in other studies that failed to incorporate PBMH but used the same profile-based measure. This is regarded as only the second-best approach as the utilities obtained in this manner are limited to the range of the utilities calculated from the original generic PBMH.

An alternative solution would be to develop disease-specific PBMH. Such measures would have the dual advantages of disease-specificity and ease in obtaining utilities, as opposed to directly evaluating health state utilities. Developing such a measure involves formulating a disease-specific health classification system along with a valuation function that will enable the calculation of utilities for the all health states that can be described by the classification system. While such systems can be based on input from comprehensive primary research for content, they can also benefit from established disease-specific profile based measures. In their current format, many existing profile based measures will give rise to classification systems that are too complex from which to meaningfully elicit utilities. However, it is feasible to suitably modify them for the purpose of developing a concise disease-specific classification system and then develop a valuation function to calculate utilities for the health states it describes. Based on existing knowledge of published research, there currently is no diabetes-specific PBMH providing utilities that can be directly used in economic evaluations.

1.3 The Proposed Study

1.3.1 Objectives of the Study

The Audit of Diabetes Dependent Quality of Life (ADDQoL) is a diabetes-specific measure of that provides a descriptive assessment of quality of life in patients with diabetes. Independent reviews have recommended the instrument for QoL assessments of both Type 1 and Type 2 patients with diabetes, and have described it as a brief and recent instrument generated with patient input, with good reliability, internal and external construct validity (Garratt, Schmidt, & Fitzpatrick, 2002; Wildes, Greisinger, & O'Malley, 2003). The instrument was not designed to allow respondents to indicate their individual preferences for different health states. Consequentially, the score on the ADDQoL is not preference-based and cannot be used in economic evaluations of healthcare interventions in diabetes.

There are two key components in constructing PBMH: 1) Designing a health-state classification system (or health state descriptive system) that is composed of attributes or domains that influence health; and 2) Estimating a valuation function that provides numerical values of utility for all health states that can be described by the classification system (Feeny, 2002a). Given the satisfactory reports of validity and reliability of the instrument and encouraging reviews, the content of the ADDQoL can be used in place of extensive primary research on diabetes-specific issues that is normally necessary as a part of the development of a new PBMH, and supplemented with input from a clinical expert panel. Like any other new instrument, a new PBMH will need to be tested for the validity of its measurements. It is also important to know what causes the variation in utilities – whether it is due to real differences between people or if it is due to construct-irrelevant variation (Lenert & Kaplan, 2000).

The purpose of the proposed study is to develop and validate a brief diabetes-specific PBMH. The specific objectives are:

1. To develop a diabetes health state classification system;
2. To derive a valuation function for the diabetes health state classification system; and
3. To assess the validity and predictors of the utilities derived from the diabetes health state classification system.

1.3.2 Research Plan & Conceptual Framework of the Study

The conceptual framework of the proposed study is presented in Figure 1.

Objective One:

A combination of theory, subjective and statistical approaches was employed to identify attributes and severity levels for a diabetes-specific health state classification system. The statistical approaches used were guided by the principles of both classical and modern theories of measurement. The results from the statistical analyses were discussed with a clinical expert panel to ascertain adequate coverage of issues important to patients with diabetes, and to make suitable modifications using an established consensus-gathering technique. Finally, the attributes and levels of severity chosen were adequately worded to create a diabetes-specific classification system, christened the Diabetes Utility Index (DUI), and were pilot tested on a patient samples.

Objective Two:

Phase One: The Multi-Attribute Utility Theory (MAUT) was used as the framework to develop a valuation function for the DUI. The MAUT guided the selection of health states from the classification system developed in Objective One. Health state descriptions were constructed from the attributes and severity levels of the DUI and the utility elicitation tasks were designed in accordance with existing measurement guidelines. The scenarios and tasks were pilot tested on a general population sample.

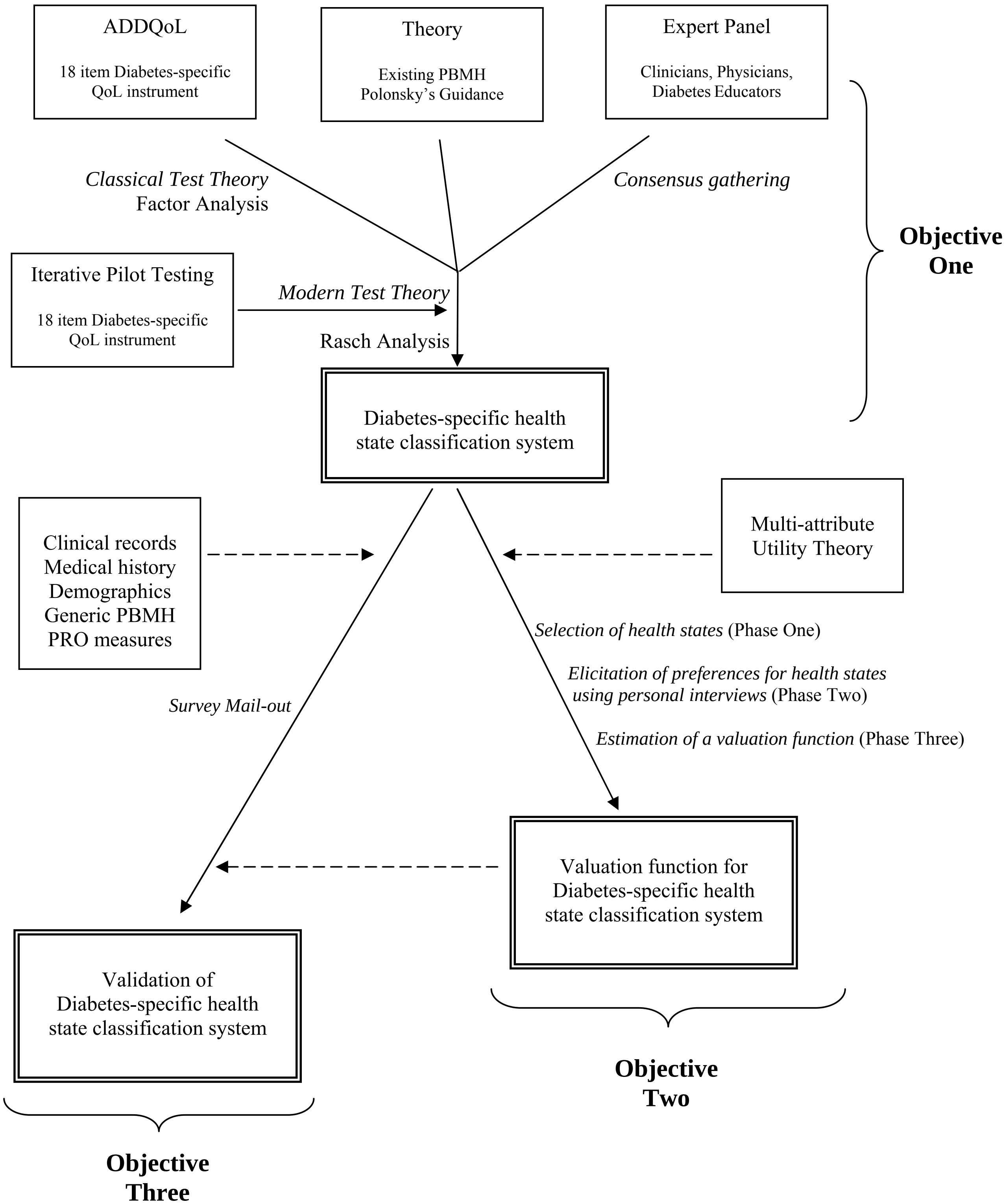
Phase Two: Patients with diabetes were administered the utility elicitation tasks designed in Phase One using one-on-one interviews.

Phase Three: The data collected in Phase Two were used to calculate the necessary metrics required as per the MAUT framework. This led to the formulation of a valuation function, also called as a multi-attribute utility function (MAUF), which would yield utility values for all health states described by the DUI.

Objective Three:

Objective Three was accomplished alongside Phases Two and Three of Objective Two, and involved administering the DUI to a separate outpatient sample of patients with diabetes in the form of a mail survey. The utilities of patients for their current health states were obtained using the MAUF, and were merged with a clinical and demographic data to conduct construct validation analyses and to understand the predictors of patients' utilities. In addition, the concurrent validity of the DUI was analyzed by comparing its scores with those obtained from other generic PBMH.

Figure 1: Conceptual Framework of the Study



1.3.2 Significance of the Study

There is increasing interest in the development of disease-specific PBMH for use in economic evaluations. This study will be the first to report the development of a diabetes health state classification system using inputs from an existing diabetes-specific measure of quality of life in combination with expert panel inquiry and statistical analyses. In addition to the classical approach to reviewing item content, the study incorporated techniques based on the Modern Test Theory as well as subjective approaches based on expert review. The valuation function for the classification system was based on the MAUT, an established theory to model utilities for health states defined by multi-attribute health status classification systems. Finally, this study reported on the feasibility of the development of a PBMH specifically for use in economic evaluations of diabetes interventions.

CHAPTER TWO: Literature Review

2.1 Health-Related Quality of Life as a Humanistic Outcome

Humanistic outcomes are increasingly being used along with traditional outcome measures as end-points in clinical trials of healthcare interventions. While one of the traditional outcomes in the number of years or 'quantity of life', humanistic outcomes help understand the 'quality' aspect of life of patients in terms of how it is influenced by disease and its treatments. Patient-reported humanistic outcomes are of several types, differing in terms of the range of issues to which they pertain; one such important outcome is health-related quality of life (HRQoL).

Hareendran (2005) summarizes the value that patient reported outcomes like HRQoL can add to the evaluation of new treatments: They are key measures of treatment outcomes when there are no objective markers of symptoms and no objective markers of the impact of symptoms. They also complement traditional endpoints to evaluate the significance of a treatment effect from a patient's perspective. Information on HRQoL outcomes can facilitate patients' involvement in treatment decision-making. Finally, HRQoL outcomes can provide guidance for health care decision-making by enabling a better understanding of the burden of illnesses and in making healthcare allocation decisions.

Descriptive measures of health, called by a variety of names including profile-based measures, health status measures, and HRQoL measures, are used to assess the impact of diseases and treatments on health. Profile-based measures of HRQoL are descriptive in that they help identify impairments in health in general or in specific domains of life that affect health. However, the scores on these instruments do not incorporate preferences of users of healthcare (Brazier, Roberts, & Deverill, 2002). Hence it is difficult to incorporate profile-based scores, of these instruments into cost-effectiveness analyses (Revicki, 1996). Additionally, when such measures generate a profile of scores rather than a single score, it can becomes difficult to interpret the multiple and potentially conflicting cost-effectiveness ratios (Revicki, 1996; Revicki, Leidy, Brennan-Diemer, Sorensen, & Togias, 1998a; Revicki, Leidy, Brennan-Diemer, Thompson, & Togias, 1998b). Another issue hindering the use of profile- based scores has to do with the lack of interpretability of changes on these scores in economic evaluations, within the context of relative cost of therapy (Revicki, 1996).

2.2 Utility as a Humanistic Outcome

In comparison to HRQoL, a Quality-Adjusted Life Year (QALY) is an outcome measure that takes into account both the quantity as well as the quality of the added life provided by a healthcare intervention; it is the arithmetic product of the added life expectancy and the quality of the remaining years (Drummond, O'Brien, Stoddart, & Torrance, 1997a). In other words, the added number of years of life afforded by an intervention is adjusted or weighted for the quality of those years as perceived by the patient. Quality-Adjusted Life Years constitute the denominator to calculate cost per QALY gained in cost-utility analysis (CUA) (Drummond et al., 1997a). In fact, one of the key features of a CUA is its use of the QALY concept. The use of QALYs allows the direct economic analysis of interventions that have more than one important outcome measure, and also facilitates the comparison of interventions that have diverse outcome measures (Drummond et al., 1997a). Cost-effectiveness analysis (CEA) is inappropriate in both these cases since the technique requires that the outcomes be measured in program-specific units. Hence CUA and CEA are similar on the cost side, but differ on the outcomes side. Additionally, CUA can be used only with final outcomes data (e.g. lives saved, disability-years averted) and not intermediate outcomes data because the latter cannot be converted into QALYs (Drummond et al., 1997a).

The adjustment factors or quality weights needed to calculate QALYs in economic evaluations are measured preferences for health states in an intervention, interchangeably called 'values', or 'preferences', or in economic terms, 'utilities' (Gold et al., 1996a), all of which are used under the umbrella term 'preferences' (Drummond et al., 1997a). In general terms, it is acceptable that the terms be used in place of each other, since there is more utility attached to preferable outcomes (Drummond et al., 1997a). However, in technical terms, there are differences arising due to the way preferences are measured – when measured under risk they are referred to as utilities; when measured under certainty they are referred to as values. Preferences are also classified as 'ordinal' or 'cardinal' – in the former, health states are simply rank ordered whereas the latter requires a number to be attached to it that represents the strength of the preference for that health state in comparison with others (Drummond et al., 1997a).

Utility values are gathered by asking people about their preference for health states, specifically those health states of interest occurring among patients in the course of the healthcare intervention being assessed. Utilities are required to meet some conditions: 1) they should meet individuals' preferences for health states and not merely provide a descriptive assessment of health as provided by profile-based measures (Brazier et al., 2002); 2) they must be measured on an interval scale in which equal intervals have equivalent interpretation (Gold, Siegel, Russell, & Weinstein, 1996b); 3) they must be anchored on measures of perfect health and death. Traditionally, a scale ranging from 1 (indicating perfect health) to 0 (indicating death) is used (Gold et al., 1996b) with allowance for negative values for health states perceived worse than death.

2.3 Methods to Obtain Preferences/ Utilities

There are two approaches to obtaining utility values for use in economic evaluations: referred to as direct and indirect utility elicitation techniques. In a direct utility-elicitation exercise, respondents are given comprehensive descriptions of specific health states under consideration, and techniques such as the rating scale, the standard gamble (SG), or the time trade-off (TTO) are used to obtain their preferences for being in those health states. Although guidelines are available to undertake each of these techniques, the actual exercises differ in terms of the health states presented and other information describes to the respondents. The indirect approach makes use of preference-based measures of health (PBMH). Health states do not have to be presented to respondents, because these PBMH contain unique standardized multi-attribute health-state descriptive systems. Utility values are obtained from respondents' answers to the health-state descriptive system by using a pre-determined scoring algorithm. Both of these approaches are discussed in detail in the following sections.

2.3.1 Direct Approaches

In the rating scale approach, respondents are asked to assign a single number to a health state on an anchored scale where the lowest number corresponds to the worst imaginable state (like death), and the highest number corresponds to the best imaginable state (usually perfect health). The rating scales have been given different names like 'visual analog scale' (VAS) and 'feeling thermometer', among others.

A standard gamble (SG) begins with presenting a written description of the health state under evaluation. The respondent is then asked to imagine a hypothetical situation in which a choice has to be made between two alternatives: to continue living in a health state of interest which is not perfect health (the certain outcome) or to take a gamble with two possible outcomes – perfect health or death (or another negative state, which could be worse than death), with probabilities 'p' and '1-p' respectively. The probabilities in the gamble are systematically altered until the respondent cannot choose between the described health state and the gamble, (i.e. the respondent is indifferent about the two alternatives). The probability at this point of indifference is considered the utility of the health state (Krabbe, Essink-Bot, & Bonsel, 1997). Or, the expected value of the gamble at this point is an estimate of the utility of the standard relative to perfect health or death (Gold et al., 1996a). Thus, it can be seen that the SG method differs from the rating scale in two important ways: the SG method incorporates the trade-off concept by introducing a choice between alternatives; also one alternative contains uncertainty and hence incorporates risk and probabilities.

The time trade-off (TTO) technique measures how much time a subject is willing to trade off to avoid a specific health outcome. A TTO begins with presenting a written description of the health state under evaluation. The respondent is then asked to imagine a hypothetical situation in which a choice has to be made between two alternatives: to continue living in the health state described (the poorer state) for a defined time, or to choose perfect health for a lesser amount of time. The time in the healthier (and more desirable) health state is varied (shortened) until the respondent is indifferent between the two alternatives, which becomes the indifference point. The TTO utility for the health state under consideration is equal the remaining number of years in perfect health (life expectancy) at the point of indifference divided by the length of remaining life with that health state (Gold et al., 1996a). It incorporates two easily understood aspects of health: morbidity and longevity. As compared to the rating scale, and like the SG, the TTO method incorporates the trade-off concept. However, since the outcomes presented to the respondent are understood to occur with certainty, the concepts of risk or probabilities are not associated with a TTO task.

2.3.2 Comparison of Direct Approaches

Rating scales can be used to obtain preference values, but which are not unanimously thought of as having interval scale properties. Rating scales only ask subjects to indicate the relative value of a health state under consideration compared to a worst and perfect health state, and thus has no utility-theoretic basis (Louviere, 1988). For use in economic evaluations, utility weights that represent individual's relative preferences for different health states are needed (like those obtained from other techniques like TTO or SG). This had lead researchers to doubt the validity of the preference values generated using rating scales as a measure of strength of preference (Bleichrodt & Johannesson, 1997; Schwartz, 1998; Robinson, Loomes, & Jones-Lee, 2001). Hence, rating scale values need to be converted to utility weights that incorporate some concept of tradeoffs—how much of one thing an individual is willing to give up in return for something else. The relationship between rating scale values and utility estimates from the TTO and SG have been found to be highly non-linear (O'Leary, Fairclough, Jankowski, & Weeks, 1995; Stiggelbout et al., 1996; Furlong et al., 1998; Lenert, 2000), while there are reports the relationship is linear (Mrus et al., 2003).

Measurement biases may also occur depending on the respondents' capabilities to perform the rating task. Respondents may space out health states presented on the rating scale irrespective of the severity of those states, suggesting that the preference values obtained are relative rather than absolute. In other words, the valuation of a health state is influenced by the presence of another state in the same rating task. Also, respondents may not use the ends of the rating scale, leading to end-of-scale bias that needs to be corrected for (Furlong, Feeny, Torrance, Barr, & Horsman, 1990; Torrance, Feeny, & Furlong, 2001). In order to overcome this, an end-of scale correction bias has been suggested (Furlong et al., 1998).

Rating scales are quick to complete, less expensive and are said to be the least burdensome among preference elicitation techniques. Thus, the technique lends itself to self-administration better than other techniques, and is easier to implement in a mail survey format (Torrance et al., 2001). They can also be used as a warm-up technique, to familiarize respondents with the descriptions of health states before moving on to another preference-elicitation technique, and to obtain ordinal preferences for health states (Torrance et al., 2001).

Because of their ease of use, rating scales are used as a proxy for other preference elicitation methods (O'Leary et al., 1995), and are best used in conjunction with other methods. Adopting such an approach, rating scales can be used to obtain ordinal preferences for those health states in a study that are not valued by the SG or TTO techniques. The values obtained in a rating scale have been converted to SG or TTO utility values using an appropriate conversion formula (Schackman et al., 2002; Mrus et al., 2003; Raat, Bonsel, Hoogeveen, & Essink-Bot, 2004). Noting that a standard conversion equation that could be applied universally has evaded researchers thus far, Torrance (2001) recommends that such equations be developed specific to each study so that the conversion is appropriate to the study context. However, there are reports that relationships between SG and rating scale scores are not stable (Robinson et al., 2001).

The main advantage of the SG method is that it provides utility estimates using a method that is consistent with von Neumann and Morgenstern's (vN-M) expected utility theory (von Neumann, Morgenstern, & Rubinstein, 2004). Standard Gamble is the one utility elicitation method that has a well-developed basis in utility theory, and is hence considered as a gold-standard against which other methods are compared (Torrance, 1976b). Feeny (2002b) has summarized the reasons for the widespread use of the vN-M theory as the foundation of health utility assessments: one, it makes only simple assumptions to provide a scale with interval-scale properties; and two, vN-M utilities deal with risk which naturally exists in the context of healthcare decision-making.

The virtues of the SG method do come at a cost. It is better administered in an interview format with well-trained interviewers and effective props, which makes it more expensive than the rating technique. The technique, with an iterative process of changing probabilities, is more cognitively demanding than the rating scale and other techniques (Drummond, O'Brien, Stoddart, & Torrance, 1997b; Furlong et al., 1990). Further, the relative risk associated between sets of probabilities presented in the SG may not be understood by respondents as expected by researchers. Standard Gamble is affected by risk-aversion – a respondent who is risk averse does accept the trade-offs no matter what the probabilities be. Risk aversion leads to an upward bias in SG utilities (Bleichrodt, 2002; van Osch, Wakker, van den Hout, & Stiggelbout, 2004). Also, respondents may find the risk of death to be unacceptable whatever its probability, especially in cases of acute conditions where a full recovery is expected. Naturally, for such conditions, respondents generally are not expected to take any risk of death.

Another bias affecting SG scores is due to loss-aversion. This bias is introduced when respondents are more sensitive to losses than to gains (Tversky & Kahneman, 1992), leading to losses weighing more heavily in decisions and causing an upward bias to SG utilities (van Osch et al., 2004). van Osch and colleagues (2004) also point out that people tend to over-weight small probabilities and under-weight large probabilities. This leads to an upward bias in SG utilities (Bleichrodt, 2002).

As compared to the SG, a TTO may be less cognitively demanding and more comprehensible to a respondent. The utilities may be easier to obtain using the TTO than SG because the TTO does not require an iterative process with changing probabilities. However, like with the SG, the method may have problems for acute conditions where a person might not be willing to give up any life expectancy in exchange for an immediate return to perfect health. The TTO approach is based on the assumptions that the utility time is linear (Torrance, Thomas, & Sackett, 1972; Stiggelbout et al., 1995), which has been stated as untrue (Pliskin, Shepard, & Weinstein, 1980), and that the perception of the severity of illness is independent of the time spent in this state (Rosser & Kind, 1978). In addition, the trade-off concept is difficult for many people to understand.

Overweighting of traded future years in TTO leads to a downward bias in TTO utlities (van Osch et al., 2004). These issues do not affect SG measurements. In the TTO, loss-aversion causes people to be more reluctant to give up life-years, and hence causes an upward bias into measurements (van Osch et al., 2004). Probability weighting is not an issue with the TTO since they do not play a role in TTO measurements. Bleichrodt (2002) contends that since SG is mainly susceptible to upward bias and TTO to both upward and lower biases, TTO is hence preferable overall.

2.3.3 Indirect Approaches

Direct measurement of preferences for health states can be a very resource and time consuming, complex task. An alternate approach bypasses the direct measurement task by indirectly obtain preferences using PBMH. All these PBMH contain a health-state classification system, covering attributes or domains that describe aspects of health that are influenced by disease and treatments, or both. The attributes and levels of severity together form the instrument’s health state classification system. The developers of the PBMH have already established a valuation algorithm that provides utilities for each multi-attribute health state that is conceivable in the health-state classification system. These utilities have been typically obtained by using one of the direct techniques discussed earlier: the rating scale, the SG or the TTO.

The PBMH indirectly provide utilities in two steps. In the first step, the respondent self-rates the extent of impairment, if any, in each attribute using the scales provided in the PBMH. This is the only measurement to be undertaken, on the basis of which the respondents is assigned to a multi-attribute health state. In the second stage, the pre-scored valuation algorithm is applied to provide utilities for the respondent's self-reported current health state.

The most popular PBMH are the Quality of Well-Being (QWB) scale, the EuroQoL-5D (EQ-5D), the Health Utilities Index (HUI), and the SF-6D. Following is a description of these instruments in terms of their classification systems and the utility elicitation technique employed in the formulation of their valuation algorithms:

The QWB (Kaplan & Anderson, 1988; Kaplan & Anderson, 1996) consists of four attributes: mobility, physical activity, social activity, and symptom-problem complex, represented by 38 items. This system describes 1,170 health states. The scoring function of the QWB uses category scaling measurements, and hence provides values and not utilities. The EQ-5D (Brooks, 1996; Kind, 1996) includes five attributes represented by one item each : mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Each attribute has three levels; thus the descriptive system of the EQ-5D can define 243 health states. Both the VAS and TTO have been separately used to obtain scoring functions for the EQ-5D. The TTO was employed in a U.K. community sample (Dolan, Gudex, Kind, & Williams, 1995; Dolan, Gudex, Kind, & Williams, 1996). Recently, a scoring function was formulated for the U.S. population as well (Shaw, Johnson, & Coons, 2005).

The HUI (Feeny, Furlong, Boyle, & Torrance, 1995) is a family of three measures: the HUI Mark 1 (HUI-1), the HUI Mark 2 (HUI-2), and the HUI Mark 3 (HUI-3). The original HUI (Torrance, Boyle, & Horwood, 1982; Boyle, Torrance, Sinclair, & Horwood, 1983) was based in part on the QWB, and has been modified since then. The six attributes covered in the HUI 2 include sensory, mobility, emotion, cognitive, self-care, pain, and fertility, with levels of severity ranging between three and five (Torrance et al., 1996). With 30 items in all, the HUI 2 system describes 24,000 heath states. The HUI 3 system is closely based on the HUI 2, dropping out the application specific attribute – fertility, and adding sensory attributes. The eight attributes now covered in the HUI 3 are vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain, with levels of severity ranging between five and six (Feeny et al., 2002). With 45 items in all, the HUI 3 system describes 972,000 heath states. The HUI 2 and HUI 3 scoring functions were obtained using the VAS transformed into SG using a power conversion algorithm (Furlong et al., 1998).

The SF-6D (Brazier et al., 2002; Brazier & Roberts, 2004b) consists of six attributes which include physical functioning, role limitation, social functioning, pain, energy, and mental health, with levels of severity ranging between four and six. With 36 items in all, the SF-6D system describes 18,000 heath states. The SG was used in the development of the scoring function of the SF-6D.

All the PBMH described above are generic in content that differ mainly in the types of attributes and the degrees of impairment included in their respective health state classification systems. One important difference is in the scope of attributes that they include. In the development of the HUI, a 'beneath of the skin' approach was chosen, that confines the descriptions of items in the measure to individuals. The HUI 2 and HUI 3 do not include a social dimension or any health condition incorporating social interaction or role definition (Feeny et al., 1995; Torrance, Furlong, Feeny, & Boyle, 1995). This approach eliminates problems arising due to complex environmental factors (Gold et al., 1996a) that may often be difficult to measure. In the development of the

QWB, the EQ-5D, and the SF-6D on the other hand, the 'out of the skin' approach was employed. This approach recognizes that aspects of an individual's role functioning and social activities do influence individuals performing a valuation task (Gold et al., 1996a; Brazier, 2005), as this enables them to better understand scenarios presented to them.

2.4 Source of Utilities

The methods used to obtain utilities have been discussed in a preceding section. Sometimes, researchers may not be able to perform such assessments, and hence obtain utilities simply by taking the opinion of clinicians or other experts instead. The primary justification for using clinicians or other healthcare professionals is that they are sufficiently knowledgeable about the health states and may be more accessible, and thus serve as reasonable proxies for patients.

The two main sources of respondents to the direct utility elicitation techniques or the PBMH, however, are patients and the general community. A choice between the two is important because significant differences in the utility estimate provided by these two sources could lead to changes in the results of economic evaluations that use them. There is no consensus on whether community members value a given health state the same as patients who are experiencing that health state. Some studies have shown that patient-based and community-based utilities do not differ (Llewellyn-Thomas et al., 1984; Jenkinson et al., 1997), while some others have shown otherwise (Furlong et al., 1990; Gabriel et al., 1999; Postulart & Adang, 2000).

Since patients have directly experienced a health state, one view is that they can better assess its effect on their life and express a true preference. Thus studies of new therapies are most often conducted from this perspective (Torrance et al., 1982). On the other hand, members of the community may provide more objective evaluations without any bias. Another point of view is that utilities should be obtained from the community as decisions pertinent to public policy should be based on the general population that ultimately incurs the cost of resource allocation decisions (Drummond et al., 1997b; Dolan, 1999).

2.5 Disease-Specific Utilities Using PBMH

In the assessment of HRQoL, two broad categories of instruments have emerged – generic or global, and disease-specific measures. Generic HRQoL scales are designed to be used with any population regardless of the specific disease, and hence allow comparison across diseases. However, the domains or attributes contained in generic measures may have little or no relevance to a specific patient group or disease entity in which they are intended to be used. Generic measures may also lack the items that are necessary to gain a complete understanding of patients' health conditions. Generic measures offer the benefit of being applicable across disease categories, but this can be a disadvantage when research questions pertaining to specific diseases are being studied.

Disease-specific instruments, on the other hand, focus on specific problems posed by a particular illness and reflect the restrictions associated with specific disease conditions. They can include aspects of health considered by patients and clinicians to be of greatest importance. The targeted focus of disease-specific instruments has the potential to make them more responsive to changes in health as a result of improvement in disease prognosis.

These arguments comparing the content of instruments apply to the measurement of health state preferences using generic PBMH as well. Applicable across many different disease areas, they are much easier to administer than obtaining preferences using preference elicitation interviews, and also can be mailed to respondents. However, like generic measures of HRQoL, a general limitation of generic PBMH is that they may lack sensitivity to important differences in particular diseases (Gold et al., 1996a). The validity of the preference weights derived from such measures thus depends upon the coverage of issues important in a particular disease in them, as well as the underlying utility elicitation technique employed.

Most existing disease-specific measures are profile-based and are descriptive in nature, without any assessment of preferences or utilities for health states they describe. There is increasing interest in estimating preference weights for disease-specific measures with the twin objectives of utilizing their disease-specificity, and to enhance their use in economic evaluations. Two distinct approaches have been used in this regard: mapping of disease-specific measures on to PBMH, and developing disease-specific PBMH. These two approaches are described in the following sections.

2.5.1 Mapping Disease-specific Measures on to PBMH

Disease-specific measures can be mapped on to existing generic PBMH to overcome the lack of preference-weights in the former. The 'mapping' process is generally a statistical approach, using regression models, to predict utility scores from scores on a disease-specific measure. In most of the early attempts in this direction, general populations have been considered rather than disease-specific groups.

One variant of this approach involves identifying individual items on health status questionnaires or surveys, such as those from publicly available national databases, which may be considered similar to items on existing generic PBMH. Hence, the simultaneous availability of PBMH is not warranted. For example, identical data elements from the National Health and Nutrition Examination Survey (NHANES) were matched on the HUI-1 index (Gold, Franks, & Erickson, 1996). In other studies, data from the National Medical Expenditure Survey (NMES) were matched on to two multi-attribute preference-based measures, the HUI-1 (Rizzo, Pashko, Friedkin, Mullahy, & Sindelar, 1998), and the EQ-5D (Rizzo & Sindelar, 1999).

The rationale for these studies was to obtain interval level utility values without having to develop scales that provide such assessments. These studies demonstrated that items from the NHANES and the NMES could be grouped and linked in such a way as to obtain health state utility values. The usefulness of this approach, however, depends upon the extent to which items on the surveys can be suitably matched on to items on existing PBMH.

Another variant of this approach requires the simultaneous availability of scores on both the PBMH as well as the disease-specific HRQoL measure in one dataset. A regression equation or algorithm that can predict utility scores based on the scores on the HRQoL measure is then estimated. It is however desirable that the algorithm be obtained from a dataset with a population comparable to the one in which the predicted utilities will be employed. The usefulness of this approach lies in that such an algorithm can be used to perform economic evaluations in the context of other studies where only scores on a descriptive measure is available (Brazier, Deverill, Green, Harper, & Booth, 1999;

Tsuchiya, Brazier, McColl, & Parkin, 2002). Possible reasons for the non-inclusion of PBMH in clinical studies range from the desire to limit patient burden to resource constraints.

In one of the earliest studies using this approach, an empirical equation allowing prediction of the utility scores on the QWB from HRQoL scores using the SF-36 was formulated (Fryback, Lawrence, Martin, Klein, & Klein, 1997) from data in a community-based study. More recent studies have developed mapping algorithms to estimate EQ-5D index scores from the SF-12 in nationally representative US populations (Sullivan & Ghushchyan, 2006; Franks, Lubetkin, Gold, Tancredi, & Jia, 2004) using preferences originally obtained from a community sample in the UK population (Dolan, 1997). A similar mapping study involving the SF-12 (Sullivan et al., 2006) was performed using EQ-5D scores derived from preference weights for the U.S. population (Shaw et al., 2005). As can seen from these studies, utility weights were obtained for generic HRQoL measures.

Similar conversion algorithms have been estimated involving disease-specific descriptive measures. In a community-based study involving patients participating in a weight-loss program, the Impact of Weight on QoL-lite (IWQOL-lite), an obesity-specific HRQoL measure was mapped on to the SF-6D (Brazier, Kolotkin, Crosby, & Williams, 2004a). In another study, the Asthma Quality of Life Questionnaire (AQLQ) was mapped on to the EQ-5D (Tsuchiya et al., 2002). Both studies reported that it was possible to produce robust algorithms that predicted preference scores based on the generic preference measures used in those studies.

However, this may not be the best available solution to the problem of conducting economic evaluations in the absence of preference weights for health states. This is because the preference weights obtained in this manner are limited to the range of the preference weights of the original generic PBMH. There may be aspects of the condition captured by the disease-specific measure that may not be covered by the generic multi-attribute measure and vice versa (Tsuchiya et al., 2002). The formulation of the two types of measures, in terms of the extent to which they address disease-specific issues, can explain these differences. If the spectrum of coverage of these two types of instruments can be considered to be different, it is likely that important dimensions of health may not be valued appropriately (Brazier, 2005).

Brazier (2005) explains two additional assumptions made under this approach - that items representing attributes included in the instruments all have the same importance, and that equal importance is accorded to the intervals included in rating scales of the instruments. Since the extent of convergence between the two measures in a mapping exercise will depend on the type of instruments used and the condition being considered, Brazier and colleagues (2004a) argue that the ultimate utility of this approach is influenced by the degree of this overlap.

2.5.2 Developing Disease-specific PBMH

Mapping exercises, such as those discussed in the previous section, have been described as a second best option compared to directly obtaining utilities using preference-weighted instruments (Tsuchiya et al., 2002; Brazier, Roberts, Platts, & Zoellner, 2005). This suggestion excludes the option of direct elicitation of utilities using techniques like rating scales, standard gamble, and time-trade off because these techniques are time-intensive and expensive. While mapping exercises seem easier to perform in comparison, Brazier and Fitzpatrick (2002) caution that the validity of the approach in estimating utility values for disease conditions will depend on the extent to which the preference-weighted measure used provides coverage of issues regarding the disease and its treatment. On the basis of these inputs, one can infer that in order to obtain utility values that reflect disease-specificity, and that too in a timely manner, disease-specific preference-weighted instruments should be used.

The process of developing new disease-specific measures, like any other measure in any field of research, is iterative. It involves a combination of theory, primary data collection, and statistical analyses. Theory helps in the formulation of a framework to describe the construct being measured. In the development of disease-specific measures, it is necessary to build a pool of items that reflect the areas of impairment caused by the disease (Juniper, Guyatt, & Jaeschke, 1996). In addition to interviewing patients, conducting focus groups, and consulting experts, the process can include a review of literatures and existing instruments (Juniper et al., 1996; Juniper, Guyatt, Streiner, & King, 1997). Primary data collection at the pilot stage helps instrument developers to identify aspects of the instrument that may need modifications. This sets the stage for another round of primary data collection from a larger sample that lends itself to robust statistical analyses, which may include the testing of reliability, validity, and responsiveness of the instrument. Needless to say, the process of developing a new disease-specific PBMH is time-consuming.

2.5.3 Disease-specific PBMH Based on Profile Measures

Feeny (2002a) has summarized the two key components in constructing preference-based measures: 1) designing a health-state classification system that describes health states composed of attributes or domains that influence health; and 2) estimating a valuation function that provides numerical values of utility for all health states that can be described by the health-state classification system.

Profile-based HRQoL measures are descriptive in nature, as discussed previously. Many of them contain multiple domains, with several items measuring the construct represented by each domain. Such descriptive systems may be very complex and it may be very challenging to design valuation functions for them (Brazier et al., 2002). However, it may be appropriate and feasible to develop preference-based measures that are based on the descriptive system offered by existing valid and reliable profile-based measures. This avoids the need to perform extensive research on content identification and item pool composition, thereby aiding the faster completion of component one described above. In a review of methods used to shorten existing HRQoL instruments, Coste and colleagues (1997) discuss these as key reasons considered by researchers who developed shortened versions of existing instruments. This approach was used to generate the classification system for a generic, preference-based measure, later called the SF-6D, from a widely used measure of HRQoL, the SF-36 (Brazier, Usherwood, Harper, & Thomas, 1998; Brazier et al., 2002), and later from the SF-12 (Brazier et al., 2004b) as well. Similarly, a menopause-specific instrument, the QualiPause Inventory (QPI) was used to develop a preference-based measure for the condition (Brazier et al., 2005).

Traditional approaches of identifying item content more applicable to disease-specific areas continue to be used for the purpose of designing disease-specific PBMH. Contemporaneous to the SF-6D was the development of the Asthma Symptom Utility Index (ASUI) (Revicki et al., 1998a) and the Rhinitis Symptom Utility Index (RSUI) (Revicki et al., 1998b). More recently, preference-based instruments have been developed for use in conditions such as stroke (Poissant, Mayo, Wood-Dauphinee, & Clarke, 2003) and erectile dysfunction (Torrance et al., 2004; Casey, Tarride, Keresteci, & Torrance, 2006).

Based on the existing literature, the process involved in directly eliciting preference weights for any condition-specific profile-based measure can be summarized as consisting of the following basic steps:

1. The first step in the process involves making suitable modifications to an existing measure to obtain a more compact disease-specific health state classification system, on the lines of those found in generic PBMH;
2. In the next, a sample of health states derived from such systems are then valued using one of the direct preference elicitation techniques like rating scale, standard gamble or time-trade off; and
3. Finally, these preference weights are modeled using statistical techniques to produce a valuation function that can generate preference weights or utilities for all states described the disease-specific health state classification system.

Steps 2 and 3 together constitute the process of modeling a valuation function for the PBMH. Once a valuation function is made available, the disease-specific preference measure can even be administered, in the form of interviews or even as a survey, to obtain preference weights for the health state of respondents. The following sections discuss these steps in detail.

2.6 Step One: Designing a Health State Classification System

2.6.1 Theories in Measurement

O'Connor (2004) summarizes two different views measurement-related views that influence the development of patient-reported outcomes instruments. One view contends that most psychological and behavioral constructs, although ordinal, can be considered to be interval in nature for the purpose of statistical assessments. An alternate view emphasizes the need for psychological constructs to possess interval properties in order for assessments to be deemed scientific. It has been suggested that the main reason that impedes the acceptance of the analysis of human behavior as a science is the perceived lack of need among researchers in ensuring that these instruments possess interval properties (Wright, 1999; O'Connor, 2004). These diverse viewpoints have spawned comparisons between theories that they inspire – Classical test theory (CTT) and Modern test theory (MTT).

2.6.1.1 Classical Test Theory

Classical Test Theory has been used as the foundation for the development of instruments for several decades now. The theory does not have a single founder, and neither was it explicitly laid out with a set of principles. Rather, CTT refers to a broad range of measurement principles that have been used in the past, and that guided instrument development. Over the years, several instruments have been developed and improvised using the principles of CTT, and have performed well under existing standards of assessments.

However, as in any other field, advancements in the field of measurement of behavior point to some drawbacks to CTTs. One issue pertains to the assumption that it is possible to infer a person's standing on the attribute being measured by summing responses to individual items in an instrument, and calculating a total score. This is a type of additive scaling model that does not take item hierarchy into account (Nunnally & Bernstein, 1994). Inherent in model is the erroneous assumption that each item on that instrument contributes equally to the total score (Prieto, Thorsen, & Juul, 2005). This approach to additivity is said to be a direct derivation from CTT (Crocker & Algina, 1986; Nunnally et al., 1994).

It has been suggested that some of the item performance feature used in CTT like item difficulty, item discrimination as well as reliability are sample-dependent. Item difficulty in CTT can be explained as the proportion of individuals who endorse or pass an item, or give a correct response (which refers to more of the attribute) to a particular item (Kline, 2005). Thus, items with high item difficulty are easy items and vice-versa. Item discrimination in CTT indicates the capacity of an item to discriminate between high and low scorers in the attribute that the instrument measures. Hence, CTT estimators are not generalizable across populations. One of the biggest limitations of CTT is that it does not provide measures with interval properties, but assumes that they can be treated that way for statistical purposes.

2.6.1.2 Modern Test Theory

An alternative scaling approach to the prevailing ideas of the CTT was proposed by the Danish mathematician Georg Rasch (Rasch, 1960), leading to the birth of a new measurement and ideas that are now termed as MTT, which is anchored around Item Response Theory (IRT) models. The Rasch model, also known as the one-parameter logistic model (1PL), is one of the simplest of IRT models. The model works under the condition of unidimensionality, i.e. it is applied to the measurement of a single construct or dimension (O'Connor, 2004). Rasch analysis is a probabilistic model of analysis based on IRT (Rasch, 1960) in which a response to each item in is defined as the result of a linear probabilistic interaction of a person's ‘ability’ and a question's ‘difficulty’. Hence, according to the model, individual items are considered more likely to be endorsed (or rated higher) by subjects at a given level of the construct that is being measured (like HRQoL), while other items are considered more likely to be rated lower. The terms ‘ability’ and ‘difficulty’ come from the educational field, where IRT was first applied.

Summarizing these features, Fisk and Doble (2002) point out that in the Rasch model, the ordering of items is not determined a priori by such means as ‘expert’ opinion but is dictated by empirical data obtained from respondents’ answers. Using such criteria, the Rasch model places items hierarchically and provides the following fit statistics (O'Connor, 2004): item fit, that indicates the extent to which the responses, across persons, fit a particular item; person fit, that indicates the of fit of the responses by a person, across items; and the overall fit of the data to the model. When using the

Rasch model, the emphasis is on the fit of the data to the unidimensional model, not of the model to the data (O'Connor, 2004).

O'Connor (2004) summarizes the advantages that the Rasch model offers over CTT in the measurement of patient-reported outcomes:

1. Both the classical estimates of difficulty and discrimination are sample-dependent. In contrast, one of the unique features of Rasch model is that it provides sample-free measurement and test-free measurement;
2. The main advantage of adopting this approach is that it recognizes that ignoring interval properties leads to poor items. When data fit the Rasch model, it transforms item data from ordinal scores into interval level measurement with the logit (log odds unit) as the unit of measurement based on empirical evidence; and
3. Rasch models are more robust to missing data, in comparison to CTT where data have to either be discarded or imputation employed (Bond & Fox, 2001).

2.6.1.3 Using Theories of Measurement in Instrument Design and Modification

The basic structure of generic PBMH has been discussed in an earlier section. Existing PBMH vary in their coverage of attributes, and hence the number of health states they represent. Structurally, profile-based measures are similar, in that they contain items representing various domains that are important in the measurement of the overall construct they measure, like HRQoL. However, they are typically longer than PBMH. The items in a profile-based measure and the levels of those items together can generate descriptive systems, like those in PBMH, but the number of health states would be too large. Consequentially, there would be problems in valuing such a large number of health states. This was an issue that Brazier and colleagues (2002) reported in the process of developing a generic PBMH, the SF-6D, from a generic measure of HRQoL, the SF-36. The approach chosen was to reduce the number of items of the parent HRQoL instrument without loss of descriptive information.

Efforts to produce shorter HRQoL instruments, in order to reduce respondent burden, have typically involved reducing existing instruments. This has mostly been done using various statistical methods available for this purpose, including correlations between item and total scores, Cronbach's alpha per scale, stepwise regression, and factor analysis (FA) (Coste, Guillemin, Pouchot, & Fermanian, 1997); most of these are based on CTT. However, there is lack of standardization in the methodology of shortening existing HRQoL instruments (Coste et al., 1997; Prieto, Alonso, & Lamarca, 2003).

Among these techniques, FA has been widely used in the development and validation of HRQoL instruments. One of the objectives of applying the technique to HRQoL instruments is for construct validation (Fayers & Machin, 2000). When little is known about the structure of an instrument, FA helps identify the existence of latent constructs, called factors, as well as the associations of individual items to those factors. When there is enough available information on the structure of HRQoL instruments, FA helps confirm the number of latent factors and to confirm the items in the instrument that are better associated, or that contribute to, the measurement of that factor (Fayers et al., 2000). Since FA is a parametric method requiring interval level data, doubts have been raised about the use of this technique with measures that do not have interval-level properties (Doward, Meads, & Thorsen, 2004).

Rasch analysis, introduced in the previous section as an alternative scaling approach, can also be used as a reduction procedure (Rasch, 1960). When used with HRQoL instruments, categorical items are mapped on to a continuous latent scale which can be said to be a continuous measure of HRQoL. The use of the technique for the purpose of item reduction of existing HRQoL measures is growing. Rasch analysis has been recommended as a preferred technique in reducing redundant items in an HRQoL instrument (Conrad & Smith, Jr., 2004; Tennant, McKenna, & Hagell, 2004). Some examples of instruments that have been shortened using the technique are the Nottingham Health Profile (Prieto, Alonso, Lamarca, & Wright, 1998), the Fatigue Impact Scale (Fisk & Doble, 2002), the Asthma Quality of Life Questionnaire (Young, Yang, Brazier, & Tsuchiya, 2005), and a dermatology-specific index called Skindex-29 (Nijsten, Sampogna, Chren, & Abeni, 2006).

A couple of studies have provided some useful insight by comparing the Rasch model and the CTT approach in reducing HRQoL instruments. In one study using data on the Nottingham Health Profile, the two shortened versions using the two different approaches were found to be comparable in measurement properties to the original instrument but they well differed in the items chosen from the original instrument (Prieto et al., 2003). Another study using Impact of Psoriasis Questionnaire data reported similar results with regards to item selection, but suggested the use of the Rasch-shortened version largely due to its unidimensionality and resultant ability to produce an outcome measure in a single numeric (Nijsten, Unaeze, & Stern, 2006).

2.6.2 Patient-Reported Outcomes of Diabetes and its Management

Type 1 and Type 2 diabetes are chronic illnesses that can influence patient-reported outcomes because the treatments are burdensome and the complications can be debilitating and/or life-threatening. For those with diabetes, the disease and the demands of its day-to-day management can be very challenging. Patients have to deal with their diabetes almost every instant of their life and have to make continuous decisions that interfere with living a normal life. The management of diabetes itself imposes considerable demands on patients and their families, and affects patients both physically and psychologically. Patients with diabetes may feel overwhelmed by the management of the disease. Apart from the emotional and social burdens this may cause, they face the acute physical distresses of hypoglycemia or hyperglycemia and chronic physical distress of diabetes-related complications. Thus, diabetes has major psychosocial implications and it influences self-management behavior in terms of diminished self-care, leading to worsened glycemic control in the long run.

Satisfactory diabetes control can be achieved when this interdependence between physical and psychological well being is addressed (Eiser & Tooke, 1993). In this context, humanistic assessments, using validated and reliable instruments of health status (or HRQoL), QoL, and other patient-reported outcomes, can play a role in predicting individuals' capacities to manage their diabetes and stay healthy in the long run. Preference-based measures of health additionally provide the necessary metric to conduct the economic evaluations of diabetes interventions. Polonsky's (2000) suggestions can be a useful source of guidance in determining the classification for a diabetes-specific PBMH. The following three discussions deliberate on the impact of diabetes on physical, psychological and social functioning, and list Polonsky's suggestions on how best to measure these outcomes from the perspective of an individual with diabetes.

Impact of Diabetes on Physical Functioning

Diabetes can negatively affect physical well-being in three major ways. The most important factor is the development of long-term complications like vision loss, kidney damage, peripheral neuropathy resulting in chronic pain, amputation, and/or difficulty walking. Other complications include sexual dysfunction, autonomic neuropathy problems, and acute conditions like ketoacidosis. The Pittsburgh Epidemiology of Diabetes Complications Study showed that patients with macrovascular disease or nephropathy reported significantly poorer QoL compared with those who were free from all complications, and that QoL significantly deteriorated according to the presence of multiple complications (Lloyd, Matthews, Wing, & Orchard, 1992; Lloyd, Wing, Orchard, & Becker, 1993).

The second factor is short-term complications and physical symptoms. Elevated blood glucose levels may lead to increased fatigue, sleep problems, and other associated problems. Tight glycemic control may lead to unwanted weight gain, hypoglycemia, and/or loss of hypoglycemic warning signs. The third major factor is the lifestyle changes resulting from the demands of the diabetes regimen.

Polonsky (2000) suggests that to assess the impact of diabetes on physical functioning most effectively, evaluation should focus on a patient's perceived distress due to diabetes-specific symptoms as well as the perceived loss of physical function, interference with common activities and loss of independence due to diabetes.

Impact of Diabetes on Psychological Functioning

Diabetes care can have a short-term and long-term impact on mood of patients. Frustration can emerge out of the fact that the disease may not seem to respond in spite of sincere efforts by patients. Cycles of elevated blood glucose levels and hypoglycemic episodes can be exhausting, and can worsen already dampened spirits. Depression is not generally listed as a complication of diabetes, but is widely prevalent in patients with diabetes. Lloyd and colleagues (Lloyd et al., 1992) reported greater depressive symptoms in patients with macrovascular disease; greater number of complications were found related to higher depression symptom scores. There is some suggestion that the stress of depression may lead to neglect of diabetes care. Polonsky (2000) suggests that to assess this dimension, evaluation should focus on a patient’s perceived emotional distress due to diabetes-related symptoms, self-care, and broader diabetes issues.

Impact of Diabetes on Social Functioning

The management of diabetes itself poses many challenges to a patient, as this may necessitate changes in daily habits in order to manage the illness most effectively. For instance, some patients are embarrassed to check their blood glucose or inject insulin in front of others. For some, the requirement of meal planning may affect food choices at social events that may be different from family/friend preferences. Thus, a patient with diabetes may not receive all the cooperation from family and friends in social settings, be it home or outside of home. Polonsky (2000) suggests that to assess this dimension, evaluation should focus on a patient's perceived emotional distress due to diabetes-related social situations.

2.6.3 The Audit of Diabetes Dependent Quality of Life

Among diabetes-specific QoL measures, the Audit of Diabetes Dependent Quality of Life (ADDQoL) is an individualized instrument designed to measure individuals' perceptions of the impact of diabetes on their QoL (Bradley et al., 1999; Bradley & Speight, 2002). The design of the ADDQoL is influenced by the development of the Schedule for the Evaluation of Individual Quality of Life (SEIQoL), an interview-based approach to QoL measurement (O'Boyle, McGee, Hickey, O'Malley, & Joyce, 1992). The SEIQoL method involved asking the respondents to generate domains of life that are important to them, evaluate how good or bad each aspect was currently felt to be, and indicate the importance of each for their own QoL. This approach was adapted to address diabetes-specific issues and presented in a questionnaire format, resulting in the creation of the ADDQoL. The last published version of ADDQoL is a 18-item instrument that presents a comprehensive list of 18 life domains that diabetes might affect (Table 1). Two additional items report estimates of overall quality of life, comparing life with and without diabetes.

The ADDQoL evaluates diabetes-specific QoL from an attributional perspective, i.e. how diabetes may be perceived as interfering with well-being. In contrast, the majority of the other diabetes-specific instruments assess QoL from an intrinsic perspective, i.e. how the different aspects of diabetes may be perceived as burdensome. Rather than asking about the degree to which problems associated with diabetes are occurring, the ADDQoL asks patients to imagine how life might be different without diabetes and compares it to their current QoL with diabetes. This is a more complex task, one step removed from direct questions about diabetes-specific quality of life (Polonsky, 2000). One advantage of this approach is that it is unbiased in that it allows respondents to indicate how diabetes may be having a positive effect in certain domains.

In answering the ADDQoL, respondents rate how diabetes impacts individual items on a seven-point scale (the impact rating), as well as how important the individual items are to their QoL on a four-point scale (the importance rating). A 'not applicable' (N/A) option is provided for three items that may not be applicable to a given individual. Impact ratings when multiplied by the respective importance ratings yield scores ranging from –9 to 9 for each item (Table 2). An average weighted impact score is derived by summing the weighted impact scores for each item and dividing the number of applicable items. Thus, the patient's 18 scores can then be arithmetically weighted, such that the total score is more strongly influenced by those items that a patient has selected as being most important. Bradley and colleagues (1999) assert that none of the other existing diabetes-specific QoL measures contain this property.

Patients and diabetes experts were involved in the generation and confirmation of items (Bradley et al., 1999), contributing to face and content validity. Reliability for the ADDQoL, using Cronbach's alpha coefficient has been found to be in the range of 0.85 to 0.92 (Bradley et al., 1999; Sundaram et al., 2006). This property, along with the results of the forced one-factor factor analysis, indicated the feasibility of combining the weighted items into a single ADDQoL score (Bradley et al., 1999; Bradley et al., 2002). As a measure of the construct validity of the measure, insulin-users generally reported greater negative impact on the ADDQoL than non-users, and patients with diabetes complications reported significantly greater negative impact of diabetes on QoL than did those without complications (Bradley et al., 1999; Bradley et al., 2002). Recently, Sundaram and colleagues (2006) demonstrated the utility of the instrument in measuring QoL in patients with Type 2 diabetes in the presence of conditions commonly co-morbid to the disease. The instrument has been reported to be able to detect negative influence of diabetes on QoL even as satisfaction with the treatment was high (Bradley et al., 2002), and to detect significant changes in QoL over time due to the effects of a diabetes educational intervention (The DAFNE Study Group, DAFNE Study Group, 2002).

Table 1: Content of the ADDQoL

Item Number	Item Description
1	Working life
2	Family life
3	Social life
4	Sex life
5	Physical appearance
6	Physical activities
7	Holidays/ leisure
8	Ability to travel
9	Confidence in ability
10	Motivation
11	Society reaction
12	Worries about Future
13	Finances
14	Dependence
15	Living conditions
16	Freedom to eat
17	Enjoyment of food
18	Freedom to drink

Table 2: Summary of the Scoring of the ADDQoL

Weighted ratings	=	[unweighted rating (-3 to +3)] x importance rating (0 to 3) for each domain]
		Unimportant domains score 0, regardless of magnitude of effect of diabetes. Domains unaffected by diabetes score 0, regardless of their importance for QoL. Any non-applicable domains are not scored.
ADDQoL score	=	$\frac{\text{Sum of weighted ratings of applicable domains}}{\text{N of applicable domains}}$
Scores vary from:		-9 (maximum negative impact of diabetes)
to:		+9 (maximum positive impact of diabetes)

New Scoring of the ADDQoL (unpublished)

Weighted ratings	=	[unweighted rating (-3 to +1)] x importance rating (0 to 3) for each domain]
		Unimportant domains score 0, regardless of magnitude of effect of diabetes. Domains unaffected by diabetes score 0, regardless of their importance for QoL. Any non-applicable domains are not scored.
ADDQoL score	=	$\frac{\text{Sum of weighted ratings of applicable domains}}{\text{N of applicable domains}}$
Scores vary from:		-9 (maximum negative impact of diabetes)
to:		+3 (maximum positive impact of diabetes)

In separate reviews, the ADDQoL has been recommended for use in QoL assessments in both Type 1 diabetes and Type 2 diabetes populations, and has been described as a brief and recent instrument generated with patient input, with good reliability, internal and external construct validity (Garratt et al., 2002; Wildes et al., 2003). As an established diabetes-specific instrument with reported use in different populations, the ADDQoL can be a suitable candidate to determine item content for a proposed diabetes-specific PBMH.

2.7 Steps Two & Three: Modeling Health State Valuations

Once a health-state classification system has been described, the next task involves incorporating stated preferences into it. One logical way to proceed would be to list all the health states that can be described by the classification system and then obtain preferences weights or utilities for those health states. However, the classification, depending on the number of attributes or domains contained in them, may be large to the extent that the valuations of all the health states it defines may be considered infeasible. The total number of such health states will depend upon the number of attributes and the level of severity assigned to each attribute. For example, the classification systems in the EQ-5D, SF-6D, and the HUI-3 describe 243, 18,000, and 972,000 health states, respectively. Valuing such a high number of health states, especially for the SF-6D and the HUI-3 is a complex task. The approach used so far has been to elicit preferences only for a sample of health states and then extrapolate them to all health states (Feeny et al., 1995; Dolan, 1997; Brazier et al., 1998).

There are two basic approaches to model health state valuations, although the literature does not use standardized terminology for these approaches. Froberg & Kane (1989) and Dolan (2000; 2002) refer to them as the decomposed versus the composite approach, Feeny (2002b) uses the terms decomposed versus statistical methods of interference, while Brazier and colleagues (2002) describe them as methods based on multi-attribute theory versus econometric methods, respectively. For the purpose of standardization, the terms decomposed and composite will be used henceforth. Both approaches have been used in modeling health state valuations for existing preference-based instruments, and according to Feeny(2002b), the two approaches are not mutually exclusive.

2.7.1 Modeling Considerations

The following discussions will help better appreciate the difference between the two modeling approaches.

First, it will be useful to revisit the content of preference instruments. These instruments are made up of attributes or domains that describe aspects of health that are influenced by disease and treatments, or both. Respondents are allowed to rate each attribute on a rating scale that denotes the extent to which disease and treatments, or both, influence that attribute. The attributes and levels of severity together form the instrument's health state descriptive system. Since the combinations of attributes and severity levels give rise to several hundreds or thousands of health states, called composite health states since they include all aspects of health described by the instrument, it will be difficult to value them entirely.

Second, the kind of interactions between attributes needs to be discussed. Three such interactions are possible: no interactions, synergistic interactions, and antagonistic interactions (Torrance et al., 1995; Furlong et al., 1998). When attributes are complements of one another, the combined preference weights will always be more than the sum of weights for each attribute alone. This is a synergistic preference interaction in that the total is greater than the sum of effects of it parts. In other words, to enjoy one attribute, you may also need to have the other attributes (Keeny & Raiffa, 1976; Keeny & Raiffa, 1993). When attributes are substitutes for one another, the combined preference weights will always be less than the sum of the weights for each level. This is an antagonistic preference interaction in that the total is less than the sum of effects of it parts. In other words, either attribute is good enough, since one substitutes the other (Keeny et al., 1976; Keeny et al., 1993).

Third, it will also be helpful to define some functional forms governing the relationship between attributes. Torrance and colleagues (1995) explain that one of the following types of utility independence is necessary, as an additional assumption, to be able to extend the traditional vN-M theory to multi-attribute outcomes that we come across in healthcare applications:

1. The additive functional form (also referred to as additive utility independence) assumes attributes to be independent, permitting no interaction. This is the most restrictive form (i.e. it makes the strongest assumption) in which preference weights for each level of attributes are simply added to one another. In measurement terms, only the preferences for the levels of individual attributes need to be considered, ignoring the manner in which levels of different attributes are combined (Torrance et al., 1995).

Hence, additive models are considered appropriate only if the interactions in preferences among attributes are not important. Data from health status applications have generally rejected this model due to this constraint (Furlong et al., 1998);

2. In the multiplicative functional form (also referred to as mutual utility independence), interactions between attributes are allowed but the interactions are forced to be of the same kind (Furlong et al., 1998; Dolan, 2002). This functional form specifies no interaction between preferences for levels on some attributes and fixed levels for other attributes (Torrance et al., 1995). In other words, the interaction is limited in that it is the same between all attributes and for all levels of each attribute (Furlong et al., 1998). The multiplicative functional form is said to be the least complex among non-additive models (Furlong et al., 1998); and

3. The multilinear form (also referred to as first-order utility independence) specifies no interaction between preferences for levels on any one attribute and fixed levels for the other attributes (Torrance et al., 1995). This functional form is the least restrictive functional form, allowing some pairs of attributes to be complements and other pairs to be substitutes; in other words, it is the weakest form of utility independence (Torrance et al., 1995). However, these are very complex models requiring too large number of valuations (Furlong et al., 1998).

Details of the decomposed and composite approaches are discussed in the forthcoming sections.

2.7.2 The Composite Approach

In the composite approach, a subset of such composite health states is valued. Related to this approach is the method of directly valuing a profile-based measure (Brazier, 2005) in which the measure is administered along with a self-administered valuation technique, like Lundberg and colleagues who used the TTO along with the SF-12. When the profile-based score is regressed against the utility score derived from the valuation technique, preference weights for the profile-based score can be obtained. However, the drawback of this technique is that the utility scores can only be obtained for health states that naturally exist in the population, which may exclude rare health states, thereby impact the prediction capability of the regression model (Brazier, 2005).

Hence, the composite approach advocates the selection of composite health states from the descriptive system by statistical design. One way to do this is by using an orthogonal design for estimating an additive model (Brazier, 2005). The composite approach was used in the estimation the EQ-5D (Dolan, 1997), the SF-6D (Brazier et al., 2002) and in developing a menopause-specific preference-based index (Brazier et al., 2005), among others. In the development of the menopause-specific index, for example, a statistical program was used to generate an orthogonal array of health states that had to be valued to fit an additive model. While the program selected 49 health states as essential for the modeling valuation exercise, 47 more were randomly indicated to enhance the number of degrees of freedom (Brazier et al., 2005). These many health states had to be valued for the menopause-specific descriptive system that had seven attributes and between three and five levels per attribute. In the SF-6D, with six attributes that had between two and six levels per attribute, the number of health states selected was 249, of which 200 were randomly selected (Brazier et al., 2002). The requirement will be greater for more complex descriptive systems.

2.7.3 The Decomposed approach

The 'decomposed' approach advocates valuing a health state at the level of its constituent attributes, and modeling the values of individual attributes, by making theoretically sound assumptions, to obtain preferences for the overall health state. In the decomposed approach, each level within a particular attribute or domain from the health state classification system is valued, keeping the levels of all other attributes constant. This approach makes simple assumptions about the relationship between the attributes (Dolan, 2002) that are consistent with the multi-attribute utility theory (MAUT) (Keeny et al., 1976; Keeny et al., 1993). The MAUT is an extension of conventional utility theory that can be applied to model preference scores for health states defined by multi-attribute health status classification systems (Torrance et al., 1982; Torrance et al., 1995; Torrance et al., 1996). The theory specifies that the preferences should be measured within the axioms of the expected utility theory, and can be used to fit additive, multiplicative and multilinear models depending on the type of preference interaction (or independence) among attributes (Torrance et al., 1995). In the decomposed approach, MAUT is used to determine the sample of states to be valued, with the most commonly used specifications being the additive and multiplicative forms (Brazier, 2005).

The valuation process under MAUT comprises the following steps (Furlong et al., 1998):

1. *Single-attribute level states* are valued. In a single-attribute level state for attribute A, attribute A is at less than full function and all other attributes are set at their best levels. If attribute A has five levels, then it has four single-attribute level states, each containing varying levels of attribute A (except the best level) and all other attributes at their best level;
2. *Corner states* are valued. Corner states are those states where one attribute is at the lowest level and all other attributes are their best level. There will be as many corner states as there are attributes. The use of corner states makes the necessary MAUT-stipulated calculations easier (explained in detail in section 3.3.2.2). A corner state is a special case of a single-attribute level state where, using the previous example, A is at level five (lowest level for A) and all other attributes are at their best level ;
3. A group of *methodological marker states* are valued. These are a set of multi-attribute determined by the model specification, and spanning a wide range of severity across attributes; and
4. Other *anchor states* valued are dead, perfect health (highest level of function on all attributes) and pits (lowest level of function on all attributes).

The MAUT has been used in the development of two generic PBMH, the HUI-2 and the HUI-3 (Torrance et al., 1995). For the HUI-3 (having eight attributes and between five and six function levels for each attribute), a total of 43 health states were valued to fit a multiplicative functional form (Furlong et al., 1998; Feeny et al., 2002). Two of the earliest disease-specific preference measures to be valued using MAUT were the Asthma Symptom Utility Index (ASUI) (Revicki et al., 1998a) and the Rhinitis Symptom Utility Index (RSUI) (Revicki et al., 1998b); both were based on a multiplicative model. In the valuation of the ASUI, consisting of four symptoms/attributes, two dimensions (frequency and severity), and with a four point rating scale for both these dimensions, a total of 23 health states were valued (Revicki et al., 1998a). A same number of health states were valued for a similarly structured RSUI (Revicki et al., 1998b).

The single-attribute level states, corner states, methodological marker states and anchor states can be valued using any of the established preference elicitation techniques. For the HUI 2 and the HUI 3, VAS rating scales were used to perform most of the valuations which were then converted to SG utilities, which were also collected for a sample of the health states in the study, using a power transformation curve (Furlong et al., 1998; Torrance et al., 2001). This approach was also employed by Revicki and colleagues in the estimation of valuation functions for the ASUI (1998a) and the RSUI (1998b).

2.7.4 Choosing the Modeling Approach

An advantage of the composite approach is that it accounts for heterogeneity in the data (Dolan, 2002) by using statistical modeling using random effects techniques or by modeling mean health state values, as noted by Brazier (2005) citing methods used for valuation in the EQ-5D (Dolan, 1997) and the SF-6D (Brazier et al., 2002). Few *a priori* restrictions need to be placed on models using this approach (Dolan, 2002). Also, there is no need to specify corner states, which may be difficult for respondents to comprehend (Brazier, 2005). In general, the composite approach requires a large sample of health states, as can be seen from the example cited above, because of which each respondent has to value more states. Brazier (2005) contends that with the composite approach, there are difficulties in modeling interaction terms that require a random addition of health states to be valued, as was discussed with the example of the menopause-specific measure. Also, it is important that the health states selected be widely spread over the valuation space so that many combinations of attributes and levels be included (Dolan, 2002).

One of the main advantages of the decomposed approach is that the complexity of the valuation task is reduced (Dolan, 2002; Brazier, 2005). One of the reasons contributing to the simplicity is that for the most part in the valuation exercise, respondents have to consider only one attribute with impairment at a time rather than several different attributes at the same time (Dolan, 2002). However, respondents may also find it difficult to value combinations that they find infeasible – for example, impairment in one attribute and the best levels in other attribute, especially if the attributes are related (Dolan, 2002; Brazier, 2005). By making simplifying assumptions about the relationships between attributes, this approach requires less number of valuations, and can be used to model almost an infinite number of health states (Dolan, 2002). Additionally, these assumptions are consistent with a well-specified, theoretical model like the MAUT (Dolan, 2002).

In both the decomposed and composite approaches, the required sample size for the utility elicitation process will depend upon the extent of restrictions placed on the models that are to be estimated. The less restrictive the models, the more states that will have to be valued (Dolan, 2002). Considering that there needs to be a limit on the number of health states that can be valued per subject to minimize respondent burden, a greater number of health states will require planning to recruit more respondents. Feeny (2002b) recommends that while it is important to ascertain the suitability of the functional form chosen, another desirable feature would be the requirement of fewer number of health states to be valued to estimate the valuation function. Hence a discussion on functional form is needed when designing valuations for a health state descriptive system since this has a bearing on planning of available resources to perform this task.

In choosing the right modeling approach, it is useful to discuss the relative merits and demerits of the composite and the decomposed approaches, as discussed above. Those who attach importance to an approach with a theoretical basis should choose the decomposed approach, whereas the composite approach would be the choice when the ability to estimate interactions between attributes is needed (Dolan, 2002). Another important practical consideration is regarding the time and resources available to do the modeling task. If there are limitations on the availability of subjects to perform the valuations, then the method of choice is the decomposed approach. The composite approach is suited to situations where the ratio of direct values (utilities for measured health states) to estimated values (utilities estimated by modeling) is relatively high (Dolan, 2002). Finally, available expertise to conduct the research may be a factor in choosing one approach over the other – the composite approach requires use of econometric-type modeling, while the decomposed approach requires an understanding of the MAUT to health-services research.

CHAPTER THREE: METHODOLOGY

3.1 Study Design and Population

For Objective One of the study, a combination of exploratory and secondary research was used to develop a brief diabetes-specific health state classification system. For Objectives Two and Three, cross-sectional designs using representative samples of patients from their respective study populations were employed.

The study population of interest for the pilot surveys of the newly developed diabetes-specific health state classification system that were part of Objective One, and for the one-on-one utility elicitation interviews in Objective Two included patients with diabetes, either Type 1 or Type 2, in a combination of outpatient and community settings in North Central West Virginia (WV). Following are descriptions of the venues where the study populations were recruited:

1. The Diabetes Education Center at Ruby Memorial Hospital, Morgantown, WV, is a department of West Virginia University Hospitals (WVUH) and is located in the hospital building. The Center offers inpatient and outpatient diabetes education by Certified Diabetes Educators (CDE), dietitians, and nurse educators. Outpatient education is offered in both a one-on-one setting for those requesting it and group classes. Group classes are offered for those patients with Type 2 diabetes and those patients on insulin, whether they have Type 1 or Type 2 diabetes. The group classes contain three to four sessions over a three to six-month time frame, with up to six patients in each group. On an average, two new classes are held each month;

2. Milan Puskar Health Right Clinic is a free clinic located in downtown Morgantown, WV. One of the services provided at the clinic is a diabetes education program. Two 1-hour classes are held each Tuesday afternoon. There are 1 to 8 participants in each class with some participants being newly diagnosed with diabetes and others having had the disease for several years. Patients new to the classes complete a series of six sessions covering the basics of diabetes self-management. Following these six classes, participants return once a month for continued education and support;

3. Monongalia General Hospital is a general, acute care hospital located in Morgantown, WV. Diabetes education programs at this site are conducted by CDEs, dietitians and nurses. Two different kinds of education programs are offered. Classes for those patients with Type 2 diabetes who do not use insulin are group classes, with four sessions conducted over a three-month period. The classes for those patients who are on insulin (whether they have Type 1 or Type 2 diabetes), are one-on-one classes with four sessions conducted over a four-month period. On average, two new classes begin each month, with eight to ten patients participating in each session;

4. The Rite Aid Pharmacy #1982 is a community pharmacy located in Grafton, WV. The Pharmacy Care Center located in the pharmacy currently utilizes the Rite Solutions Diabetes Self-Management Education programs for diabetes education. The Grafton Rite Aid employs five registered pharmacists, one dietician/CDE, and one registered nurse/CDE to provide patient education. Diabetes education classes are offered both on an individual basis and in groups of three to five, and target patients with Type 2 diabetes;

5. Other outpatient venues included the diabetes education programs conducted by the Departments of Family Medicine and Endocrinology at WVU; and

6. Diabetes Support Group meetings conducted in a community setting in Morgantown, WV, and Fairmont, WV. These meetings are typically conducted once a month by diabetes education centers operating in the local area at community settings such as churches.

The study population of interest for Objective Three included patients with diabetes receiving care at the outpatient clinics of the West Virginia University Diabetes Institute (WVUDI). The WVUDI is a collaborative effort to standardize outpatient diabetes education and care imparted by the department of Family Medicine and the Physician Office Center's Medical General Practice (MGP) and Endocrinology clinics. While the department of Family Medicine sees patients of all ages including children, the MGP and Endocrinology clinics provide care for patients older than 18 years of age. In a mailed survey, patients receiving care at the WVUDI provided responses to the health state classification system developed in Phase One.

3.2 Sampling Design and Patient Recruitment

Objective One commenced with the analysis of a dataset containing responses of 385 patients with Type 2 diabetes to the ADDQoL from previous research (Sundaram et al., 2006) and consultations with experts. Different versions of the health state classification system were then piloted on convenience samples of people with diabetes, either Type 1 or Type 2, at the sites described above, in the form of self-administered surveys.

Objective Two of the study similarly utilized a convenience sample of patients who responded to flyers posted in the outpatient settings and to handouts distributed in the community settings described in the previous sections. In order to participate in the study interviews, patients had to a) have diabetes, either Type 1 or Type 2, b) be at least 18 years of age, c) be without significant visual or hearing impairment, and d) be conversant in written and spoken English. The flyers indicated a participant reward of $15 for individuals meeting these criteria and completing the interviews.

Objective Three of the study included a survey mail-out to a convenience sample of patients receiving care for their diabetes, either Type 1 or Type 2, at the outpatient clinics of the WVUDI. An individual from the WVU Hospital's Office of Medical Staff Affairs (OMSA) served as a coordinator for identifying patients with diabetes from the WVU hospital's medical records, while maintaining compliance to the Health Insurance Portability and Accountability Act (HIPAA). The physicians currently providing care to these patients were contacted and informed about the purpose of this study. Letters of endorsement were obtained, in which the respective physician's signature was affixed to a personalized letter addressed to individual patients on OMSA letterheads. The letters indicated a participant reward of $5 for individuals completing the survey.

Separate approvals were obtained from the WVU Institutional Review Board to conduct the pilot surveys (Objective One), the utility elicitation interviews (Objective Two) and the survey mail-out (Objective Three).

3.3 Data and Measures

3.3.1 Objective One

The methodology involved in the process of finalizing the diabetes health state classification system is summarized in Figure 2. Objective One commenced with the use of a dataset with ADDQoL responses from 385 patients with Type 2 Diabetes collected in previous research (Sundaram et al., 2006). In answering the ADDQoL, respondents rated the impact of their diabetes on (those applicable among) 18 items representing domains of life, and also indicated how important they considered those domains. Three variables were available for each ADDQoL item in the dataset – a seven-point un-weighted impact rating (-3 through zero to +3) and a four-point importance rating or weight (0 to 3), both indicated by the respondent, and a calculated weighted impact rating (-9 through zero to +9). This dataset was used to perform the necessary statistical analyses that informed the process of gathering input for the classification system from the items contained in the ADDQoL.

In addition to performing statistical analyses, subjective input was obtained from an expert panel in order to ensure adequacy of content. The expert panel consisted of diabetes clinicians, diabetes educators, registered nurses and a psychometrician. The seven members were: Betsy Elswick, PharmD, and Tara Whetsel, PharmD from the Clinical Pharmacy department at the WVU School of Pharmacy; Charlotte Nath, RN, EdD, CDE, from the Family Medicine department at the WVU School of Medicine; Elizabeth Quintana, RD, CDE, from the Endocrinology section of the department of Medicine at the WVU School of Medicine; Andrea Hasley, RD, CDE, from the WVU Diabetes Center; April Lombardo, RN, CDE, from the Monongalia General Hospital

Diabetes Education Center; and Clarice Hayes, EdD from the Global Health Outcomes group at Eli Lilly & Company. A modified method based on the Delphi technique was used in order to gather consensus on the item content of the proposed diabetes health state classification system. The Delphi technique, which was originally developed by the Rand Corporation for technological forecasting, is commonly used in medical, nursing and health services research (Jones & Hunter, 1995; Hasson, Keeney, & McKenna, 2000).

One of the reasons for the choice of this technique over other consensus methods, such as the Nominal Group Technique, was that it does not require that the panelists be brought together physically for group discussions (Jones et al., 1995; Hasson et al., 2000). Hasson and colleagues (2000) report in their review on the Delphi technique that while the classical Delphi had four rounds, two to three rounds are preferred, and that the number of panelists involved in studies employing the technique were as low as 15 to as high as 60. Although the technique can be administered using any combination of face-to-face interviews, and communication by postal mail, email, and facsimile, it has been recommended that personal interviews be employed in the first round (McKenna, 1994).

The objective of constituting the expert panel was to obtain subjective input on: 1) arriving at a four to five-attribute classification system based in part on the results of the statistical analysis of the ADDQoL data; and 2) describing the attributes and severity levels of the classification system in a concise manner using situations relatable to people with diabetes. Each member was individually contacted in order to explain health outcomes concepts including ‘utility’ and ‘value’ and the above-mentioned objectives in one-on-one discussions. As a handy reference, they were emailed a written document

(Appendix 1) containing an explanation of the concepts and links to useful websites that describe them as well. In addition, the document listed the specific expectations from the panel members in the form of distinct action items, as well as examples of health state classifications systems from both generic and disease-specific PBMH. This was followed by qualitative-input gathering exercises conducted using structured formats explained in the paragraphs below.

In Round One, Panel members were emailed a document (Appendix 2) containing a list of plausible attributes based on statistical analysis of the previously collected ADDQoL dataset and our own interpretation and intuition. Each panel member was requested to perform the following tasks electronically within the document:

1. Rate the importance of each aspect of life (represented by the respective ADDQoL item, on a scale where '0' is 'not at all important', and '3' is 'very important') in terms of its influence on the utility that a person with diabetes may place on his or her health condition;
2. Describe briefly in words how, using their own experience in dealing with persons with diabetes, this aspect of life may be influenced by diabetes;
3. Suggest an attribute (from the plausible list provided in the document or any another attribute that the panel member may consider appropriate) that the ADDQoL item under consideration best fits under;
4. Indicate other items not included in the ADDQoL, but which may be needed to be considered, and perform the above three tasks for those additional items as well.
5. Update names of the attributes and additionally provide a brief description for each of the selected attributes; and

6. Describe up to four attribute levels, in sentences, for each of the selected attributes.

The answers to the structured qualitative input-gathering exercise in Round One were pooled together to create a suggested version (Suggested Version One, Appendix 3) of the heath state classification system. As a general format, it was decided to incorporate the following components in all versions of the heath state classification system under development:

a. A title followed by an introduction to the questionnaire;
b. A table summarizing the three aspects related to diabetes care that the respondent is requested to keep in mind (wherever applicable) while answering all questions: Symptoms, Complications, and Daily Care needs; and
c. One question each pertaining to the attributes selected for inclusion in the classification system and including:
    i. A one-sentence question;
    ii. A bulleted paragraph containing a description of the attribute under consideration, put together by listing the items that were thought of as comprising the attribute from the qualitative analysis conducted at the end of Round One; and
    iii. Sentences describing severity levels for the attribute.

In Round Two, Suggested Version One was emailed to the panel members. Each panel member was requested to perform the following tasks electronically within the document:

1. Substitute words or phrases that could potentially bias patients' perceptions of the impact of diabetes with alternative words or phrases, keeping in mind the need to remind patients, via the classification system, of the potential challenges (or responsibilities) that diabetes may introduce into their lives;
2. Substitute complex words or phrases with simpler words or phrases (seventh grade readability level was desired) and make the classification system less verbose without compromising its diabetes-specificity; and
3. Comment on the appropriateness of the grouping of items into attributes and the manner in which attributes were described on the basis of their constituent items.

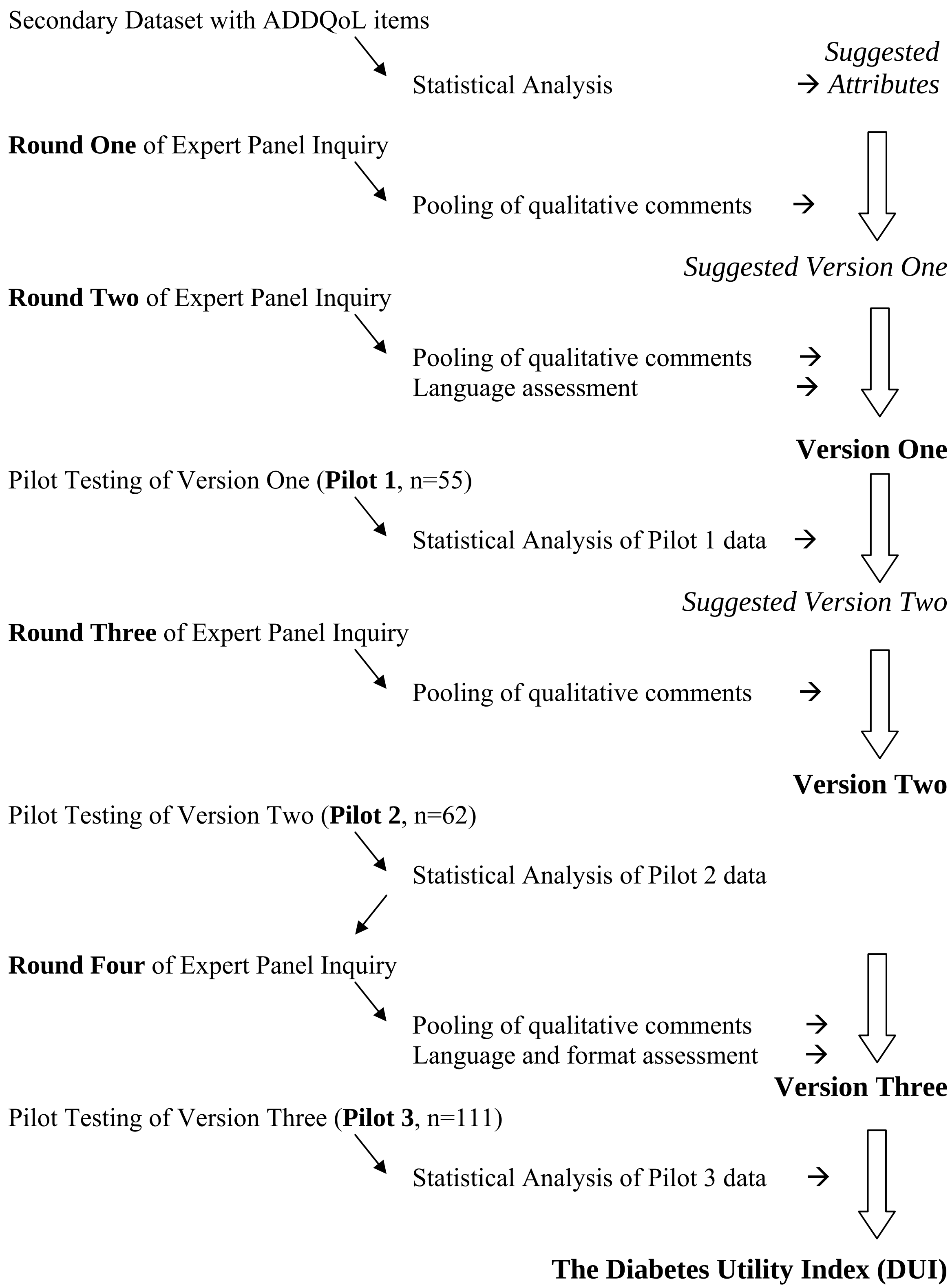
The answers to the structured qualitative input-gathering exercise in Round Two were pooled together to create Version One of the health state classification system (Appendix 4) which was then sent to an English expert to make any suitable changes in wording and grammar in line with the expectations from the expert panel in Round Two of inquiry. It was then pilot tested (Pilot 1, n=52) in the outpatient and community settings described in sections 3.1 and 3.2. Flyers were placed at prominent locations in these settings requesting patients with diabetes to pick up and answer the classification system along with which a cover letter signed by the Principal Investigator (PI) and Co-Investigator (Co-I) was attached. Responses were anonymous and no demographic information or any other kind of information (other than the venue) was collected. Respondents were requested to drop the answers in specially designated boxes.

Statistical analysis of the Pilot 1 responses to Version One informed the necessary modifications to be made leading to the development of another suggested version (Suggested Version Two, Appendix 5) of the classification system. In order to discuss the changes from the previous version and to explain the reasons thereof, the expert panel was contacted over the phone or in person. Given the extent of the prior involvement of the panel, this step in the process of receiving qualitative input was relatively faster, and led to the development of Version Two (Appendix 6) of the classification system that was pilot tested (Pilot 2, n=65) in a manner similar to the methodology adopted in Pilot 1.

Statistical analysis of the Pilot 2 responses informed the necessary modifications to be made to Version Two. Instead of making suggested changes and then submitting it the expert panel, the panel was consulted beforehand due to the nature of the modification suggested in this version (dropping of a redundant attribute and replacement with another, discussed in detail in the Results Chapter). The resultant classification system was reviewed by an English expert for content and by a graphics designer for layout and related formatting, leading to the development of Version Three that was pilot tested (Pilot 3, n=111). Due to the satisfactory results obtained from the analysis of Pilot 3 results, Version Three was declared the Final Version (Appendix 7) of the diabetes-specific health state classification system. In this final version, four attributes (Physical Ability & Energy Level, Relationships, Mood & Feelings, and Enjoyment of Diet) were designated with four severity levels each and the fifth attribute (Satisfaction with management of changes) was designated with three severity levels. The classification system was accommodated in two standard-sized pages, and titled the 'Diabetes Utility Index' (DUI).

Health states can be described from the DUI simply by choosing levels of severity on each DUI attribute. For example, '1,1,1,1,1' represents a Perfect Health state where each DUI attribute is at its best level, and '4,4,4,4,3' represents and all-worse health state where each DUI attribute is at its worst level. Health states such as these can then be incorporated into preference elicitation interviews, as discussed in the next section. The diabetes-specific classification system of the DUI can describe 768 unique health states in all (calculated by multiplying the number of levels of severity in each attribute, or 4x4x4x4x3).

Figure 2: Summary of the Steps Involved in the Development of the Final Version of the Diabetes-Specific Health State Classification System.



3.3.2 Objective Two: Phase One

3.3.2.1 Overview of the Multi-Attribute Utility Function

Phase One of Objective Two involved selecting and describing health states from the classification system designed in Objective One, based on the framework of the Multi-attribute Utility Theory (MAUT). The relative merits and demerits of the decomposed approach, using the MAUT, have been discussed in the previous Chapter. The decomposed approach was chosen as the health state modeling technique in this study due to a combination of factors. The technique requires relatively fewer valuations as compared to the composite approach (Dolan, 2002; Brazier, 2005). Another important consideration that went into the choice of method was that the necessary expertise to guide this research study using the MAUT approach was available. Finally, the decomposed approach enabled the study to be feasibly conducted, given the prevailing time and resource constraints for completing the study.

A general multiplicative Multi-Attribute Utility Function (MAUF) for the five DUI attributes can be given by:

$$u = [1 / k] \left[\prod_{j=1}^{5} (1 + k * k_j * u_j) - 1 \right] \quad \text{…. Equation (1)}$$

where:

1. the subscript 'j' refers to a DUI attribute, where j=1 through 5, and $\prod$ is a multiplication sign indicating the product of all $(1+ k * k_j * u_j)$ from j_1 to j_5;
2. 'u' is the cardinal utility score of a multi-attribute health state on the worst-healthy scale, a scale on which the utility score of the all-worst state (Pits state, where all the five DUI attributes at their worst level: 4,4,4,4,3) is 0.00 and the utility score of Perfect Health (all the five DUI attributes at their best level:1,1,1,1,1) is 1.00;

3. 'u_j' refers to the single-attribute utility function for attribute 'j', on a scale where the worst level of attribute 'j' has a utility score of 0.00 and the best level has a utility score of 1.00. Each level of an attribute has a value for 'u_j', representing the utility attached to being in a particular level on that attribute, and is obtained from the preference-elicitation interviews;
4. 'k_j' are constants, with one value for each attribute (k_1 through k_5), and are obtained from the preference-elicitation interviews. They represent the weight attached to that attribute and its importance in calculating the utility of the overall health state;
5. 'k' is a constant, a scaling parameter estimated from the data, which captures the preference interaction between attributes, and is calculated by solving Equation (2) below using the values of k_1 through k_5 $$1 + k = \prod_{j=1}^{5} (1 + k * k_j) \text{ …. Equation (2)}$$

The multiplicative functional form, as discussed earlier, specifies one type of preference interaction between all attributes: they are all either preference complements or preference substitutes.

If $\sum_{j=1}^{5} k_j > 1$, then $-1 < k < 0$ (Equation 3a) (i.e. 'k' is negative), and the attributes are all preference substitutes, i.e. the interaction between attributes is synergistic, and if $\sum_{j=1}^{5} k_j < 1$, then $k > 0$ (Equation 3b) (i.e. 'k' is positive), and the attributes are all preference complements, i.e. the interaction between attributes is antagonistic. However, if $\sum_{j=1}^{5} k_j = 1$ (Equation 3c), then $k = 0$. This can be considered as a special case of the multiplicative function, known as the linear additive function where there are no interactions in preferences between attributes. It follows that the additive model fails when 'k' is not equal to zero, and that other non-additive models should be considered.

In order to simplify the measurement task and to make calculations easier (Torrance et al., 1996; Furlong et al., 1998; Feeny et al., 2002), a 'disutility' (disutility = 1 minus utility) approach was considered. The resultant Multi-Attribute Disutility Function (MADUF) for the five DUI attributes can be specified as:

$$u` = [1 / c] [ \prod_{j=1}^{5} ( 1 + c * c_j * u`_j) – 1] \quad …. \text{Equation (4)}$$

where:

1. the subscript 'j' refers to a DUI attribute, where j=1 through 5, and ∏ is a multiplication sign indicating the product of all $(1+ c * c_j * u`_j)$ from j_1 to j_5;
2. 'u`' is the disutility score of a multi-attribute health state on the worst-healthy scale, a scale on which the disutility score of the all-worst state (Pits state, where all the five DUI attributes at their worst level: 4,4,4,4,3) is 1.00 and the disutility score of Perfect Health (all the five DUI attributes at their best level:1,1,1,1,1) is 0.00;
3. '$u`_j$' refers to the single-attribute disutility function for attribute 'j', on a scale where the worst level of attribute 'j' has a disutility score of 1.00 and the best level has a disutility score of 0.00. Each level of an attribute has a value for '$u`_j$', representing the disutility attached to being in a particular level on that particular attribute which is obtained from the preference-elicitation interviews;
4. 'c' and 'c_j' are scale parameters, the counterparts of 'k' and 'k_j' explained earlier such that:

$$1 + c = \prod_{j=1}^{5} (1 + c * c_j) \quad …. \text{Equation (5)}$$

The interpretation will differ in that an interaction that is synergistic on the utility scale will be antagonistic on the disutility scale and vice versa (Furlong et al., 1998). Thus a negative value of 'c' is indicative of attributes that complement each other.

The simplicity of the disutility approach results from the ease in the calculation of the c_j's, as is explained in the next paragraph.

3.3.2.2 Selection of Health States for the Preference Elicitation Interview

Phase Two was based on the approaches described by Feeny and colleagues (2002) in the development of the generic HUI-3, and Revicki and colleagues in the development of the asthma-specific ASUI (1998a) and the rhinitis-specific RSUI (1998b). The number of health states required to be specified to estimate a MAUF depends on the structure of the health state classification system of the instrument. Thus, for a hypothetical five-attribute classification system with three severity levels in each attribute, about 13 to 16 states in total will need to be described as illustrated in Figure 3. In accordance with these approaches, the following types of health states were selected:

1. *Single-attribute level states*

In a single-attribute level state, a given attribute is at less than full function and all other attributes are set at their best levels. For example, a given attribute with three levels would have two single-attribute level states other than the best level and Perfect Health (which is common for all attributes), such that each health state contains a varying level of the given attribute, and all other attributes within the classification system at their best level. Each such single-attribute level state was described holistically explaining the extent of loss of function in the attribute being measured (i.e. severity level of the attribute in question) as well as the best levels for the other attributes. Valuations of these single-attribute level states provided the data to calculate '$u`_j$' in the MADUF explained in Equation 4.

2. *Corner states*

Corner states can be considered to be a special kind of single-attribute state. These are health states where one attribute is at its lowest level and all other attributes are at their best level. Hence, there will be as many corner states for a health state classification system as there are attributes. For example, the DUI classification system with five attributes had five corner states. Their primary role is to provide values for 'c' and 'c_j'. The specification of the five corner states makes calculations easy because Equation 4 reduces to the form c_j = u` when it is solved for each of the corner states. In other words, c_j's are the disutility scores of the corner states on a scale where the all-worst state, Pits, has a disutility of 1.00 and Perfect Health has a disutility of 0.00. The five values of c_j (c_1, c_2, c_3, c_4, and c_5, i.e. one for each DUI attribute) then enable the calculation of the scaling parameter 'c'.

The nature of the description (where the one attribute is at the lowest level) of these states affords the use of the term 'disutility corner states'. These type of corner states are said to be easier for respondents to imagine and are easier to discriminate among as compared to corner states that are described with one attribute at the best level and all other attributes at their worst levels (Torrance et al., 1996; Furlong et al., 1998).

3. *Methodological marker states*

These are a set of multi-attribute states with attributes at different levels of severity. There is not a fixed number of marker states required to be used, but the states chosen should span a wide range of severity across attributes. Three such states were reportedly chosen in the development of the valuation function for the eight-attribute HUI-3 (Furlong et al., 1998; Feeny et al., 2002), and five were chosen for the five-attribute ASUI (Revicki et al., 1998a) and the RSUI (Revicki et al., 1998b). Marker states were not used in the calculation of MAUF parameters directly; their purpose was to provide the data to facilitate the conversion of values into utilities for the single-attribute level states and corner states for which only values, not utilities were obtained (a more detailed explanation is available in Section 3.3.3). A variety of marker states were tested, allowing the severity levels for the five DUI attributes to vary. Expert panel members rated the marker states on a VAS where 100 equals Perfect Health and Pits equals zero. Among the marker states tested, three easy-to-understand states were chosen such that on average:

- the first marker state was rated to be in the top one-third of the VAS (between 100 and 67 on the VAS - the mild marker state);
- the second marker state was rated to be in the middle one-third of the VAS (between 66 and 34 on the VAS – the moderate marker state); and
- the third marker state was rated to be in the bottom one-third of the VAS (between 33 and 0 on the VAS - the severe marker state).

4. *Scaling anchor states* - These were Dead, Perfect Health (all attributes at their best level), and Pits (all attributes at their worst level). Perfect Health was fixed as the best possible state with a value of 100 (i.e. a utility of 1.00).

Applying the above principles, the estimation of a MAUF for the DUI involved describing twenty health states in all, as illustrated in Figure 4, out of the 768 health states that could be described by the instrument. Fourteen single-attribute level states, which included five corner states, were required to be described in addition to the three scaling anchors of Dead, Perfect Health and Pits, and three Marker States. They were described in a manner that would enable respondents to conceptualize the health situations they represent. To do so, the relevant sentences representing the severity levels for the five attributes contained in the DUI were used. Since Dead is undefined in the

DUI (it is not one of the 768 health states that the DUI can describe), it was simply pictorially represented. More detail on the presentation format of these cards is provided in the next section (Section 3.3.3).

Figure 3: Representation of Health States to be Valued as per MAUT for a Hypothetical Classification System.[1]

	A	B	C	D	E	State	Group
1	A 1	B 1	C 1	D 1	E 1	**Perfect Health**[2]	
2	**A 2**	B 1	C 1	D 1	E 1		**Single-Attribute Level States**[4]
3	**A 3**	B 1	C 1	D 1	E 1	**Corner State**[3] **for A**	
4	A 1	**B 2**	C 1	D 1	E 1		
5	A 1	**B 3**	C 1	D 1	E 1	**Corner State for B**	
6	A 1	B 1	**C 2**	D 1	E 1		
7	A 1	B 1	**C 3**	D 1	E 1	**Corner State for C**	
8	A 1	B 1	C 1	**D 2**	E 1		
9	A 1	B 1	C 1	**D 3**	E 1	**Corner State for D**	
10	A 1	B 1	C 1	D 1	**E 2**		
11	A 1	B 1	C 1	D 1	**E 3**	**Corner State for E**	
12	A 3	B 3	C 3	D 3	E 3	**Pits**[5]	
13	A 2	B 2	C 1	D 1	E 1		**Marker States**[6]
14	A 2	B 1	C2	D1	E 1		
15	A 1	B 2	C 1	D 1	E2		
14	A 3	B 1	C 1	D 2	E 1		
15	A 1	B 1	C 3	D 2	E 3		
16	Undefined					**Dead**	

[1]The Hypothetical Health State Classification System consists of five attributes: A, B, C, D & E. Each attribute consists of three levels, where 1 is the best and 3 is the worst for that Attribute.

[2] Perfect Health is a health state where all attributes are at their best levels.

[3] Corner states are health states where one attribute is at the lowest level and all other attributes are at their best level.

[4] In a single-attribute level health state for a given attribute, that attribute is at less than full function and all other attributes are set at their best levels.

[5] Pits is the health state where all attributes are set at their lowest or most severe levels.

[6] Marker states are a set of multi-attribute states with attributes at different levels of severity, spanning a wide range of severity across attributes.

Figure 4: Representation of Health States Derived from the Diabetes Utility Index That Were Valued as per MAUT. [1]

1	P1	R1	M1	D1	S1	**Perfect Health**[2]	
2	P2	R1	M1	D1	S1		**Single-Attribute Level States**[4]
4	P4	R1	M1	D1	S1	**Corner State**[3] **for P**	
5	P1	R2	M1	D1	S1		
7	P1	R4	M1	D1	S1	**Corner State for R**	
8	P1	R1	M2	D1	S1		
10	P1	R1	M4	D1	S1	**Corner State for M**	
11	P1	R1	M1	D2	S1		
13	P1	R1	M1	D4	S1	**Corner State for D**	
14	P1	R1	M1	D1	S2		
15	P1	R1	M1	D1	S3	**Corner State for S**	
16	P4	R4	M4	D4	S3	**Pits**[5]	
17	P1	R1	M2	D2	S1	MA	**Marker States**[6]
18	P2	R2	M3	D3	S2	MB	
19	P3	R3	M4	D3	S2	MC	
20	Undefined					**Dead**	

[1] The DUI system consists of five attributes: Physical Ability & Energy Level (P), Relationships (R), Mood & Feelings (M), and Enjoyment of Diet (D) with four severity levels each and Satisfaction with management of changes (S) with three levels.

[2] Perfect Health is a health state where all attributes are at their best levels.

[3] Corner states are health states where one attribute is at the lowest level and all other attributes are at their best level.

[4] In a single-attribute level health state for a given attribute, that attribute is at less than full function and all other attributes are set at their best levels.

[5] Pits is the health state where all attributes are set at their lowest or most severe levels.

[6] Marker states (MA, MB, and MC) are a set of multi-attribute states with attributes at different levels of severity, spanning a wide range of severity across attributes.

3.3.3 Objective Two: Phase Two

Phase Two involved conducting one-on-one interviews to obtain patients' preferences for the health states selected and described in Phase One. The interviews used a combination of the VAS and SG techniques used by Feeny and colleagues (2002) and Revicki and colleagues (1998a; 1998b) in the development of valuation functions for the PBMH that were designed in their respective studies. The choice of the SG is necessitated since the MAUT specifies that preferences be obtained in a manner consistent with the expected utility theory. Interview participants were contacted based on their expressed interest (via their response to study-related flyers and handouts) to take part in these interviews, and appointments were scheduled to conduct the interviews either at their private residence or any available convenient, quiet and private meeting rooms in the Health Sciences Center of WVU.

The interviews commenced with the names and descriptions of the five DUI being verbally narrated to the respondent in the manner of a free-wheeling conversation. This provided an opportunity for the respondent to gain an understanding of the attributes to the satisfaction of the interviewer. Respondents were also shown five pictures, one each representing a DUI attribute and a color scheme in which one color each corresponded to a severity level. The color scheme was standardized across attributes (level three for the Satisfaction attribute, the last level for that attribute, was designated the color corresponding to the color used to represent level four on the other attributes). The description of the attributes and their representative pictures as well as the color scheme was color-printed and laminated on a handout that was available to the respondents at all times during the interview. The respondent then provided their answers to the DUI.

Respondents were then asked how many additional years they expected to live. Respondents were asked to keep this expected life-expectancy in mind in all the health state scenarios presented to them from that point onward. They were allowed to expect everything else in their lives to be just the same, barring the extent to which each attribute was affected, which would be in the manner contained within a health state description.

In the course of the interview, the following major measurement tasks, based on the those outlined in a detailed technical report on the valuations conducted for the HUI-3 (Furlong et al., 1998), were performed:

1. A VAS was established such that Perfect Health, the most desirable state, was assigned a value of 100 and the respondent choose either Pits or Dead as the least desirable state and placed it at 0;
2. On the appropriate 0-100 VAS established above, the respondent valued the methodological marker states and also valued either Pits or Dead (whichever was not chosen as the least desirable state);
3. On the same 0-100 VAS, the single-attribute level states and corner states were valued in sets. Each set represented single-attribute level states and corner states for one attribute in the classification;
4. The methodological marker states were valued using SG, on a Pits=0/ Perfect Health=1.00 scale; and
5. Depending on the choice of Pits or Dead as the least desirable state, SG measurements of Dead on the Pits=0.00/ Perfect Health=1.00 scale, and Pits on the Dead=0.00/ Perfect Health=1.00 scale were undertaken.

When administering these tasks, respondents were requested to focus only on the descriptions provided, and assume that other factors were all constant. Before wrapping up, demographic information and respondents' perceptions regarding the interview were also collected. The answers were recorded anonymously in a booklet (see Appendix 8).

In summary, the single-attribute level states and the corner states were valued in sets, in VAS tasks. Data for the Multi-Attribute Utility Function (MAUF) required to provide the valuation function of the health state classification comes mainly from the single-attribute level states and the corner states (Furlong et al., 1998; Feeny et al., 2002). The marker states (and the applicable anchors, Pits or Dead) were presented in both VAS as well as SG tasks. While Perfect Health was fixed as the best possible state with a utility of 1.00, Dead and Pits were valued using both VAS and SG tasks. The primary role of the marker states (and the applicable anchors, Pits or Dead) was to provide data for a value-to-utility conversion model, since both VAS value scores and SG utility scores were available only for these health states (Furlong et al., 1998; Feeny et al., 2002). Figure 5 provides an illustration of the type of valuation tasks that were performed for the health states that were described from the DUI.

Figure 5: Representation of the Valuation Tasks Performed as per MAUT for the Diabetes Utility Index.[1]

1	P1	R1	M1	D1	S1	**Perfect Health**[2]		
2	P2	R1	M1	D1	S1		VAS [7]	**Single-Attribute Level States**[4]
4	P4	R1	M1	D1	S1	**Corner State[3] for A**	VAS	
5	P1	R2	M1	D1	S1		VAS	
7	P1	R4	M1	D1	S1	**Corner State for B**	VAS	
8	P1	R1	M2	D1	S1		VAS	
10	P1	R1	M4	D1	S1	**Corner State for C**	VAS	
11	P1	R1	M1	D2	S1		VAS	
13	P1	R1	M1	D4	S1	**Corner State for D**	VAS	
14	P1	R1	M1	D1	S2		VAS	
15	P1	R1	M1	D1	S3	**Corner State for 3**	VAS	
16	P4	R4	M4	D4	S3	**Pits**[5]	VAS & SG [8]	
17	P1	R1	M2	D2	S1	MA	VAS & SG	**Marker States**[6]
18	P2	R2	M3	D3	S2	MB	VAS & SG	
19	P3	R3	M4	D3	S2	MC	VAS & SG	
20	Undefined					**Dead**	VAS & SG	

[1] The DUI system consists of five attributes: Physical Ability & Energy Level (P), Relationships (R), Mood & Feelings (M), and Enjoyment of Diet (D) with four severity levels each and Satisfaction with management of changes (S) with three levels.
[2] Perfect Health is a health state where all attributes are at their best levels.
[3] Corner states are health states where one attribute is at the lowest level and all other attributes are at their best level.
[4] In a single-attribute level health state for a given attribute, that attribute is at less than full function and all other attributes are set at their best levels.
[5] Pits is the health state where all attributes are set at their lowest or most severe levels.
[6] Marker states (MA, MC, and MC) are a set of multi-attribute states with attributes at different levels of severity, spanning a wide range of severity across attributes.
[7] VAS=Visual Analog Scale task, administered using the Feeling Thermometer prop.
[8] SG=Standard Gamble task, administered using the Chance Board prop.

In the estimation of MAUFs, either an individual-level approach or a person-mean level approach can be chosen. The former involves calculation of a multi-attribute function for each respondent. This study employed a Person-Mean approach, i.e. the mean of all respondents' values and utilities for each health state were calculated (details regarding the calculations involved are discussed in detail in the Data Analysis section) before estimating an MAUF. This approach permits two strategies in allotting health states to respondents: 1) all the health states (corner states, single-attribute level states, marker states and anchor states) be required to be completed by each respondent; or 2) health states be randomly assigned – one way to accomplish this is by assigning corner states and single-attribute level states for two or more attributes to one respondent, and so on. While the former approach was selected, the suggestion by Torrance and colleagues (2001) to limit the number of health states on a single VAS task to about six, and to value health states in sets, one set at a time, was also considered.

The instrumentation and props for conducting these valuation tasks were based on those developed and used by the HUI group and illustrated in detail in the technical reports published by this group (Furlong et al., 1990; Furlong et al., 1998). For the VAS tasks in this study, the Feeling Thermometer prop was used, while for the SG tasks, the Chance Board prop was used. Both were designed employing suggestions from the HUI manuals (Furlong et al., 1990) with appropriate revisions to the color scheme and materials as necessary. The health states were printed in color on cards, including the severity level-statements (borrowed from the DUI) along with a representative picture for each attribute, and then laminated. Production of the props was contracted to the Department of Graphic Design at WVU.

3.3.4 Objective Two: Phase Three

The methodological Marker States and Dead (or Pits, depending on which state was considered the worst) were valued using both VAS and SG. This provided the necessary data to estimate a model for converting VAS scores for health states that were not valued using SG (single-attribute level states and corner states), into SG utility scores. The VAS valuations of the single-attribute level states and corner states, when converted into utilities using the conversion model, provided the measurements necessary to estimate a multiplicative Multi-attribute Utility Function (MAUF) (Furlong et al., 1998; Feeny et al., 2002) for the DUI. The purpose of the MAUF is to convert individuals' responses to the DUI questions into a score that estimates the utility that the individual places on his or her life. In other words, the MAUF was intended to provide utility estimates for each of the 768 health states that can be described by the diabetes-specific health state classification system of the DUI without having to individually perform preference-elicitation tasks for each of these health states. These calculations are explained in greater detail in the Data Analysis section.

3.3.5 Objective Three

As mentioned in section 1.3.1, it is important to know what causes the variation in utilities – whether it is due to real differences between people or if it is due to construct-irrelevant variation. However, like with other humanistic constructs, a majority of the variation in utilities may go unexplained in research. Lenert and colleagues (2000) suggest that researchers may be more confident in the ability of utilities to reflect actual patient preferences if the variation in utilities can be explained by demographic, clinical and QoL-related factors. Demographic and clinical variables were therefore obtained for the survey respondents in order to validate the utilities estimated by the MAUF of the DUI, as well as to account for the variation in utilities. In addition, other PRO questionnaires were also included that will aid in the validation of the classification system and MAUF of the DUI.

The Diabetes Symptoms Checklist Revised (DSC-R) (Grootenhuis, Snoek, Heine, & Bouter, 1994) was designed to assess the frequency and severity of diabetes-associated symptoms that can impact both physical and psychological aspects of health. Symptoms are grouped into the following subscales comprising 34 items in all: cardiovascular, hyperglycemic, hypoglycemic, neurologic, psychological, and visual. A higher score on the instrument is indicative of greater burden of diabetes-related symptoms, and could be expected to be associated with the more severe levels on the 'Physical Ability and Energy Level' and 'Moods & Feelings' attributes of the DUI. The instrument was hence included in order to compare its scores against the severity classification and the utilities associated with the 'Physical Ability and Energy Level' and 'Moods & Feelings' attributes of the DUI, as well as the overall utility reported by the DUI.

The Well-Being Questionnaire (WB-Q) was originally developed to provide a measure of depressed mood, anxiety, and various aspects of positive well-being to evaluate new diabetes treatments. The instrument was further developed into the 12-item version of the WB-Q (WB-Q12), offering a balanced selection of positive and negative items and providing negative well-being, positive well-being, and energy sub-scales that can be totaled for a total general well-being score (Bradley, 2000; Riazi, Bradley, Barendse, & Ishii, 2006). Although originally developed and validated in diabetes populations (Bradley, 1994), the item content of the WB-Q12 is not diabetes-specific and is hence not intended to be used in diabetes populations alone. A higher score on the instrument is indicative of greater general well-being, and can be expected to be associated with the less severe levels on the 'Moods & Feelings' attribute of the DUI. The instrument was hence included in order to compare its scores against the severity classification and the utilities associated with the 'Moods & Feelings' attribute of the DUI.

The Diabetes Empowerment Scale (DES) was developed to measure the psychosocial self-efficacy of people with diabetes via its three measured subscales: managing the psychosocial aspects of diabetes, assessing dissatisfaction and readiness to change, and setting and achieving goals (Anderson, Funnell, Fitzgerald, & Marrero, 2000). A brief assessment of overall self-efficacy encompassing these concepts can also be made using the validated eight item short form of the DES (DES-SF) (Anderson, Fitzgerald, Gruppen, Funnell, & Oh, 2003). A higher score on the instrument is indicative of greater self-efficacy, which can be expected to result in a greater sense of satisfaction that people with diabetes experience in relation to what needs to be done to manage the condition on a daily basis. The instrument was hence included in order to compare its scores against the severity classification and the utilities associated with the 'Satisfaction with management of changes' attribute of the DUI.

The EuroQol EQ-5D (EQ-5D) is a generic PBMH consisting of five dimensions: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression (Brooks, 1996; Kind, 1996). An overall preference-weighted summary score or utility, represented as the $EQ\text{-}5D_{index}$, is obtained from the pattern of responses to the EQ-5D questions by using previously derived valuation functions, that essentially 'weight' the answers to the EQ-5D questions using social preferences obtained empirically in tested populations (Dolan, 1997; Shaw et al., 2005). A greater $EQ\text{-}5D_{index}$ score indicates greater utility or better HRQoL. Since the EQ-5D is a widely used generic PBMH, the $EQ\text{-}5D_{index}$ scores can be directly compared to the utilities obtained using the MAUF of the DUI.

The Medical Outcomes Study Short-Form 12 (SF-12) is a widely used 12-item measure covering eight domains of health status or HRQoL (Ware, Jr., Kosinski, & Keller, 1996). The SF-12 yields two scores: the Physical Component Score -12 (PCS-12, or physical health status), and the Mental Component Score (MCS-12, or mental health status) (Ware, Jr., Kosinski, Turner-Bowker, & Gandek, 2002). The SF-12 was used to generate the classification system for a generic PBMH called the SF-6D (Brazier et al., 2004b) that consists of six attributes including physical functioning, role limitation, social functioning, pain, energy, and mental health. Because of this relationship between the two instruments, SF-6D scores, which are preference-based or utility scores, can be directly obtained from responses to the SF-12 itself (Brazier et al., 2004b), and can then be compared to the utilities obtained using the MAUF of the DUI.

Additionally, responses to the SF-12 have also been used to predict utility scores obtained from the HUI-3, another widely used generic PBMH (Sengupta, Nichol, Wu, & Globe, 2004). Since the SF-12 is not directly related to the HUI-3 like it is to the SF-6D, the predicted HUI-3 scores can be considered as 'derived' HUI-3 scores, in order to distinguish them from the utility scores that are obtained by the direct use of the instrument. Nevertheless, using the SF-12 thus allows the calculation of health status scores, the PCS-12 and MCS-12 in addition to the SF-6D utility scores and derived HUI-3 utility scores, while at the same time minimizing the survey booklet-induced respondent burden since the SF-6D and the HUI-3 do not have to be separately included.

At the time interviews were being conducted in Phase Two of Objective Two, there was a concurrent survey mail-out containing:

1. a personalized, signed cover letter from the patient's specific physician on an OMSA letterhead, explaining the purpose of the study and providing the contact information of the research investigators of the current study, and the co-investigators at OMSA;
2. a battery of PRO instruments arranged in a booklet including the DUI, SF-12, DSC-R, EQ-5D, WB-Q12, and the DES-SF ; and
3. a postage-paid return envelope.

Follow-up reminder post cards were sent two weeks after the commencement of the mailing. The above materials were sent a second time to patients who responded to the reminder cards and requested an additional copy.

The MAUF developed in Phase Three of Objective Two was used to obtain the survey respondents' utilities for their current health conditions without having to conduct lengthy preference elicitation interviews. These utilities were merged with a collection of retrospective clinical, medical history and demographic data in order to examine the nature of associations between these variables, as stated earlier. A set of three different databases, linked by patients' medical record numbers (MRN) were used to obtain the necessary clinical and medical history information: 1) The Diabetes Clinic Database can be described as a lab-value database consisting of clinical lab values of patients who consult physicians at the WVUDI clinics for the management of their diabetes; 2) The University Health Associates (UHA) database derives its information from the IDX billing system, which is a software package that tracks patient office visit appointments and billings. After receiving care from a provider, patients receive a fee slip that is tracked by the IDX system within two days of the visit. In addition to the charges, the fee slip also consists of diagnosis information that is later coded into ICD-9 CM format by the coding department; 3) The WVU Hospital database provides information on hospital and ER encounters. It tracks the bills that patients receive upon discharge from the hospital, and among others, it contains ICD-9 CM diagnosis codes associated with each hospital and ER visit.

The co-investigators at OMSA coordinated the collection process of patient responses with the clinical and medical history data to be used in the study using respondents' MRN. In order to de-identify patient information, a unique identification number was assigned to each survey respondent.

The following is a summary of the data that was collected in Objective Three:

1. Self-reported data:
   a. Demographic information - including age in years, gender, education, race, marital status, and annual income;
   b. Duration of diabetes in years ;
   c. Diabetes type;

d. Insulin use; and

e. Height (in feet) and weight (in pounds) – this was used to calculate BMIs using the formula (CDC, 2006) : BMI = [Weight in pounds / (height in inches)^2] * 703.

f. Number of physician office visits, ER visits and hospitalizations related to diabetes in the past one year

2. Data obtained from electronic records:

a. A1C - An A1C test result reflects the average blood glucose level for the previous 2-3 months, in comparison to blood glucose testing which identifies control on a day-to-day basis. Not only can A1C allow healthcare professionals to judge how well a patient's diabetes treatment plan is working, it can also assist in evaluating the long-term effects of diabetes management. In order to assess the long-term impact of glycemic control on utility, the average of A1Cs in a 12-month period prior to the survey was recorded for each patient. The A1C results were made available from the Diabetes Clinic Database;

b. ICD-9 CM Diagnosis codes - Diagnosis codes for office visits, emergency room visits and hospital admissions, made available from the UHA database and the WVU Hospital database, were used to calculate the following:

i. Index of Co-morbidity - A Charlson Co-morbidity Index based on ICD-9 CM codes for medical claims in a 12-month period prior to the survey was used to measure co-morbidity for each respondent. The required ICD-9 CM codes were made available by the OMSA. The Charlson Index consists of a list of 19 medical conditions, with each condition being assigned a weight from one to six. This weight was derived from relative risk estimates of proportional hazard regression models using clinical data

(Charlson, Pompei, Ales, & MacKenzie, 1987). The Charlson index for an individual is the sum of weights for all prevalent conditions (among the list of 19 conditions) during a specified time period. A version of the index by D'Hoore and colleagues (d'Hoore, Bouckaert, & Tilquin, 1996), one of the few Charlson indices for use with administrative claims data, was used in this study. This version is based on only the first three digits of ICD-9 CM codes. Diagnosis codes related to diabetes will be excluded in the computation of this index; and

ii. Index of diabetes complications - A diabetes complication score based on ICD-9 CM codes for medical claims in a 12-month period prior to the survey was calculated for each respondent. The required ICD-9 CM codes were made available by the OMSA. Specific ICD-9 CM diagnosis codes under the series 250.xx have been designated for conditions related to diabetes and its associated complications. Since the 250.xx series of codes are not always the only ones used to code for conditions that result from complications arising due to diabetes, the risk of underestimating the burden due to diabetes-related complications by using only the 250.xx series cannot be ruled out. In order to address this issue, a list of ICD-9 diagnosis codes for diabetes-related complications was consolidated using a set of ICD-9 CM codes compiled by the Endocrine Society (Dickey, 2005) as well as by identifying other ICD-9 CM codes, not included in this list, that indicate a condition potentially arising due to a complication directly related to diabetes, through consultations with an Endocrinologist practicing at the WVU Hospitals. The diabetes complications score was finally calculated as a sum of the number of diabetes-related long-term complications (renal, ophthalmic, neurological, cardiac, cerebrovascular, and peripheral circulatory disorders, and depression) as well as short-term complications (hypoglycemia, diabetic coma, hyperosmolar coma, and ketoacidosis). Thus, the diabetes complications score can take values from zero to 11.

3.4 Data Analysis

3.4.1 Objective One

3.4.1.1 Classical Methods

The impact ratings were first subject to item analysis using standard statistical procedures based on Classical Test Theory (CTT). The classical index of discrimination, the item-total correlation coefficients (r) for an item's impact score with the total average impact score (Nunnally et al., 1994), was obtained for each item. Those items with an r <0.4 were marked for exclusion (Streiner & Norman, 1989). The remaining items were subjected to factor analysis using the SPSS Version 15 software package (SPSS, 2007). Exploratory factor analysis (EFA) attempts to explain the number of factors that are required to explain the relations among a set of items (Fayers et al., 2000), and is used when the factor structure of an instrument is unknown. Confirmatory factor analysis (CFA), on the other hand, is used to test the number of factors purported to be underlying the relations among a set of items in an instrument (Fayers et al., 2000).

In previous research using the ADDQoL, it was observed that very few respondents endorsed impact ratings of +2 and +3 (indicating increasing amounts of positive impact of diabetes on QoL), across the 18 ADDQoL items. Hence, the impact ratings of +2 and +3 were replaced with a rating of +1. This is similar to the approach adopted in the newer version of the ADDQoL (not published), in that it uses a 5-point impact rating scale i.e. -3 (through zero) to +1. While the importance weights remained unchanged, the weighted impact scores were now expected in the range of -9 (through zero) to +3. The un-weighted impact ratings were used to perform the factor analysis that further informed the process of gathering input for the classification system from the items contained in the ADDQoL, for the following reasons:

One, a single-factor structure for the ADDQoL has been established on the basis of a combination of the results of factor analysis and reliability testing using internal consistency reliability (Bradley et al., 1999; Bradley et al., 2002) on the weighted impact ADDQoL scores. This was confirmed by Sundaram (2005); who, however, reported using CFA that the single factor (obtained from a forced one-factor solution) explained only 45.63% of the variance from the 18 ADDQoL items. Factor loadings for each item on that single unrotated factor showed that 15 items loaded greater than 0.5 on that factor, with factor loadings for the other three items being 0.492, 0.481, and 0.375, respectively. Since the purpose of this study was to explore the creation of a brief diabetes-specific health state classification system based on the content of the ADDQoL, it was necessary to further explore the dimensionality among the items of the instrument. An alternate approach to assessing dimensionality to the ones already described above would involve using un-weighted impact ratings, using the modified rating scale, in place of weighted impact scores; and

Two, the proposed diabetes-specific health state classification system was being designed as a PBMH, with intentions to also calculate a different set of weights for the items included in it – aggregated preferences of a diabetes population to be on different levels of the items of the classification system. Hence, it was decided to explore the ADDQoL items without applying any other kind of weighting system.

Two tests were initially conducted in order to assess the adequacy of the data for conducting factor analysis. In the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy, high values (close to 1.0) generally indicate that a factor analysis may be useful with the data. The second test is the Bartlett test of sphericity - very small values (less than 0.05) indicate that significant relationships among variables probably exist. Exploratory Factor Analysis (EFA), employing Principal Axis Factor (PAF) extraction and Promax rotation, were performed on the items other than those marked for deletion. Promax rotation is a type of Oblique or non-Orthogonal rotation which is employed when the underlying latent factors are expected to correlate somewhat with one another (DeVellis, 2003).

A broad overview on the prevailing dimensionality of the ADDQoL items was obtained by examining the following: the scree plot, the total percentage variance explained by the factors, and the loading of items or subscales on the factors emerging from the rotated solution. This enabled the estimation of the number of factors that adequately and parsimoniously capture the original information contained within the ADDQoL items as well as making an initial observation on the constitution of those factors in terms of individual ADDQoL items. A secondary reduction was then performed by deleting those items with communalities < 0.3, those items showing its highest factor loading on the main factor to be lower than 0.4, and those items with similar loadings on different factors (such that the differences in loadings across factors is ≤ 0.1) (Prieto et al., 2003). This second step helped identify the major underlying factor that each item contributed in explaining, so that any one factor could be described in terms of its constituent items. The interpretation of the results of these analyses were presented to the clinical expert panel, who provided subjective input on items and reviewed pilot version of the diabetes-specific classification system, as outlined earlier.

3.4.1.2 Rasch Analysis

Rasch analysis was employed in order to provide guidance in the creation of broad attributes and in the construction of severity levels for each attribute of the diabetes health state classification system. Fit of the data from the resultant classification system to the Rasch model would impart desirable measurement characteristics to the instrument: uni-dimensionality, interval-level measurement, additivity and sample-free measurement. After the initial determination of attributes using CTT (factor analysis) and qualitative input from experts, Rasch Analysis was employed on data obtained from the pilot testing of three versions of the classification system (n1=52, n2=65, n3=111) using a software package called Winsteps (Linacre, 2007). As a confirmatory measure, the technique was also employed on data obtained from the larger validation survey (n=396) that included the classification system that was finalized after three rounds of pilot inquiry.

Rasch Analysis also has applications in determining dimensionality in data, which means that it could be used in place of, or along with, Factor Analysis based on CTT. While Rasch and factor analysis produce similar results, the two have been reported to suggest different factors in case the factor structure is vague. Since there is relatively greater uniformity in the interpretation of the CTT Factor Analysis, it was the preferred technique in the assessment of dimensionality. Rasch Analysis, on the other hand has been widely acclaimed as a valuable tool in developing items and response scales for instruments, and was the preferred technique to provide guidance in the development of the diabetes health state classification system.

Statistics generated by Rasch analysis was used to determine the extent to which individual respondents and each of the items on the classification system fit the expectations of the Rasch measurement model (Wright & Masters, 1982). The following tests were conducted:

i. Chi-square fit statistics were used to determine how well each item contributed to the measurement of the same underlying construct. The most commonly used Chi-square tests are known as Outlier-sensitive Fit (Oufit) and Information-weighted Fit (Infit), and are reported as Mean-Squares (MNSQ), that is, the chi-square statistics divided by their degrees of freedom (Wright et al., 1982). Item Outfit or Infit MNSQ values of about 1 are ideal by Rasch model specifications, while items with Outfit or Infit MNSQ values greater than 1.3 are usually diagnosed as potential misfits to Rasch model conditions and considered for deletion (Smith, Schumacker, & Bush, 1998). Wright and Linacre (1994) have recommended that standardized Infit and Outfit mean square values greater than 1.4 be used as the criterion to define those items and respondents that fail to demonstrate acceptable goodness-of-fit with the measurement model. Fit statistics exceeding these values for individual subjects indicate that the subjects failed to respond to the scale items in a manner consistent with the measurement model. On the basis of the summarized recommendations in the Winsteps manual (Linacre, 2007):

a. Items were considered adequately fitting the Rasch model if the Infit MNSQ was in the range 0.8-1.2 and the Outfit MNSQ was in the range 0.6-1.4;
b. Items were considered to be overfitting if the Infit and Outfit MNSQs were less than 0.6 and 0.8 respectively; and

c. Items were considered to be misfitting if the Infit and Outfit MNSQs were greater than 1.2 and 1.4 respectively.

2. Several Rasch Analysis criteria were used to assess the adequacy of the performance of the severity levels employed for the attributes of the diabetes health state classification system (Linacre, 1999):

a. Average measures for each rating category were required to advance evenly up the severity level, so that higher categories indicate greater amount of the latent variable, .i.e. severity on an attribute;
b. Average measures are reasonably close to their expected values, which are the average measures predicted in case of fit of the data to the Rasch model;
c. Step calibrations were required to advance in an orderly manner such that the distance between adjacent categories was between 1.5 logits and 5.0 logits, indicating categories that are neither too narrow a segment nor too broad a segment of the latent variable, i.e. severity on an attribute; and
d. Outfit MNSQs for category levels were required to be < 2.0, indicating less randomness and unexpected use of category levels that could lead to a non-performing measurement system where responses appear more similar.

Rasch item difficulty statistics (in logits) were obtained to examine the efficiency of measurement of the impact of diabetes by the DUI attributes, with a greater spread in difficult estimates of attributes indicating a wider measurement range. The Item-Separation Index was examined as a range of item difficulty, with larger values indicating a greater item spread. Item-Reliability, a Rasch estimate that is independent of the length of the instrument was also obtained. Rasch Person-Reliability of the DUI, analogous to

Cronbach's alpha, was calculated along with the Person-Separation Index, an indicator of the extent to which the DUI attributes could distinguish between different levels of impact of diabetes.

3.4.3 Objective Two: Phase Two

After the diabetes health state classification system was finalized at the end of three rounds of piloting and Rasch analysis in Objective One, select health states were described for use in preference elicitation interviews in Phase One of Objective Two, as discussed earlier. The analysis plan for the data obtained in Phase Two was based on the strategy outlined in the technical report of the valuation function of the HUI-3 (Furlong et al., 1998):

1) The respondents were classified into groups depending on their choice of state as least desirable in the first VAS task (see section 3.3.3): those reporting Pits as least desirable belong to Group A, while those reporting Dead as least desirable belong to Group B;

2) The preference measurements (value and utility scores) were summarized using the 'trimmed mean' measure of central tendency. A trimmed mean is calculated by discarding a certain percentage of the lowest and the highest scores and then computing the mean of the remaining scores. It is used in order to minimize the effects of outlier scores and is less susceptible to the effects of extreme scores than is the arithmetic mean. In this study, a 10% trimmed means (5% trimmed off of each end of the distribution) were used for all states valued on the VAS. Wherever stated, a Person-Mean score will denote the trimmed mean for a health state aggregated across respondents, and Person-

Mean(A) and Person-Mean(B) scores will denote the aggregated trimmed means for Groups A and B, respectively. Person-Mean(A) and Person-Mean(B) value and utility scores were calculated for the methodological marker states, the corner states, the single-attribute level states and the scaling anchors Pits and Dead;

3) An End-of-scale bias adjustment (EOSBA) was applied to Person-Mean(A) and Person-Mean(B) trimmed mean value scores to correct for end-of-scale bias (EOSB). This type of bias, also referred to as end-aversion bias or central tendency bias, refers to the tendency of respondents to avoid using extreme ends of the rating scales like the VAS, as reported in an earlier section. It was corrected for in this study by using an EOSBA factor of 1.78. The adjustment was applied in sets, only if the value score of a state was greater than 75 (on a scale of 100), and for the second-ranked state (just below Perfect Health, a conceptual anchor placed at 100). In the present study, health states were valued on the VAS in sets – first involving the anchor and marker states, followed by separate sets for each attribute's single-attribute level states and corner states, with the previous set of health states removed from the VAS FT board before a new set was introduced.

The EOSBA was applied in the following manner for each set: a) If the difference between the score for the highest ranked state (Perfect Health, anchored at 100) and the Person-Mean score for the second-ranked state was 10, then this difference was divided by 1.78 to obtain an EOSBA difference of 5.6 units; b) The EOSBA difference was subtracted from the highest score, i.e. 100, to obtain the EOSBA for the Person-Mean score of the second-ranked state at 100 - 5.6 = 94.4; c) The Person-Mean values of all other health states valued on the same set were rescaled using positive linear transformations, so that the relative distances between health states were maintained;

4) The VAS value ratings and SG utility scores obtained for the methodological marker states were utilized to estimate value-to-utility conversion models using a regression technique. Two such models were estimated, one each using EOSB-corrected Person-Mean(A) and Person-Mean(B) scores, respectively. Furlong and colleagues (1998) report from their prior work that a simple power function fared as good as or better than regressions involving transformations, non-linear techniques, and spline functions. A review of conversion models reported in the literature noted that while Torrance's (1976a) power transformation was replicable (Stiggelbout et al., 1996), researchers have also reported simple linear relationships between VAS ratings and SG utilities (Torrance et al., 2001). The present study utilized the best statistical approach possible for the data collected to obtain a study-specific conversion;

5) The value-to-utility conversions models were used to calculate Person-Mean(A) and Person-Mean(B) utility scores for the single-attribute level states and corner states (for which only VAS value scores are previously available);

6) Using positive linear transformations, the Person-Mean(B) utility scores were rescaled from a Dead=0.00/Perfect Health=1.00 scale to a Pits=0.00/Perfect Health=1.00 scale for the single attribute-level states and corner states; and

7) The above transformation facilitated the calculation of overall Person-Mean utility scores using Person-Mean(A) and Person-Mean(B) utility scores for the single-attribute level states and corner states using a weighted approach. The weights for Groups A and B were the respective prevalence proportion of each group. The Person-

Mean(A) and Person-Mean(B) utility scores were multiplied by their respective weights to yield overall Person-Mean utility scores.

3.4.3 Objective Two: Phase Three

The overall Person-Mean values (henceforth referred to as Person-Mean values) were used to fit a MAUF based on a strategy outlined in the technical report for the valuation function of the HUI-3 (Furlong et al., 1998), with the chosen approach of describing 'disutility' corner states.

1) The overall Person-Mean utility scores were converted to disutility scores, and the disutility scores for the corner states were designated as c_j's (where j=1 through 5);

2) The scaling parameter 'c' was calculated by iteratively solving the equation:

$$1 + c = \prod_{j=1}^{n} (1 + c * c_j)$$

where,

n(=5) is the number of attributes;

$\prod_{j=1}^{n}$ is the product of all $(1 + c * c_j)$ from c_1 to c_n ; and

c_j is the Person-Mean disutility for the corner state.

3) Using positive linear transformation, the Person-Mean single-attribute level utility scores for each attribute were converted from a Pits=0.00/ Perfect Health=1.00 scale into a scale where the lowest level of attribute=0.00 and the highest level of the attribute=1.00. On the transformed scale, the highest level was Perfect Health with a utility=1.00 and was fixed;

4) The Person-Mean single-attribute level utility scores for each attribute were converted to disutility scores (on a scale where the disutility of Perfect Health=0.00 and

the disutility of the lowest level of attribute=1.00) that were designated as $u`_j$'s (where j=1 through 5);

5) Fit to the additive model was then tested using the following condition:

If $\sum_{j=1}^{5} c_j = 1$, then c = 0 and the additive model holds.

The strategy employed was that if the additive model did not hold, the multiplicative model would be fitted;

6) The general multiplicative MADUF on a Pits=1.00/ Perfect Health=0.00 scale is given as:

$$u` = [1 / c] [ \prod_{j=1}^{n} ( 1 + c * c_j * u`_j) - 1]$$

where (as indicated previously in Equation 4),
n (=5) is the number of DUI attributes,
u` is the disutility of the health state on the Perfect Health =0.00 / Pits =1.00 scale,
c_j is the Person-Mean disutility for the corner state obtained in (1),
c is a scaling parameter obtained in (2),
$u`_j$ is the single-attribute level disutility score on the lowest level=1.00 /Perfect Health=0.00 scale obtained in (3).

3.4.3 Objective Three

An instrument is said to be valid to the extent that it measures what it purports to measure. The DUI was tested for its Concurrent Validity as well as Construct Validity.

Concurrent validity is demonstrated when a test correlates well with a measure that has previously been validated. In the context of the DUI, the two measures may be for the same construct (diabetes-specific utility), or for different, but presumably related (diabetes-specific utility and generic utility) constructs. In order to test the concurrent validity of the DUI utilities, the correlation between utility scores on the DUI, EQ-5D, SF-6D and HUI-3 (derived) was assessed.

Construct validation of an instrument or measure is concerned with validity of inferences made by that instrument about unobserved variables (the constructs) on the basis of observed variables (their presumed indicators) (Pedhazur & Schmelkin, 1991). One way to assess construct validity is to identify probable high and low scoring groups on an instrument on logical grounds a priori, and then demonstrate that certain groups obtain higher scores on that instrument than other groups (Pedhazur et al., 1991). In the present study, construct validity of the DUI utilities were examined by investigating differences in utilities within groups based on external indicators such as the extent of diabetes-related complications, and insulin use, using ANOVAs and t-tests.

The severity levels of the DUI classification system were assessed by studying the differences in scores on other relevant measures across these levels, using one-way ANOVAs: physical health status (using the SF-12) was compared between the levels on the 'Physical Health and Energy' attribute, mental health status and well-being (using the SF-12 and the W-BQ12 respectively) were compared between the levels on the 'Mood and Feelings' attribute well-being, while DES-SF (hence forth referred to as DES) scores were compared against the levels on the Satisfaction attribute of the DUI. Similarly, the total and subscales scores of the DSC-R were compared: the total symptom score, the neurological subscale score and the cardiovascular subscale score were compared between the levels on the 'Physical Health and Energy' attribute; and the psychological subscale score was compared between the levels on the 'Mood and Feelings' attribute. Finally, the correlations between single-attribute scores of the five DUI attributes and relevant measures listed above were also examined.

Univariate and multivariate statistics were used in order to identify the demographic, clinical, and medical history variables significantly influencing utility. Hierarchical regression models were built in order to explain the relationship between utility and the various demographic, clinical, and medical history predictors. It is used in explanatory situations when there is some basic knowledge on the relationships between at least some of the variables being used. In hierarchical regression, the order of entry of variables is assigned by the researcher according to logical or theoretical considerations (Tabachnick & Fidell, 2001). One point of view is to accord early entry to variables of greater theoretical importance. The opposite view is to introduce the relatively unimportant variables first, and then test the added prediction obtained by the introducing the remaining variables in the model. In this study, covariates were entered sequentially into the model in the following order:

Block 1: A1C, Insulin use, Index of diabetes-related complications, Duration of diabetes, Diabetes symptom burden (DSC-R total score);

Block 2: BMI, Index of co-morbidity; and

Block 3: Demographic variables: Age, Gender, Race, Education, Marital status, Income.

The regression model was checked for violations of assumptions: multi-collinearity, heteroscedasticity, autocorrelation, and non-normality of residuals. The list of variables that were used in the univariate and multivariate analyses, including information on the source, type, and categorization for analysis, is presented in Table 3.

Table 3: Description of Variables Used in Univariate and Multivariate Analyses

Category	Variable Name	Source of Information	Variable Type	Categorization
Demographic	Age (years)	Self-report	Continuous	N/A
	Gender	Self-report	Categorical	Male, Female
	Race	Self-report	Categorical	White, Not White
	Education	Self-report	Categorical	Less than college, College degree or more
Clinical	A1C	EMR	Continuous/ Categorical	A1C < 7.0 (in control), A1C ≥ 7.0 (not in control)
	BMI	EMR	Categorical	BMI < 30 (not obese), BMI ≥ 30 (obese)
Medical History	Insulin Use	Self-report	Categorical	Insulin user, Insulin non-user
	Duration of diabetes (years)	Self-report	Continuous	N/A
	Index of Co-morbidity (Charlson score)	EMR	Continuous	N/A
	Index of diabetes-related complications	EMR	Continuous/ Categorical	No complications, one complication, two or more complications
	Diabetes Symptom burden (DSC-R total score)	Calculated based on responses from Self-report	Continuous	N/A

3.5 Sample Size Calculations and Power

Power is the likelihood that a study will detect a true effect of a given magnitude if it actually exists (a true positive). Power can also be broadly termed as the probability that a statistical significance test will reject the null hypothesis for a specified value of an alternative hypothesis. Hence, power of a study is also the probability of avoiding a beta error, in which we fail to reject a null hypothesis that is false. It is important to consider power in research designs because studies with low power may be inconclusive.

A priori power analyses are conducted in order to ascertain the sample size required to perform the analyses necessary for a study at a level of power desired prior to the start of the study. For this purpose, it is necessary to decide upon the alpha level (the probability of making an alpha or Type 1 error) for the analysis, the desired power (1 minus the probability of making a beta or Type 2 error), and the effect size. Effect size can be conceived of as a measure of the distance between the null hypothesis and the alternate hypothesis. Hence, effect size refers to the underlying population rather than a specific sample. Although desirable to specify an effect size in research that has clinical implication, it is typical to decide the degree of deviation from the null hypothesis that is large enough to be clinically relevant.

In order to perform power calculations for the analyses in the study, a software package called G-Power (Faul & Erdfelder, 1992; Erdfelder, Faul, & Buchner, 1996) was used in conjunction with any available statistical guidelines. For *a priori* power calculations that could not be performed using G-Power, statistical guidelines were used. In addition, more practical time and resource considerations were taken into account.

Objective One:

There is no consensus on the sample size required to perform factor analysis, and methodologists differ in this regard. One rule of thumb suggests at least 10 cases for each item in the instrument being used. Bryant and Yarnold (1995) recommend that the subjects-to-variables ratio be no lower than 5. With a dataset containing 385 cases for an 18-item instrument, the study had adequate sample size to conduct factor analyses in Phase One.

Objective Two:

A very practical consideration in undertaking studies of this type pertains to the time and resources available to do the modeling task. One of the reasons for choosing the decomposed approach over the composite approach was its ability to model a large number of health states by necessitating fewer valuations (Dolan, 2002). If the health states that are required to be valued are randomly assigned to respondents, then it is necessary to ensure sufficient numbers of valuations for those health states. The MAUF is fitted mainly by using valuations for the corner states, single-attribute level states and anchor states. In order build in precision into models that provide the VAS to SG conversion curve, it was necessary that the marker states be valued using both VAS and SG tasks by a sufficient number of respondents as well.

In this regard, it is useful to look at studies that have designed PBMH using the MAUT framework. In the valuation study for the HUI-3 classification system, a sample size of 256 respondents was arrived at using a 2^8 fractional factorial design (Furlong et al., 1998). This was arrived at considering a couple of factors, including the need to maintain precision of the multiplicative MAUF of the HUI-3 in comparison to other instruments in the HUI family. All of the 256 respondents valued the marker states, while only 64 valued the single-attribute level states and corner states (Furlong et al., 1998). In the development of the ASUI (Revicki et al., 1998a), 161 patients participated in the study, while 100 provided valuations for the RSUI (Revicki et al., 1998b). Respondents completed valuations for all the necessary health states in these two studies.

In the present study, the necessary health states required to develop the valuation function for the health state classification system of the DUI, were each valued by a total of 100 subjects.

Objective Three:

The *a priori* calculation of required sample size necessitated making a decision on the anticipated effect size index for regression, f^2 (Cohen, 1988). f^2 reflects the proportion of variance accounted for by some source in the population (PVs) relative to the residual variance proportion (PVe), such that $f^2 = PVs / PVe$. For multiple regression with a set of predictors, the hypothesis tested was that the correlation of a set of predictors with a dependent variable is zero in the population. While f^2 is equal to $R^2 / 1-R^2$ (where R^2 is the coefficient of determination obtained from the regression using the set of predictors), the required power was decided as 0.8 with an alpha level of 0.05. For a set of predictors explaining 20% of the variance in the dependent variable, f^2 would be 0.25, and with 12 predictors, a sample size of 81 was needed to achieve a power of 0.8. Another conservative estimate necessitated at least 15 subjects per variable; for 12 predictors the required sample size was 300.

CHAPTER FOUR: RESULTS

4.1 Objective One

4.1.1 Classical Test Theory

4.1.1.2 Overview

The process of designing a diabetes health state classification system began with the exploratory analysis of a dataset containing responses of 385 people with Type 2 Diabetes to the 18-item ADDQoL collected in previous research (Sundaram, 2005). The profile of these respondents can be summarized as: about 57% respondents were female, 64 % in 40-69 age range, 94% were Caucasian. The mean diabetes duration was 10.2 (±9.1) years, mean A1C of respondents was 7.2 (±1.4), with about 50% experiencing at least one diabetes-related complication, and 62.1% were obese (BMI>30). About 49 % of respondents were on oral medications only, 32% on oral medications and insulin, and 13% on insulin only.

A negative score on the ADDQoL would indicate that diabetes was negatively affecting the QoL of the individual, while a positive score would indicate a positive effect of diabetes. Intuitively, a person with diabetes is expected to report a negative influence of the disease on his or her QoL. At the same time, the bipolar scale allows for some respondents to have positive scores, although these were expected to be uncommon. As shown in Table 4, the maximum negative impact of diabetes was felt on 'freedom to eat', 'enjoyment of food', and on 'finances', while diabetes least impacted 'society reaction', 'dependence', and 'living conditions'. The original and recoded means of the impact ratings for the 18 items are respectively represented in Columns 4 and 5 of Table 4.

Table 4: Mean Scores for ADDQoL Items

Item no.	ADDQoL Item	Weighted Impact Mean Score [1]	Original Un-weighted Impact Mean Rating [2]	Recoded Un-weighted Impact Mean Rating [3]
1	Working life	-1.73 (±2.52)	-1.04 (±1.05)	-1.04 (±1.05)
2	Family life	-2.17 (±2.72)	-1.00 (±1.00)	-1.00 (±1.00)
3	Social life	-1.56 (±2.34)	-0.74 (±0.91)	-0.74 (±0.91)
4	Sex life	-1.60 (±2.64)	-0.94 (±1.14)	-0.95 (±1.13)
5	Physical Appearance	-1.33 (±2.30)	-0.65 (±0.95)	-0.66 (±0.93)
6	Physical activities	-2.16 (±2.67)	-1.04 (±1.05)	-1.04 (±1.04)
7	Holidays/ leisure	-1.75 (±2.47)	-0.87 (±1.01)	-0.87 (±1.01)
8	Travel	-1.84 (±2.51)	-0.98 (±1.04)	-0.98 (±1.04)
9	Confidence in ability	-1.67 (±2.48)	-0.79 (±0.97)	-0.79 (±0.97)
10	Motivation	-1.59 (±2.38)	-0.77 (±0.94)	-0.77 (±0.94)
11	Society reaction	-0.88 (±1.84)	-0.47 (±0.80)	-0.47 (±0.80)
12	Future	-2.13 (±3.34)	-0.92 (±1.28)	-0.96 (±1.20)
13	Finances	-2.20 (±2.90)	-0.98 (±1.06)	-0.98 (±1.06)
14	Dependence	-1.23 (±2.89)	-0.55 (±1.17)	-0.59 (±1.09)
15	Living Conditions	-1.32 (±2.31)	-0.61 (±0.92)	-0.61 (±0.92)
16	Freedom to eat	-3.10 (±3.04)	-1.54 (±1.13)	-1.54 (±1.12)
17	Enjoyment of food	-2.89 (±3.10)	-1.36 (±1.13)	-1.36 (±1.12)
18	Freedom to drink	-1.89 (±2.75)	-0.99 (±1.15)	-1.01 (±1.11)

1. Original Weighted Impact Score ranges between -9 to +9
2. Original Un-weighted Impact Rating ranges between -3 to +3
3. Recoded Un-weighted Impact Rating ranges between -3 to +1

4.1.1.2 Examination of Factor Structure

The classical index of discrimination, the item-total correlation coefficients (r) for an item's weighted impact score with the total average weighted impact score was greater than 0.4 for each item, and hence, none of the items were marked for exclusion on this basis. Exploratory Factor Analysis (EFA), employing Principal Axis Factor (PAF) extraction and Promax rotation, was performed on all the 18 items. The Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy as well as the Bartlett test of sphericity indicated that FA would yield useful information from the data. The Scree plot indicated one major factor accounting for 49% of the total variance. The loading pattern of items to the dominant factor was unclear due to which the factor could not be described without ambiguity. From the loading of the items on the rotated factor solution, two smaller factors could be identified as diet-related ('food', 'eat', and 'drink') accounting for 11% of the total variance, and relationships-related ('work', 'family', 'friends', and 'sex') accounting for 6% of the total variance. Overall, these results were similar to those obtained from the EFA of the weighted impact scores.

The next steps involved iterative EFAs excluding items with:

- communalities less than 0.4 – 'worries', 'sex', 'drink', 'dependence', and 'finance';
- similar loadings on different factors (such that the differences in loadings across factors is ≤ 0.1) – 'leisure'.

Excluding the above six items again resulted in a factor structure with one dominant factor accounting for 53% of the total variance. Again, this factor could not be described without ambiguity on the basis of the loading of items, but a smaller factor (accounting for 6% of the total variance) could be identified as diet-related ('food', and 'eat'). Since the exploratory nature of the exercise was geared more towards understanding the composition of plausible factors, a forced-three factor solution was requested with the same set of (12) remaining items. In addition to the diet-related factor, the relationships-related factor could be identified again (‘work’, ‘family’, and ‘friends’), while the remaining seven items could at best be described as a functioning-related attribute. The factor loadings of the 12 items from the forced 3-factor rotated solution is represented in Table 5.

This factor structure formed the basis of the plausible attributes suggested to the Clinical Expert Panel members as part of the process of designing the diabetes-specific health state classification system. The plausible attributes submitted to the expert panel was expanded in its scope, in that they also included relevant ADDQoL items that were excluded in successive EFAs in adherence to statistical criteria (see Appendix 2).

Table 5: Factor Loading from the Forced 3-Factor Rotated Solution of ADDQoL Items

	Factor Loadings [1]		
ADDQoL Item	**Factor 1** [2]	**Factor 2** [3]	**Factor 3** [4]
Motivation	0.879	0.024	-0.037
Confidence in Ability	0.848	0.023	-0.042
Travel	0.842	-0.083	0.114
Physical	0.715	0.068	0.052
Living	0.690	0.041	0.030
Society Reaction	0.462	0.257	-0.061
Appearance	0.452	0.320	-0.149
Family	0.023	0.822	0.037
Friends	0.066	0.724	0.035
Work	0.297	0.510	0.042
Eat	0.019	-0.028	0.932
Food	-0.021	0.066	0.806

1. Factor loadings for each variable on the factors; using Principal Axis Factoring and Promax Rotation.
2. Plausible variable: Functioning-related.
3. Plausible variable: Relationships-related.
4. Plausible variable: Diet-related.

4.1.2 Input from the Clinical Expert Panel

The importance ratings provided by patients with Type 2 Diabetes to the 18 ADDQoL items (contained in a dataset from previous research, n=385) were compared with the ratings provided by the expert panel members (n=7). On the basis of these importance ratings, two lists of the ten most important items, one each generated from patient and expert panel input, were generated. Added to this pool were items not contained in the ADDQoL but which the experts considered to be important. The intent behind the exercise was to build an item pool with which to construct and describe up to five attributes for the diabetes-specific classification system, in concert with the results obtained from the statistical analyses described in the previous section.

A comparison of ratings showed that the patients and the expert panel both included five ADDQoL items in their respective lists of ten important items. These items are listed as 'top consensus items' in Table 6, and were added to the item pool discussed above. Other items that figured in either but not both lists are listed as 'top non-consensus items' in Table 6, and were also added to this pool along with those items that are not contained in the ADDQoL but which the experts considered to be important, listed as 'others' in the table. Finally, these items were organized into attributes (see Table 6) to be included in the first version of the diabetes health state classification system based on the factor loadings observed in the final iteration of EFA and expert panel input on the composition of attributes from ADDQoL and other items.

Successive pilot versions of the diabetes health state classification, titled the Diabetes Utility Index (DUI), were analyzed using Rasch Analysis.

Table 6: Pooling Items for the Classification System based on Patient and Expert Panel Input

Top Consensus Items [1,2]	Top Non-Consensus Items [2,3]	Other Items [4]	Plausible Attribute [5]
Food 2.0 (±0.79) [10] / *2.6 (±1.13) [4]*	**Eat** 2.0 (±0.78) [12] / *2.9 (±0.38) [1]* **Drink** 1.6 (±0.88) [18] / *2.1 (±0.69) [9]*		**Enjoyment of Diet**
Family 2.4 (±0.72) [1] / *2.7 (±0.49) [2]*	**Work** 2.1 (±0.76) [2] / *2.4 (±0.53)[5]* **Friends** 2.1 (±0.81) [6] / *1.9 (±1.21)[11]* **Sex** 1.9 (±0.92) [14] / *1.9 (±0.90) [12]*		**Relationships**
Physical 2.1 (±0.71) [3]/ *2.0 (±1.00) [10]*	**Dependence** 2.0 (±0.84) [13] / *2.3 (±0.76) [7]*	**Pain** **Energy**	**Physical ability & Energy level**
Confidence 2.0 (±0.78) [8]/ *2.1 (±0.90) [8]*	**Worries** 2.0 (±0.79) [9] / *1.9 (±0.90) [14]* **Motivation** 2.0 (±0.76) [7] / *1.4 (±0.98) [15]*	**Depression**	**Mood and feelings**
Confidence /2.0 (±0.78) [8]/ *2.1 (±0.90) [8]*	**Motivation** 2.0 (±0.76) [7] / *1.4 (±0.98) [15]*		**Satisfaction**
Finances 2.1 (±0.76) [4] / *2.7 (±0.49) [3]*	**Living Conditions** 2.0 (±0.74) [5] / *1.0 (±0.82) [17]* **Dependence** 2.0 (±0.84) [13] / *2.3 (±0.76) [7]* **Travel** 1.8 (±0.81) [16] / *2.3 (±0.76) [6]*	**Self-Care**	**Lifestyle freedom/ Life Situations/ Freedom Living Life**

1. Items that were ranked by both patients and experts in their Top 10 among most important attributes
2. The statistics are present as follows:
   patient importance rating: Mean (SD) [importance rank] / *expert importance rating: Mean (SD) [importance rank]*. Importance ratings are on a scale were '0'represents 'not at all important' and '3'represents 'very important'. A Rank of '1' denotes the highest possible rank.
3. Items that were ranked by either patients or experts, but not both, in their Top 10 among most important attributes.
4. Other items considered as important by the experts
5. Plausible attribute that can be explained by a combination of consensus, non-consensus, and other important items.

4.1.3 Modern Test theory

4.1.3.1 Fit of Items to the Rasch Model

Rasch Analysis was employed on data obtained from the pilot testing of three versions of the diabetes health state classification system (n1=52, n2=65, n3=111) using a software package called Winsteps (Linacre, 2007). This section also includes results from the Rasch Analysis of the dataset obtained from the Validation Survey (n=396) that was sent out after the diabetes health state classification system, the DUI, was finalized at the end of the last pilot round. Table 7 summarizes the endorsement by the survey respondents of the severity levels contained in the various versions of the classification system. The fit of the selected attributes (or items, since each attribute is also one item of the classification system) to the Rasch model was assessed on the basis of the Chi-Square based statistics of Infit and Outfit Mean Squares (MNSQs). These statistics are represented in Table 8.

The versions tested in pilots 1 and 2 differed, among others, largely in the nature of the rating scales employed. The rating scale was standardized to some extent in Pilot 2 for all the attributes. From the Infit and Outfit MNSQs in the Pilot 2 results (Table 8), the relative benefits of this standardization are not entirely evident, presumably because of issues related to the overfit of the attribute (item) ‘freedom living life’. This was the only major deviation from the Rasch model identified across the different pilot versions tested. The problem persisted even upon modification of the attribute (‘freedom living life’ in Pilot 1) to present a related construct of ‘life style freedom’ (in Pilot 2). Overfit of an item to the Rasch model indicates item redundancy, or dependence on other items included in the model which in this case would be the attribute ‘physical ability and energy level'. This dependence can be explained on the basis of the theoretical expectation that the respondents with impaired physical ability may also naturally report constraints with their lifestyle freedoms.

The issue of dependent items needed to be addressed because of two reasons. First, the classification system was being designed to be as comprehensive as logically feasible and parsimonious - this necessitated the exclusion of an attribute that may be found to be redundant with another that provided more useful information. Second, attribute independence was a characteristic desired in the classification system in order to avoid complications in the determination of its utility scoring function as noted in the development of the HUI-2 (Torrance et al., 1996). The preference elicitation of some health states required to be specified when the decomposed approach is being employed (for example, those with the worst level in one attribute, yet the best in others) could be problematic if such states are implausible because of the lack of independence among attributes (Feeny, 2002a).

Exclusion of the 'lifestyle freedoms' attribute corrected the problem of misfit that came to be identified in the related 'physical ability & energy level' attribute (compare Pilot 2 and Pilot 3 results for the attribute in Table 8), perhaps on account of the dependent nature of the association between the two. The replacement was made in concert with the opinions of the clinical expert panel and tested in Pilot 3. The attribute 'Satisfaction with management of diabetes' was found to fit the Rasch model desirably without impairing the fit of the other attributes (especially the 'Mood & Feelings' attribute, with which it can be marginally expected to be related). It was also expected that, with the inclusion of the satisfaction attribute, no major problems would arise during the preference elicitation of health states that are required to be specified in order to calculate a MAUF for the classification system using the decomposed approach.

The final DUI attributes were: 'Physical Ability & Energy', 'Relationships', 'Mood & Feelings', 'Enjoyment of Diet', and 'Satisfaction with Management of diabetes'. Results of the validation survey, included in Table 8, indicated Infit MNSQs ranging between 0.83 and 1.1, and Outfit MNSQs ranging between 0.77 and 1.07 for the five attributes indicating that the five DUI attributes in general adequately fit the Rasch uni-dimensional model. Person and item reliabilities were 0.71 and 0.96, while the respective separation ratios were 1.55 and 5.01. The 'Physical Ability & Energy Level' attribute was the most difficult and the 'Satisfaction' attribute was the easiest. The Item-Separation and Item-Reliability at 5.01 and 0.96, respectively, indicated a fair spread of items. The Person-Separation was moderate at 1.55, while the Person-Reliability was acceptable at 0.71.

Table 7: Endorsement of Severity Ratings of Versions of the Diabetes Utility Index

	Pilot 1 [1]		Pilot 2 [2]		Pilot3 [2]		Validation Survey [4]	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Attribute 1	*Physical Ability*		*Physical Ability & Energy Level*		*Physical Ability & Energy Level*		*Physical Ability & Energy Level*	
Level 1	13	25.0	23	35.4	57	51.4	183	46.2
Level 2	29	55.8	31	47.7	44	39.6	148	37.4
Level 3	8	15.4	8	12.3	7	6.3	49	12.4
Level 4	2	3.8	3	4.6	3	2.7	16	4.0
Attribute 2	*Relationships*		*Relationships*		*Relationships*		*Relationships*	
Level 1	21	40.4	18	27.7	33	29.7	127	32.1
Level 2	20	38.5	35	53.8	60	54.1	168	42.4
Level 3	9	17.3	9	13.8	17	18.3	76	19.2
Level 4	2	3.8	3	4.6	1	0.9	25	6.3
Attribute 3	*Mood & Feelings*		*Mood & Feelings*		*Mood & Feelings*		*Mood & Feelings*	
Level 1	11	21.2	2	3.1	6	5.4	38	9.6
Level 2	25	48.1	38	58.5	83	74.8	244	61.6
Level 3	13	25	21	32.3	20	18.0	101	25.5
Level 4	3	5.8	4	6.2	2	1.8	13	3.3
Attribute 4	*Enjoyment of Diet*		*Enjoyment of Diet*		*Enjoyment of Diet*		*Enjoyment of Diet*	
Level 1	12	23.1	8	12.3	31	27.9	111	28
Level 2	26	50.0	39	60.0	60	54.1	194	49.0
Level 3	12	23.1	14	21.5	15	13.5	67	16.9
Level 4	2	3.8	4	6.2	5	4.5	24	6.1
Attribute 5	*Freedom Living Life*		*Lifestyle Freedoms*		*Satisfaction with Managing Diabetes*		*Satisfaction with Managing Diabetes*	
Level 1	20	38.5	7	10.8	44	39.6	126	31.8
Level 2	22	42.3	43	66.2	48	43.2	210	53.0
Level 3	6	11.5	12	18.5	19	17.1	60	15.2
Level 4	4	7.7	3	4.6	-	-	-	-
Total N	52	100	65	100	111	100	396	100

1. Pilot data from the testing of Version One. See Appendix 4 to view this version.
2. Pilot data from the testing of Version Two. See Appendix 6 to view this version.
3. Pilot data from the testing of Final Version. See Appendix 7 to view this version.
4. Data from the Validation Survey using the Final Version.

Table 8: Item and Person Fit Statistics for Versions of the Diabetes Utility Index

	Pilot 1 [1]		**Pilot 2 [2]**		**Pilot3 [3]**		**Validation Survey [4]**		
Item Reliability	0.53		0.87		0.92		0.96		
Item Separation	1.07		2.63		3.34		5.01		
Person Reliability	0.68		0.77		0.65		0.71		
Person Separation	1.46		1.81		1.36		1.55		
	I [5]	**O [6]**	**I [5]**	**O [6]**	**I [5]**	**O [6]**	**I [5]**	**O [6]**	**Difficulty [7]**
Attribute 1	*Physical Ability*		*Physical Ability & Energy Level*		*Physical Ability & Energy Level*		*Physical Ability & Energy Level*		
	1.10	1.19	0.81	1.52	1.07	1.10	0.83	0.77	0.86
Attribute 2	*Relationships*		*Relationships*		*Relationships*		*Relationships*		
	1.08	1.08	1.02	0.95	1.07	1.02	1.05	1.02	0.12
Attribute 3	*Mood & Feelings*		*Mood & Feelings*		*Mood & Feelings*		*Mood & Feelings*		
	1.16	1.09	1.01	0.99	0.88	0.96	0.90	0.85	-0.51
Attribute 4	*Enjoyment of Diet*		*Enjoyment of Diet*		*Enjoyment of Diet*		*Enjoyment of Diet*		
	0.99	0.97	1.37	1.15	0.91	0.92	1.06	1.06	0.09
Attribute 5	*Freedom Living Life*		*Lifestyle Freedoms*		*Satisfaction with Managing Diabetes*		*Satisfaction with Managing Diabetes*		
	0.58	0.59	0.58	0.48	1.02	0.94	1.11	1.07	-0.57
Total N	**52**		**65**		**111**		**396**		

1. Pilot data from the testing of Version One. See Appendix 4 to view this version.
2. Pilot data from the testing of Version Two. See Appendix 6 to view this version.
3. Pilot data from the testing of Final Version. See Appendix 7 to view this version.
4. Data from the Validation Survey using the Final Version.
5. Infit Mean Squares
6. Outfit Mean Squares
7. Item Difficulty in logits

4.1.3.2 Rating Scale Diagnostics Using Rasch Analysis

The performance of the rating scales (severity levels) of the DUI were judged on the basis of their Average measures, Step calibrations, and Infit and Outfit MNSQs. The Average measures for each rating category were required to advance evenly up the severity level, and be reasonably close to their expected values. The Step calibrations were required to advance in an orderly manner such that the distance between adjacent categories was between 1.5 logits and 5.0 logits. Outfit MNSQs for category levels were required to be < 2.0, indicating less randomness and unexpected use of category levels.

While interpreting the results of the Rasch rating scale diagnostics (Tables 9 and 10) , one needs to bear in mind that the results are often unstable under smaller sample sizes of the kind encountered in the pilot testing of the DUI (Linacre, 2007, personal communication), chiefly in Pilots 1 and 2. The results may also be unstable in the event of poor endorsement of certain levels of the rating scale because of the small sample size overall or because of poor construction that may leave certain levels un-endorsed in spite of a reasonable sample size. Those results in Tables 9 and 10 that are not in conformance to the desired norms listed in the previous paragraph have been italicized.

The rating scales saw a major change between the versions tested in Pilots 1 and 2. The rating scale tested in Pilot 1 was directly influenced by suggestions from the clinical expert panel and was more subjective as compared to the scale employed in later versions. Results of the scale diagnostics were largely acceptable. However, a decision was made to adopt a reasonably standardized rating scale across attributes and to test it in the next pilot. The virtues expected from this change were:

1. less ambiguity in interpreting the meaning of the scale levels;

2. ease of identification of the differences between scale levels; and
3. less respondent burden from faster completion times because of familiar rating scale construction across attributes.

An added, though comparatively less important, benefit of the new rating scale construction was seen in the reduced size of the instrument. The scale levels performed satisfactorily across the DUI attributes; the result was confirmed from the Rasch rating scale diagnostics conducted on the data from the Validation Survey, as can be seen in Tables 9 and 10. In the final version of the DUI, the attributes 'Physical Ability & Energy Level', 'Relationships', 'Mood & Feelings', and 'Enjoyment of Diet' contained four levels of severity each while the fifth attribute 'Satisfaction with the Management of Diabetes' contained three levels. This classification system can describe a total of 768 (4x4x4x4x3=768) unique health states.

Overall, the Rasch rating scale diagnostic tests of data obtained on the final validation survey indicated that the severity levels of the DUI attributes performed well. The average measures were found to increase with the severity levels of each DUI attribute (Table 10). The probability are presented in Figure 6 for each attribute, with Level 1 representing the best level and Level 4 (Level 3 in attribute #5) representing the lowest (or most severe) level. The curves show distinct peaks indicating that each category (severity level) is the most probable response category for some level on that attribute. These results indicated that the rating scale statements were satisfactorily understood and differentiable. Excluding Level 2 on "Mood & Feelings" which spanned a relatively greater distance on the attribute, the step calibrations for the severity levels of other attributes were within expectations.

Table 9: Rating Scale Diagnostics for Versions of the Diabetes Utility Index (1)

	Pilot 1 [1]				Pilot 2 [2]			
	I [4]	**O [5]**	**AM [6]**	**SC [7]**	**I [4]**	**O [5]**	**AM [6]**	**SC [7]**
Attribute 1	*Physical Ability*				*Physical Ability & Energy Level*			
Level 1	0.76	0.74	-3.07	none	1.32	1.03	-3.28	none
Level 2	0.89	0.88	-1.20	-3.28	0.61	*2.68*	-0.99	-3.28
Level 3	0.84	0.83	0.79	0.87	0.45	0.37	1.83	0.79
Level 4	2.37	*3.66*	*-0.18*	2.41	0.66	0.59	3.13	2.50
Attribute 2	*Relationships*				*Relationships*			
Level 1	0.75	0.79	-2.65	none	1.64	1.46	-3.34	none
Level 2	0.85	0.76	-.69	-2.62	0.79	0.71	-1.28	-3.68
Level 3	1.63	1.85	0.31	-0.04	0.74	0.70	1.51	0.87
Level 4	1.35	1.30	0.98	2.67	0.65	0.58	3.13	2.82
Attribute 3	*Mood & Feelings*				*Mood & Feelings*			
Level 1	2.19	1.88	-1.99	none	0.92	0.77	-5.69	none
Level 2	0.59	0.64	*-2.01*	-3.09	1.18	1.19	-2.33	-5.61
Level 3	0.71	0.67	0.33	0.38	1.00	1.03	0.35	0.93
Level 4	1.15	1.11	1.14	2.71	0.82	0.78	2.44	4.68
Attribute 4	*Enjoyment of Diet*				*Enjoyment of Diet*			
Level 1	1.40	1.26	-2.52	none	2.22	1.61	-3.86	none
Level 2	0.82	0.91	-1.69	-3.21	0.96	0.88	-1.81	-4.41
Level 3	0.83	0.82	0.35	0.34	1.26	0.99	0.74	0.90
Level 4	0.81	0.77	1.69	2.87	1.22	1.10	1.89	3.51
Attribute 5	*Freedom living Life*				*Lifestyle Freedoms*			
Level 1	0.66	0.77	-2.76	none	0.77	0.58	-5.07	none
Level 2	0.73	0.67	-1.05	-2.19	0.51	0.56	-1.74	-4.89
Level 3	0.45	0.31	0.51	0.93	0.49	0.32	1.56	1.14
Level 4	0.39	0.36	1.91	1.27	0.64	0.59	3.13	3.75
Total N	52		100%		65		100%	

1. Pilot data from the testing of Version One. See Appendix 4 to view this version.
2. Pilot data from the testing of Version Two. See Appendix 6 to view this version.
4. Infit Mean Squares
5. Outfit Mean Squares
6. Average Measure
7. Step Calibration

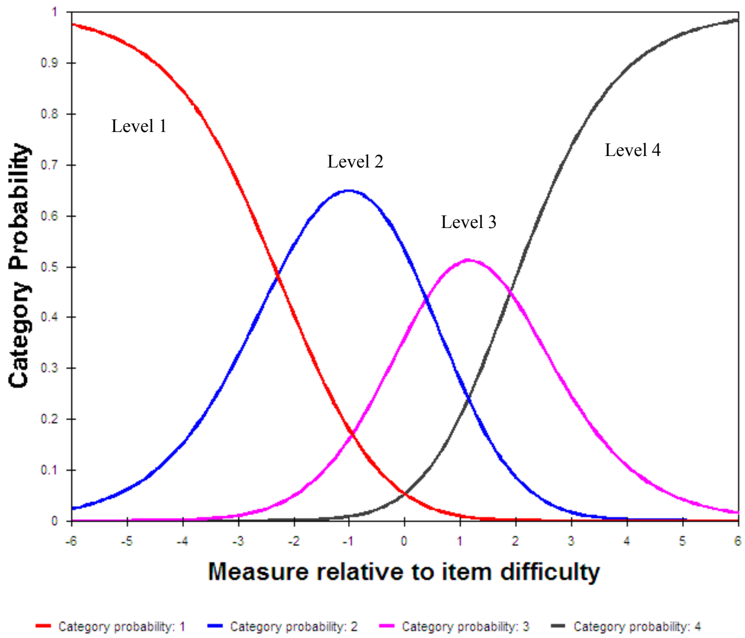
Table 10: Rating Scale Diagnostics for Versions of the Diabetes Utility Index (2)

	Pilot3 [1]				**Validation Survey [2]**			
	I [3]	**O [4]**	**AM [5]**	**SC [6]**	**I [3]**	**O [4]**	**AM [5]**	**SC [6]**
Attribute 1	*Physical Ability & Energy Level*				*Physical Ability & Energy Level*			
Level 1	0.93	0.96	-2.65	none	0.89	0.90	-2.53	none
Level 2	1.03	1.05	-0.88	-2.40	0.75	0.64	-0.62	-2.28
Level 3	0.74	0.68	1.05	0.97	0.75	0.69	1.14	0.39
Level 4	1.89	*2.56*	1.53	1.14	0.99	1.07	2.23	1.90
Attribute 2	*Relationships*				*Relationships*			
Level 1	1.11	1.09	-3.13	none	1.15	1.10	-2.76	none
Level 2	.82	.71	-1.47	-3.71	0.89	0.87	-1.19	-2.54
Level 3	.80	.79	0.84	.02	0.94	0.91	0.52	0.33
Level 4	*4.53*	*5.02*	-1.72*	3.70	1.27	1.39	1.74	2.22
Attribute 3	*Mood & Feelings*				*Mood & Feelings*			
Level 1	1.14	1.46	-3.48	none	1.30	1.10	-2.88	none
Level 2	.82	.87	-2.08	-5.41	0.74	0.77	-1.86	-4.68
Level 3	.79	.74	.43	1.21	0.77	0.71	0.68	0.76
Level 4	.90	.8	2.75	4.20	1.09	1.03	1.87	3.92
Attribute 4	*Enjoyment of Diet*				*Enjoyment of Diet*			
Level 1	1.04	1.03	-3.14	none	1.25	1.13	-2.77	none
Level 2	.90	.98	-1.64	-2.99	0.93	1.03	-1.28	-2.88
Level 3	.64	.59	.49	0.83	0.85	0.80	0.69	0.65
Level 4	1.01	1.07	1.90	2.15	1.24	1.37	1.73	2.23
Attribute 5	*Satisfaction*				*Satisfaction*			
Level 1	.84	.84	-2.98	none	1.17	1.16	-2.70	None
Level 2	1.00	.85	-1.16	-1.41	0.99	0.93	-0.88	-1.95
Level 3	1.24	1.20	0.46	1.41	1.14	1.12	0.94	1.95
Level 4	-	-	-	-	-	-	-	-
Total N	111		100%		396		100%	

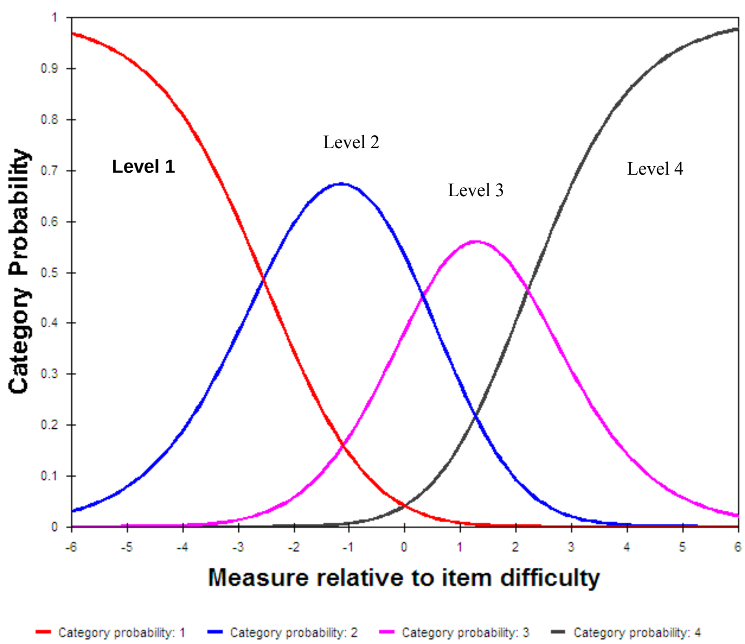
1. Pilot data from the testing of Final Version. See Appendix 7 to view this version.
2. Data from the Validation Survey using the Final Version.
3. Infit Mean Squares
4. Outfit Mean Squares
5. Average Measure
6. Step Calibration

Figure 6: **Rasch Category Probability Curves (Validation Survey)**

Physical Ability & Energy Level *

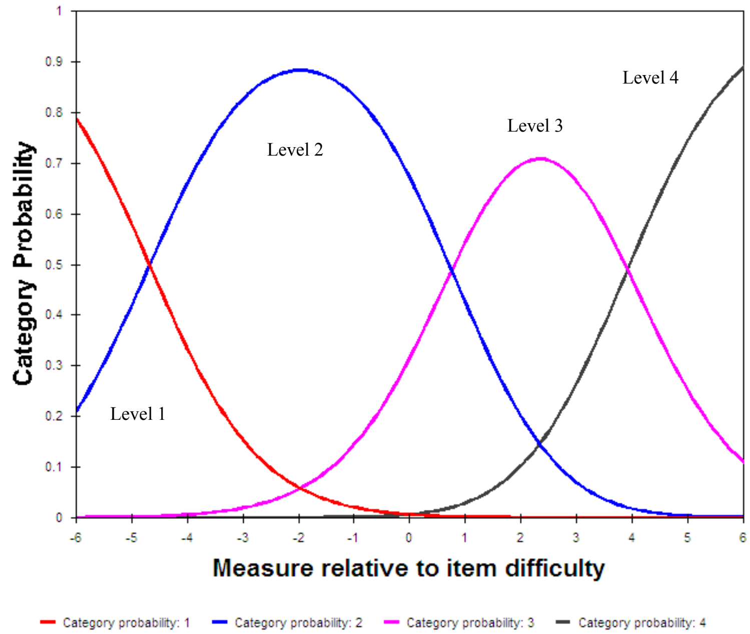


Relationships *

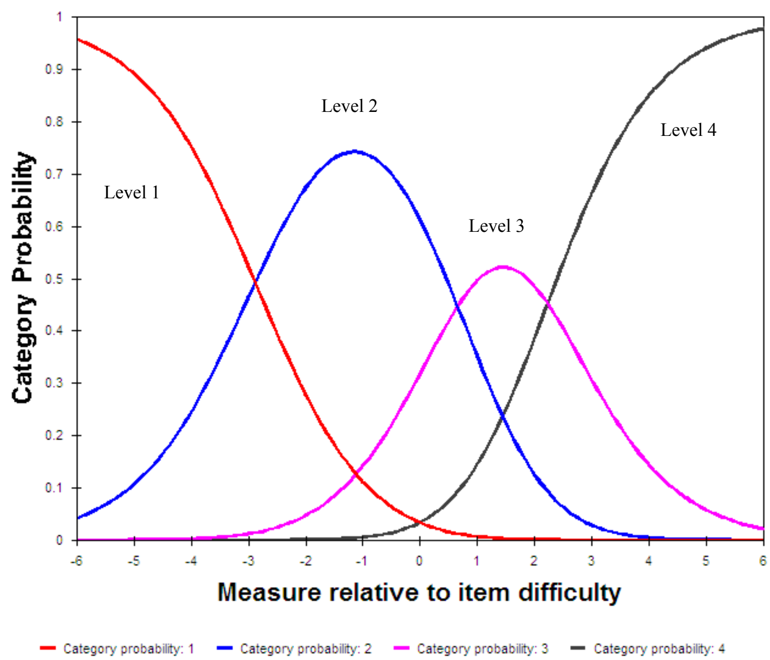


** See footnote presented after the Rasch probability curves for the 'Satisfaction' attribute*

Mood & Feelings *

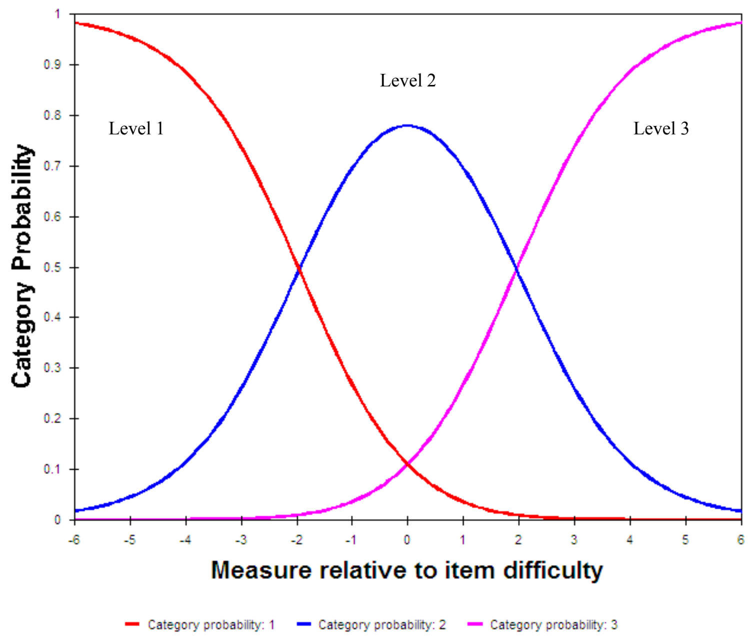


Enjoyment of Diet *



** See footnote presented after the Rasch probability curves for the 'Satisfaction' attribute*

Satisfaction with Management of Diabetes



1. For each DUI attribute, Level 1 represents the best level and Level 4 (Level 3 in attribute #5) represents the lowest (or most severe) level.
2. The x-axis (Category Probability) represents the probability of responding to a particular level on an attribute. The y-axis (Measure relative to item difficulty) represents the difference in a person's ability (severity on an attribute) and item (attribute) difficulty in logits. *For example, individuals with greater abilities (higher level of severity) on attributes relative to the difficulty of the attribute (for example, +4 on the x-axis) have a higher probability of responding to levels 3 or 4 on the attribute.*
3. The intersections of rating scale curves represent step calibrations. Step calibrations between adjacent severity levels in the range of 1.5 and 5.0 logits indicate that the severity levels are neither too narrow nor too broad.

4.2 Objective Two

4.2.1 Overview

A total of 100 persons with diabetes participated in the utility elicitation interviews, all conducted by a single interviewer. The demographic profile of the participants is described in Table 11. The interviews took 64 (±17.5) minutes to complete on an average, with the lengthiest interview lasting 135 minutes. The average number of minutes the participants took to complete the Feeling Thermometer (FT) and Chance Board (CB) tasks were 37 (±14.9) and 16 (±8.5) respectively, with the most amount of time being 85 minutes for the FT and 40 minutes for the CB. While 4% of the participants found the FT tasks to be either difficult or very difficult, 74% found the tasks to be either easy or very easy. A similar proportion (70%) found the CB tasks to be either easy or very easy while 10% rated them to be difficult or very difficult. Almost 94% of the interviews were rated by the interviewer to be either good or very good overall, with 91% of the participants understanding the questions asked to them either totally or for the most part.

Towards the beginning of the interview, 31 participants found Pits to be the least desirable condition while the remaining 69 regarded Dead as least desirable. All the participants were consistent in their choice of the least desirable state between the FT and CB stages of the interview, and hence Group A and Group B respondents, respectively, formed 31% and 69% of the sample interviewed. The data were analyzed as delineated in the methods section, the basic approach being to calculate aggregated Person-Means of the scale anchors, marker states and single-attribute utility states for the two groups. The results of these analyses are summarized in Tables 12 and 13.

Table 11: Demographic Summary of Interview Participants

	n	%
Gender		
Male	32	32%
Female	68	68%
Race		
White	92	92%
Black	6	6%
Other	2	2%
Marital Status		
Single	7	7%
Married	69	69%
Divorced/ Separated	11	11%
Widowed	10	10%
Not married/ living with partner	3	3%
Education		
High school or less	5	5%
High school graduate/ GED	23	23%
Some college/ vocational college	27	27%
College degree	22	22%
Graduate / professional degree	23	23%
Employment		
Employed	50	50%
Retired	30	30%
Home-maker	12	12%
Student	2	2%
Seeking work	3	3%
Other	3	3%
Household income		
$25,000 or less	19	19%
$25,001 - $50,000	31	31%
$50,001 - $75,000	28	28%
More than $75,000	16	16%
Refused	6	6%
Total N	100	100%

4.2.2 Transformation of Measured Preferences

The following calculations were performed separately for Groups A and B for all the health states described:

The ten percent trimmed means were calculated from the raw FT values of the scale anchors, marker states and single-attribute utility states as well as the CB utilities for marker states and anchors, and are depicted in Columns 2 and 3 of Tables 12 and 13. No further adjustments were made to the CB utilities obtained separately from the two groups.

An end-of-scale bias adjustment (EOSBA) was applied to those FT values of marker states and single-attribute level states greater than 75. Only Marker State A and second-ranked single-attribute health states (i.e. those immediately below PH in a set of single-attribute level states for one attribute) needed this adjustment; the adjusted values are depicted in Column 4.

This sort of adjustment (that changes the FT value of only one health state in a set) necessitates that the original spacing between health states in a set be restored. This was accomplished for single-attribute level states by rescaling the FT values between the EOSBA-adjusted value and the value of the lowest single-attribute level state (i.e. the corner state, using positive linear transformations). The resultant rescaled FT values are depicted in Column 5.

Table 12: Measured and Calculated Preference Scores - Group A

State		n	10% Trimmed Mean Value [1]	EOSBA-Adjusted Value [1]	Re-scaled Value [1]	Calculated Utility [2]
Physical Ability& Energy Level	**2**	31	92.5	95.8	95.8	0.914
	3	31	75.5	75.5	77.6	0.728
	4	31	45.0	45.0	45.0	0.395
Relationships	**2**	31	95.0	97.2	97.2	0.929
	3	31	77.1	77.1	78.3	0.735
	4	31	54.0	54.0	54.0	0.487
Mood & Feelings	**2**	31	92.0	95.5	95.5	0.911
	3	31	72.0	72.0	74.0	0.691
	4	31	47.0	47.0	47.0	0.415
Enjoyment of Diet	**2**	31	95.0	97.2	97.2	0.928
	3	31	75.0	75.0	75.9	0.711
	4	31	60.0	60.0	60.0	0.549
Satisfaction	**2**	31	95.0	97.2	97.2	0.929
	3	31	65.0	65.0	65.0	0.600
Dead		31	6.5	6.5	6.5	0.001
Pits		31	0.0	0.0	0.0	0.000

1. Person-Mean(A) Values on a Perfect Health=100 / Pits=0 scale, measured on Visual Analog Scale tasks using a Feeling Thermometer.
2. Person-Mean(A) Utilities on a Perfect Health=1.00 / Pits=0.00 scale, calculated using a Group A-specific value to utility conversion algorithm.

Table 13: Measured and Calculated Preference Scores – Group B

State		n	10% Trimmed Mean Value [1]	EOSBA-Adjusted Value [1]	Re-scaled Value [1]	Calculated Utility [2]	Calculated Re-scaled Utility [3]
Physical Ability& Energy Level	**2**	69	93.3	96.2	96.2	0.911	0.901
	3	69	73.1	73.1	74.7	0.715	0.684
	4	69	48.0	48.0	48.0	0.473	0.415
Relationships	**2**	69	95.0	97.2	97.2	0.920	0.911
	3	69	78.0	78.0	79.3	0.757	0.730
	4	69	55.1	55.1	55.1	0.537	0.486
Mood & Feelings	**2**	69	95.0	97.2	97.2	0.920	0.911
	3	69	75.0	75.0	76.3	0.730	0.700
	4	69	45.0	45.0	45.0	0.446	0.384
Enjoyment of Diet	**2**	69	96.0	97.8	97.8	0.925	0.917
	3	69	77.0	77.0	77.8	0.743	0.715
	4	69	62.0	62.0	62.0	0.600	0.556
Satisfaction	**2**	69	96.0	97.8	97.8	0.925	0.917
	3	69	70.1	70.1	70.1	0.674	0.637
Dead		69	0.0	0.0	0.0	0.000	-0.111
Pits		69	11.4	11.4	11.4	0.280	0.000

1. Person-Mean(B) Values on a Perfect Health=100 / Dead=0 scale, measured on Visual Analog Scale tasks using a Feeling Thermometer.
2. Person-Mean(B) Utilities on a Perfect Health=1.00 / Dead=0.00 scale, calculated using a Group B-specific value to utility conversion algorithm.
3. Person-Mean(B) Utilities on a Perfect Health=1.00 / Pits=0.00 scale.

4.2.3 Fitting Value-Utility Conversion Functions

The next set of calculations performed separately for Groups A and B involved estimating value to utility conversion functions that would enable the conversion of FT values for single-attribute utility states into Standard Gamble utilities (that were not directly obtained from the CB for those states). In order to calculate these two conversion functions, 10% trimmed means of FT values and CB utilities of the three marker states from the two groups were used separately (For Marker state MA, the EOSBA-adjusted 10% trimmed means were used). Additionally, the 10% trimmed means of CB utilities of Dead and Pits were respectively used in the calculation of conversion functions for Groups A and B, respectively.

The calculation of a conversion function was relatively simple to accomplish with Group A data where both the FT values and the CB utilities were obtained on a PH-Pits scale. While Group B FT values were obtained on a PH-Dead scale, CB utilities for the group were obtained on a PH-Pits scale. However, the proposed conversion function emanating from this sub-exercise would be applied to Group B FT values calculated for single-attribute level states for a PH-Dead scale. Hence, the CB utilities for Marker States obtained for this group had to be rescaled from the PH-Pits scale in which they were measured to a PH-Dead scale by using the CB utility for Pits obtained on a PH-Dead scale from Group B respondents. This calculation is described in Table 14.

Table 14: Linear Rescaling of Values from Pits/PH Scale to Dead/PH Scale for Group B Respondents

	Marker State MA	Marker State MB	Marker State MC
Utility on Pits/PH scale	0.90	0.43	0.14
Disutility on Pits/PH scale	0.10	0.57	0.86
Range of Pits/PH utility scale	1.00	1.00	1.00
Utility of Pits on Dead/PH scale	0.10	0.10	0.10
Range of Pits/PH utility scores on Dead/PH scale	0.90	0.90	0.90
Disutility on Dead/PH scale	0.09	0.51	0.77
Utility on Dead/PH scale	**0.91**	**0.49**	**0.23**

Two simple linear regression models, one each for Group A and Group B, employing EOSBA-adjusted FT values as the predictor variable and the 10% trimmed-CB utilities as the dependent variable, were tested on data from the three marker states (n=4). With adjusted R^2s of 0.969 and 0.947 respectively, the simple linear regression models employed with Group A and Group B data were found to predict the required utility estimates in their respective groups adequately. The data used in the regression models as well as the parameters used to calculate the value-to-utility conversion functions for the two groups are described in Table 15.

4.2.4 Calculation of Overall Person-Mean Utilities

The value-to-utility conversion functions were used to calculate Person-Mean(A) and Person-Mean(B) utility scores from the EOSBA-adjusted and rescaled FT values for the single-attribute level states and corner states of the respective groups. The calculated Person-Mean(A) and Person-Mean(B) utilities are represented in Column 6 of Tables 12 and 13. In order to combine these utilities into one aggregated variable on a common scale, Person(B) utilities were rescaled to a PH-Pits scale by positive linear transformation. The rescaled Person(B) utilities are represented in Column 7 of Table 13.

This facilitated the calculation of overall aggregated Person-Mean utility scores using the respective prevalence proportion of each Groups A and B as weighting factors. The weights employed for Groups A and B were 0.31 and 0.69 respectively (based on group sizes of 31 and 69 out of a total N of 100 participants). The overall aggregated Person-Mean utility scores are represented in Column 2 of Table 16.

Table 15: Data Used to Estimate Value to Utility Conversion Models

	Group A (PH-Pits Scale)		
Marker State	10% Trimmed Mean Value [1]	EOSBA-Adjusted Value	10% Trimmed Mean Utility [2]
MA	91.5	95.2	0.93
MB	58.6	51.1	0.43
MC	22.9	23.3	0.13
Dead	9.1	6.5	0.05

Model Fit Statistics: Adj. R^2 *=0.9828; F(1,3)=172.87; p=0.005*

Utility Conversion Function for Group A is represented as follows:
Predicted utility of health state = 0.01022159 * FT value of health state -0.0650056.

	Group B (PH-Dead Scale)		
Marker State	10% Trimmed Mean Value [3]	EOSBA-Adjusted Value	10% Trimmed Mean Utility [4]
MA	90.5	94.7	0.91
MB	57.9	50.0	0.49
MC	30.3	28.0	0.23
Pits	11.4	11.4	0.19

Model Fit Statistics: Adj. R^2 *=0.971; F(1,3)=103.52; p=0.009*

Utility Conversion Function for Group B is represented as follows:
Predicted utility of health state = 0.0090883 * FT value of health state + 0.03666548.

1. Person-Mean(A) Values on a Perfect Health=100 / Pits=0 scale, measured on Visual Analog Scale tasks using a Feeling Thermometer.
2. Person-Mean(A) Utilities on a Perfect Health=1.00 / Pits=0.00 scale, measured on Standard Gamble tasks using a Chance Board.
3. Person-Mean(B) Values on a Perfect Health=100 / Dead=0 scale, measured on Visual Analog Scale tasks using a Feeling Thermometer.
4. Person-Mean(B) Utilities on a Perfect Health=1.00 / Dead=0.00 scale, calculated from positive linear transformation of Standard Gamble tasks performed using a Chance Board originally obtained on a Perfect Health=1.00 / Pits=0.00 scale.

Table 16: Aggregated Utility Scores and MAUT Parameters

State		Weighted Person-Mean Utility[1]	Person-Mean Disutility[2]	Rescaled Person-Mean Utility[3]	Rescaled Person-Mean Disutility[4]
Physical Ability& Energy Level	**2**	0.905	0.095	0.840	0.160
	3	0.697	0.303	0.489	0.511
	4	0.408	**0.592**	0.000	1.000
Relationships	**2**	0.917	0.083	0.838	0.162
	3	0.732	0.268	0.478	0.522
	4	0.486	**0.514**	0.000	1.000
Mood & Feelings	**2**	0.911	0.089	0.854	0.146
	3	0.697	0.303	0.501	0.499
	4	0.394	**0.606**	0.000	1.000
Enjoyment of Diet	**2**	0.920	0.080	0.822	0.178
	3	0.714	0.286	0.359	0.641
	4	0.554	**0.446**	0.000	1.000
Satisfaction	**2**	0.920	0.080	0.788	0.212
	3	0.626	**0.374**	0.000	1.000

1. Overall Person-Mean Utility on a Perfect Health=1.00 / Pits=0.00 scale were calculated as [0.31 * Person-Mean(A) utility + 0.69 * Person-Mean(B) rescaled utility].
2. Overall Person-Mean Disutility (Disutility = 1 – Utility) on a Pits=1.00 / Perfect Health=0.00 scale. Bolded numbers represent the disutilities for the corner states (i.e. c_js).
3. Rescaled Overall Person-Mean Utility on a scale where the highest level on the attribute=1.00 and the lowest level on the attribute=0.00.
4. Rescaled Overall Person-Mean Disutility on a scale where the lowest level on the attribute=1.00 and the highest level on the attribute=0.00. These numbers represent the single-attribute level disutility scores, (i.e. $u`_j$s), that are used to calculate overall the DUI disutility score depending on the attribute level selected.

4.2.5 Calculation of MAUT Parameters

The parameters required to develop a utility scoring function for the DUI using the MAUT framework were obtained using a disutility approach, as discussed earlier. These parameters were five 'c_j's' (overall Person-Mean disutility scores for corner states of the five attributes of the DUI), a scaling parameter 'c', and a 15 '$u`_j$s' (single-attribute level disutility scores for every level of the five attributes of the DUI). The value of the c_j's would also determine the type of model to be employed within the MAUT framework, specifically whether a simple additive model would be adequate to develop the scoring function for the DUI.

The overall Person-Mean utility scores were converted to disutility scores. The disutilities for the five corner states were designated as c_j's (where j=1 through 5); these are represented in Column 3 of Table 16. Borrowing from Equation 3, the additive model would hold if the sum of the c_j's were equal to 1. Since the sum of the c_j's was 2.532, the additive model was rejected and the multiplicative model was chosen to describe the interaction between the five attributes of the DUI, and thereby calculate a utility scoring function for the DUI. A general multiplicative MADUF that provides disutilities on a Pits=1.00/ Perfect Health=0.00 scale was represented in Equation 4.

Towards fitting the MADUF, the value of the scaling parameter 'c' was calculated, by iteratively solving Equation 5, to be equal to -0.966. Additionally, the 'u_j`s' were to be calculated on a scale where the lowest level of an attribute had a disutility of 1.00 and the highest level of an attribute had a disutility of 0.00. The first step in this calculation was to convert the Person-Mean single-attribute level utility scores for each attribute from a Pits=0.00/ Perfect Health=1.00 scale into a scale where the lowest level of a given attribute=0.00 and the highest level of the attribute=1.00, as depicted in Column 4 of Table 16. These were converted to disutility scores (Column 5 of Table 16) on the desired scale, and were designated as u_j`s (where j=1 through 5).

4.2.6 Obtaining Utility Scores from DUI Responses

Overall Utility Function

Substituting the values of 'c', c_1, c_2, c_3, c_4, and c_5 discussed in the previous section into the general multiplicative MADUF illustrated in Equation 4, we get

$$\begin{aligned} \text{Disutility} = (1 / -0.966) * ((1 + (-0.966) * 0.592 * u_{attribute1}) &* \\ (1 + (-0.966) * 0.514 * u_{attribute2}) &* \\ (1 + (-0.966) * 0.606 * u_{attribute3}) &* \\ (1 + (-0.966) * 0.446 * u_{attribute4}) &* \\ (1 + (-0.966) * 0.374 * u_{attribute5}) -1). & \end{aligned}$$

Thus, an overall disutility (on a scale where PH=0.00 / Pits=1.00) can be calculated by substituting the appropriate values of $u_{attribute1}$ through $u_{attribute5}$ in the equation above, obtained from the u_j`s for each attribute (presented in Column 5 of Table 16), depending on the severity level for each attribute on the DUI. This disutility score, when subtracted from 1.00, yields the required overall utility score (on a scale where PH=1.00 / Pits=0.00). For example, the disutility for a health state represented as (3,3,4,3,2) can be calculated as:

$$\begin{aligned} \text{Disutility}_{PH=1.00 / Pits=0.00} = (1 / -0.966) * ((1 + (-0.966) * 0.592 * 0.511) &* \\ (1 + (-0.966) * 0.514 * 0.522) &* \\ (1 + (-0.966) * 0.606 * 1.000) &* \\ (1 + (-0.966) * 0.446 * 0.641) &* \\ (1 + (-0.966) * 0.374 * 0.212) -1) & \\ = 0.88; \text{ and} & \end{aligned}$$

$$\text{Utility}_{PH=1.00 / Pits=0.00} = (1 - 0.88) = 0.12$$

Single-Attribute Utility Functions

In addition to an overall function, the DUI provides single-attribute functions that indicate the preference associated with being on specific levels within an attribute. In disutility terms, they are the u_j`s (presented in Column 5 of Table 16), depending on the severity level for the DUI attribute considered, on a scale where the lowest level on an attribute has a disutility=1.00 and the highest level has a disutility=0.00. These disutility scores, when subtracted from 1.00, yield the required single-attribute utility scores (on a scale where the highest level=1.00 / lowest level=0.00). The five single-attribute functions, one for each DUI attribute, are summarized in Table 17.

MAUF on a PH-Dead Scale

Conventionally, utility scores are based on a scale where PH=1.00 and Dead=0.00. In order to transform the MAUF from the DUI-specific PH=1.00/Pits=0.00 scale into a conventional PH=1.00 / Dead=0.00 scale, the SG utility for Dead on a PH=1.00/Pits=0.00 scale was needed (Furlong et al., 1998). This was obtained by aggregating:

a. the 10% trimmed mean of directly measured SG utility for Dead (equal to 0.05, on a PH=1.00 / Pits=0.00 scale) from Group A respondents (n=31); and
b. the derived SG utility of Dead (equal to -0.11, on a PH=1.00 / Pits=0.00 scale) from Group B respondents (n=69), by linear transformation of the 10% trimmed mean of directly measured SG utility of Pits (equal to 0.10, on a PH=1.00 / Dead=0.00 scale).

The weighted average utility for Dead on a PH=1.00 / Pits=0.00 scale was calculated as 0.31 * (0.05) + 0.69 * (-0.11), and was found to be equal to -0.06, indicating that for the sample interviewed, overall, Dead was less preferable compared to Pits. An overall disutility score for a health state (on a scale where PH=0.00 / Dead=1.00) can be calculated as:

Disutility $_{PH=1.00\ /\ Dead=0.00}$ = Disutility $_{PH=1.00\ /\ Pits=0.00}$ / (1 – Utility of Dead $_{PH=1.00\ /\ Pits=0.00}$); and

Utility $_{PH=1.00\ /\ Dead=0.00}$ = 1 - Disutility $_{PH=1.00\ /\ Dead=0.00}$

For example, the disutility for a health state represented earlier as (3,3,4,3,2) can be calculated as:

Disutility $_{PH=1.00\ /\ Dead=0.00}$ = 0.88 / 1 - (-0.06) = 0.83; and

Utility $_{PH=1.00\ /\ Dead=0.00}$ = 1 – 0.83 = 0.17.

The calculations necessary to calculate the utility scores based on the DUI MAUF are summarized in Table 17.

Table 17: DUI Multi-Attribute Utility Function

Level	Physical Health & Energy Level u_1	Relationships u_2	Mood & Feelings u_3	Enjoyment of Diet u_4	Satisfaction with Management U_5
1	0.00	0.00	0.00	0.00	0.00
2	0.160	0.162	0.146	0.178	0.212
3	0.511	0.522	0.499	0.641	1.000
4	1.000	1.000	1.000	1.000	NA

MADUF: Formula for Disutility (PH=0.00 / Pits=1.00)

$$\text{Disutility}_{(PH=0.00\ /\ Pits=1.00)} = (1\ /\ -\ 0.966) * ((1 + (-0.966) * 0.592 * u_1) * (1 + (-0.966) * 0.514 * u_2) * (1 + (-0.966) * 0.606 * u_3) * (1 + (-0.966) * 0.446 * u_4) * (1 + (-0.966) * 0.374 * u_5) - 1)$$

MAUF: Formula for Utility (PH=1.00 / Pits=0.00)

$$\text{Utility}_{(PH=1.00\ /\ Pits=0.00)} = 1 - \text{Disutility}_{(PH=0.00\ /\ Pits=1.00)}$$

Formula for Disutility (PH=0.00 / Dead=1.00)

$$\text{Disutility}_{(PH=0.00\ /\ Dead=1.00)} = \text{Disutility}_{(PH=0.00\ /\ Pits=1.00)}\ /\ 1 - (-\ 0.06)$$

Formula for Utility (PH=1.00 / Dead=0.00)

$$\text{Utility}_{(PH=1.00\ /\ Dead=0.00)} = 1 - \text{Disutility}_{(PH=0.00\ /\ Dead=1.00)} = 1 - (\text{Disutility}_{(PH=0.00\ /\ Pits=1.00)}\ /\ 1 - (-\ 0.06))$$

4.3 Objective Three: Validating the DUI

4.3.1 Overview

A total of 396 usable responses were obtained to the Validation Survey, with a response rate of about 33%. About 52% of the respondents were female, over one-third were 65 years or older, while just under two-thirds did not have a college degree. The demographic profile of the respondents is summarized in Table 18.

The average age of respondents and duration of diabetes were 59 (±13.9) years and 12 (±11.5) years, respectively. About 80% of the respondents reported being Type 2, while 15% reported being Type 1. About 43% of the respondents reported currently using insulin overall; 36% of the respondents with Type 2 diabetes reported using insulin. The mean of average A1Cs in the past year for respondents was 7.5 (±1.6), and more than 70% had at least one diabetes-related complication in the past year, from analysis of their ICD-9 diagnosis codes. About 91% of the respondents self-reported that they did not have either an ER visit or hospitalization related to diabetes during the previous year, with the mean number of physician-office visits related to diabetes being 3 (±2.1) visits during the previous year. The average BMI was 32.4 (±7.7), with about 80% of the respondents being either overweight or obese.

Table 18: Socio-demographic Profile of the Validation Survey Respondents

	n	**%**
Gender		
Male	191	48.2%
Female	204	51.5%
Race		
White	380	96.0%
Black	9	2.3%
Other	5	1.3%
Marital Status		
Single	42	10.6%
Married	263	66.4%
Divorced/ Separated	30	7.6%
Widowed	43	10.9%
Not married/ living with partner	14	3.5%
Education		
High school or less	47	11.9%
High school graduate/ GED	125	31.6%
Some college/ vocational college	84	21.2%
College degree	69	17.4%
Graduate / professional degree	70	17.7%
Employment		
Employed	163	41.2%
Retired	158	39.9%
Retired – Disability	14	3.5%
Home-maker	38	9.6%
Student	9	2.3%
Other	13	3.3%
Household income		
$25,000 or less	128	32.3%
$25,001 - $50,000	105	26.5%
$50,001 - $75,000	77	19.4%
More than $75,000	55	13.9%
Refused	31	7.8%
Total N	**396**	**100%**

4.3.2 Distribution of DUI Scores

The scores reported for the DUI are on the PH=1.00 / Dead=0.00 scale. The DUI scores calculated from the validation survey responses ranged from 0.00 to 1.00. The mean DUI score was 0.61 (±0.24) and the median score was 0.66. The histogram of DUI scores (Figure 7a) indicates a relatively flat distribution that is skewed to the left (skewness and kurtosis were calculated to be -0.42 and -0.79 respectively). From the Normal Q-Q and box plots (Figures 7b and 7c), it can be inferred that the distribution is light-tailed, indicating less severe deviation from normality.

Rasch analysis conducted on data obtained on the endorsement of DUI severity levels from the Validation Survey indicated a Person-reliability, an equivalent of the classical 'test' reliability, of 0.71 (n=396). This statistic provides information on the number of levels to which the DUI is able to discriminate the sample into levels on a linear continuum of HRQoL, which would be about two to three for the DUI based on a person reliability of 0.71. Another Rasch reliability statistic, the 'item-reliability' (no equivalent in Classical Theory) was calculated to be 0.96 for the DUI (n=396). The fit of individual attributes (items) of the DUI to the Rasch model and the Rasch rating scale diagnostics have been discussed in an earlier section.

Figure 7: Distribution of DUI Scores

Figure 7a: Histogram

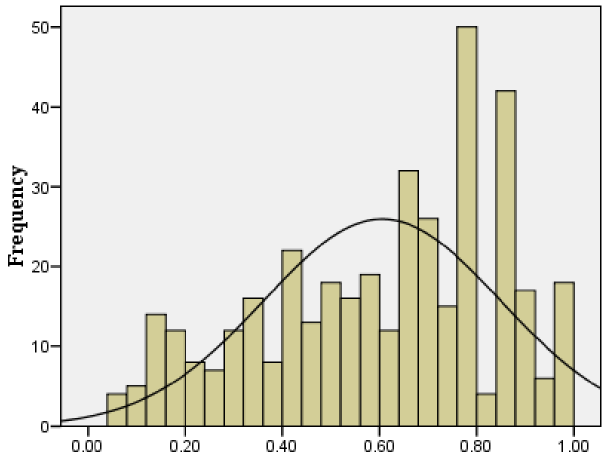
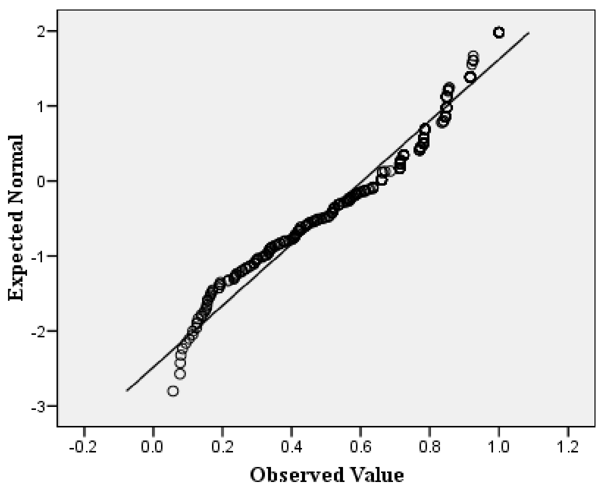
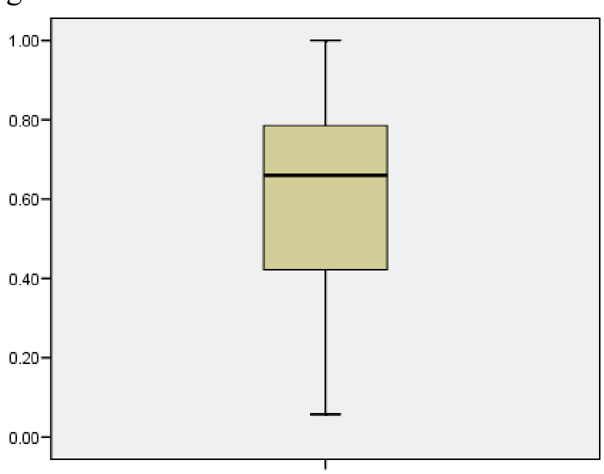


Figure 7b: Normal Q-Q Plot



This Q-Q normality plot of DUI scores shows no outliers and an S-shaped curve, with the upper end bending above a hypothetical straight line passing through the main body of the X-Y values of the plot, and the lower end bending below the line. This suggests a light-(left) tailed distribution.

Figure 7c: Box Plot



This box plot of DUI scores shows no outliers, with the tails of the box short relative to the height of the box. The mean value is below the median (the center line in the box), the median line does not evenly divide the box, and the lower tail of the box plot is longer than the upper tail. This suggests a light-tailed distribution.

4.3.3 Examination of DUI Severity Categories and Attribute-level Scores

Scores on relevant patient-reported outcomes measures were compared across the severity categories of the DUI attributes and with its single-attribute utility scores. The SF-12 Physical Component Summary (PCS-12) scores were significantly lower within higher severity groups on the Physical Ability & Energy level attribute, $F_{(3,392)}$=111.75, p<0.001. The correlation between the single-attribute utility scores for this attribute and PCS-12 scores was 0.65 (p<0.001). Those reporting higher severity groups on the Mood & Feelings attribute had significantly lower SF-12 Mental Component Summary (MCS-12) scores, $F_{(3,392)}$=82.1, p<0.001, and total Well-Being Questionnaire (WB-Q12) scores, $F_{(3,392)}$=95.3, p<0.001. The correlations for the single-attribute scores for this attribute were: 0.59 (p<0.001) with MCS-12 scores, 0.61 (p<0.001) with Total W-BQ12 scores, 0.48 (p<0.001) with the Positive well-being subscale scores, 0.52 (p<0.001) with the energy subscale scores, and -0.572 (p<0.001) with the negative well-being subscale scores. The Diabetes Empowerment Scale Short-Form (DES) scores were significantly lower within higher severity groups on the 'Satisfaction with managing changes' attribute, $F_{(2,392)}$=45.25, p<0.001. The correlation between the single-attribute scores for this attribute and DES scores was 0.38 (p<0.001).

4.3.4 Construct, Concurrent, and Predictive Validation of the DUI Utilities

The concurrent validity of the DUI was evaluated by examining the relationship between the DUI and other generic PBMH. The correlations of the DUI scores with generic PBMH were moderate: 0.62 ($p<0.001$) with SF-6D scores, 0.60 ($p<0.001$) with EQ-5D scores and 0.66 ($p<0.001$) with the derived HUI-3 scores. The correlations are summarized in Table 19.

As evidence of the construct validity of the DUI, its scores were examined against indicators of disease severity. The correlation between DUI utilities and the average of A1Cs in the past one year was small, but significant ($r=-0.30$, $p<0.001$). The correlations with the Diabetes Symptoms Checklist-Revised (DSC-R) subscale scores (see Table 19) were as follows: hyperglycemic -0.47; hypoglycemic -0.55; psychological -0.64; cardiovascular -0.47; neurological -0.58; and ophthalmologic -0.51; all statistically significant ($p<0.001$). The negative sign of these correlations indicate that lower DUI utilities are associated with higher diabetes symptom scores.

The DUI scores significantly differed between categories based on the number of diabetes-related complications, calculated on the basis of prevalent ICD-9 diagnosis codes in the last one year, $F_{(2, 389)}=9.682$ ($p<0.001$). Specifically those with two or more complications had significantly lower DUI scores as compared to those with none ($p<0.001$) or one complication ($p=0.015$). Table 20 reports the DUI scores observed within sub-categories based on clinical and medical history variables.

The DUI scores predicted by the MAUF were compared with the SG scores directly obtained for the three marker states. The MAUF underestimated SG utilities, on an average by 0.05, with the closest prediction observed for the more severe state. These differences are represented in Table 21.

There was only a small relationship between the DUI scores and age ($r=0.14$), and there was no significant difference in scores between men and women. Those without a college degree, however, had significantly lower DUI scores than those with a college degree or beyond ($t=-3.76$, $p<0.001$). Further analysis revealed that those without a college degree also had significantly higher diabetes symptom burden, as evidenced from higher scores on the DSC-R ($t=5.53$, $p<0.001$). The relationship between DUI scores and education was found to be not significant when controlled for by DSC-R scores, $F_{(2,392)}=0.232$, $p=0.630$.

Table 19: Correlations with the DUI Utilities

	Correlations [1]					
	DUI [2]	**Single-Attribute Utility Scores**				
		Physical Ability & Energy Level	**Relation-ships**	**Mood & Feelings**	**Enjoyment of Diet**	**Satisfaction with Managing Changes**
Total DSC-R [3]	-0.670	-0.649	-0.490	-0.553	-0.530	-0.358
Hyperglycemic	-0.468	-0.412	-0.308	-0.339	-0.411	-0.274
Hypoglycemic	-0.550	-0.384	-0.336	-0.561	-0.432	-0.303
Psychological	-0.643	-0.602	-0.434	-0.568	-0.475	-0.330
Cardiovascular	-0.472	-0.469	-0.346	-0.371	-0.428	-0.237
Neurological	-0.582	-0.614	-0.449	-0.445	-0.458	-0.324
Ophthalmologic	-0.509	-0.564	-0.475	-0.424	-0.380	-0.250
EQ-5D [4]	0.596	0.666	0.465	0.521	0.381	0.231
SF-6D [5]	0.623	0.608	0.456	0.550	0.381	0.275
HUI-3 Derived [6]	0.660	0.670	0.475	0.568	0.424	0.311
SF PCS-12 [7]	0.571	0.653	0.461	0.395	0.354	0.240
SF MCS-12 [8]	0.574	0.456	0.350	0.589	0.401	0.312
DES [9]	0.470	0.281	0.268	0.354	0.305	0.384
Total W-BQ12 [10]	0.641	0.521	0.371	0.614	0.449	0.373
Positive Well-Being	0.522	0.420	0.274	0.481	0.356	0.305
Energy	0.592	0.522	0.377	0.519	0.417	0.342
Negative Well-Being	- 0.530	- 0.396	- 0.302	- 0.572	- 0.378	- 0.309

1. All figures reported here are Pearson's correlations. The correlations were all found to be significant ($p<0.001$).
2. DUI = Diabetes Utility Index scores on a PH=1.00/Dead=0.00 scale.
3. DSC-R = Diabetes Symptoms Checklist – Revised total scores followed by subscale scores standardized on a 1-100 scale.
4. EQ-5D = EuroQoL EQ-5D index scores.
5. SF-6D = Short-Form 6D scores.
6. HUI-3 Derived = Health Utilities Index (HUI) Mark 3 scores derived from SF-12 items.
7. SF PCS-12 = SF-12 Physical Component Summary scores on a 1-100 scale
8. SF MCS-12 = SF-12 Mental Component Summary scores on a 1-100 scale
9. DES = Diabetes Empowerment Scale Short Form scores.
10. W-BQ12 = Well-Being Questionnaire (12-item) total scores followed by subscale scores.

Table 20: Construct Validation of the DUI Utilities

Variable	Physical Ability & Energy Level [1]	Relation-ships [1]	Mood & Feelings [1]	Enjoyment of Diet [1]	Satisfaction with Managing Changes [1]	DUI [2]
Insulin Use						
1. Insulin user	0.79	0.69	0.71	0.71	0.72	0.54
2. Insulin non-user	0.87	0.83	0.78	0.77	0.74	0.65
p-value [3]	p=0.114	p=0.002	p=0.025	p=0.058	p=0.004	p<0.001
Glycemic Control						
1. A1C < 7.0	0.86	0.83	0.78	0.77	0.79	0.66
2. A1C ≥ 7.0	0.81	0.73	0.73	0.71	0.68	0.56
p-value [3]	p<0.001	p<0.001	p=0.001	p=0.034	p=0.529	p<0.001
Diabetes Complications						
1. None	0.90	0.82	0.80	0.80	0.77	0.67
2. One complication	0.86	0.80	0.77	0.75	0.73	0.63
3. Two or more	0.76	0.71	0.71	0.70	0.72	0.54
p-value [4]	p<0.001 (1 Vs 3) p=0.001 (2Vs3)	p=0.003 (1 Vs 3) p=0.023 (2Vs3)	p=0.001 (1 Vs 3)	p=0.018 (1 Vs 3)	-	p<0.001 (1 Vs 3) p=0.015 (2Vs3)

1. Single attribute utility scores are calculated on a scale where the lowest level on this attribute has a utility of 0.00 and the highest level has a utility of 1.00.
2. DUI utility scores are calculated on a scale where the Perfect Health has a utility of 1.00 and Dead has a utility of 0.00.
3. p-value for t-test (for variables with two categories)
4. p-value for post-hoc ANOVA test (for variables with more than two categories). Non-significant differences between sub-groups are not reported.

Table 21: Predicted Validity of the DUI MAUF

	Marker State A	Marker State B	Marker State C	Average
DUI Disutility [1]	0.16	0.63	0.88	-
DUI Utility	**0.84**	**0.37**	**0.12**	**-**
SG Utility(A) [2]	0.93	0.43	0.13	-
SG Utility(B) [3]	0.90	0.43	0.14	-
SG Utility [4]	**0.91**	**0.43**	**0.14**	**-**
Difference	-0.07	-0.06	-0.02	-0.05

1. DUI MAUF predicted utility on a PH=1.00/Pits=0.00 scale.
2. Group A Standard Gamble 10% trimmed mean utility on a PH=1.00/Pits=0.00 scale.
3. Group B Standard Gamble 10% trimmed mean utility on a PH=1.00/Pits=0.00 scale.
4. Overall weighted average Standard Gamble 10% trimmed mean utility on a PH=1.00/Pits=0.00 scale.

A hierarchical regression model was constructed in order to explain the relationship between DUI utilities and the various demographic, clinical, and medical history predictors. The model was checked for violations of regression assumptions, including for the presence of multi-collinearity that was expected in view of the multiple indicators of disease severity and co-morbidity that were available from the validation survey component of this study. During this process of diagnostic checks, the variables representing 'index of complications' and 'BMI' had to excluded in order to avoid multi-collinearity. The remaining variables were entered sequentially into the model in the following order:

Block 1: Average A1C, Insulin use (categorical), Duration of diabetes, Diabetes symptom burden (i.e. DSC-R total score);

Block 2: Index of co-morbidity (i.e. Charlson co-morbidity score); and

Block 3: Demographic variables: Age, Gender and Education (all categorical).

Wherever indicated, the categories were designated as indicated in Table 3. Since indicators of co-morbidity and diabetes severity were retained in the model, the exclusion of the two variables was not expected to lead to a loss in model specification. Although 'average A1C' was retained, 'insulin use' was included since the variable was expected to additionally represent the expected burden resulting from the administration of injectible insulin. The resultant model had an adjusted R^2 of 49% ($F_{(8,291)} = 36.8$; $p<0.001$). Those with a greater burden of diabetes symptoms ($p<0.001$), indicated by higher DSC-R scores, those with higher A1C levels ($p=0.041$), and those using insulin ($p=0.05$) were found to have significantly lower DUI utility scores, as summarized in

Table 22. This indicates that insulin use had an effect on DUI utility scores independent of the effect of diabetes symptoms and glycemic control.

Table 22: Results of the Hierarchical Regression Model Predicting DUI Utilities

	Predictor Variable	**Beta**	**Std. Error**	**Test statistic (t)**	**Sig. p**
1.	Insulin Use	0.048*	0.025	1.97	0.050
2.	Diabetes duration	0.000	0.001	0.15	0.882
3.	Average A1C	-0.014*	0.007	-2.05	0.041
4.	Diabetes Symptoms Burden	-0.008*	0.001	-13.63	< 0.001
5.	Index of Co-morbidity	0.000	0.004	0.09	0.926
6.	Age in years	0.000	0.001	0.44	0.660
7.	Gender	-0.031	0.021	-1.49	0.137
8.	Education	0.020	0.022	0.89	0.374

* *significant at 0.05 level*
Model fit statistics:
$R^2 = 0.71$, adjusted $R^2 = 0.49$, $F(8, 291) = 36.8$, $p < 0.001$

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS

5.1 Objective One

The health state classification system of the DUI represents an effort in the development of a diabetes-specific PBMH involving inputs from theory, an existing diabetes-specific measure of quality of life, a clinical expert panel, primary data collection, and statistical analyses. Obtaining input from the descriptive system offered by an existing valid and reliable measure avoided the need to perform extensive research on content identification and item pool composition. In addition to the classical approach to reviewing the item content of the existing measure, the study incorporated techniques based on the Modern Test Theory as well as subjective approaches based on expert review.

In order to develop a disease-specific measure, it is necessary to build a pool of items that reflect the areas of impairment caused by the disease. Type 1 and Type 2 diabetes are chronic illnesses that can influence QoL because the treatments are burdensome and the complications can be debilitating and/or life-threatening. Patients have to deal with their diabetes almost every instant of their life and have to make continuous decisions that interfere with living a normal life. Patients with diabetes may feel overwhelmed by the management of the disease that requires them to make diet and lifestyle changes as well as taking medications (oral, injectible, or both). Apart from the emotional and social burdens this may cause, patients with diabetes face the acute physical distresses of hypoglycemia or hyperglycemia and chronic physical distress of diabetes-related complications.

At the outset, it was decided that the DUI was not intended specifically for either Type 1 or Type 2 diabetes patients, and that the out-of-skin focus would be adopted to ascertain items for the DUI given the broad range of impairments in life due to the nature of diabetes and its management. Also, this implies that the scope of the DUI would be beyond the 'within the skin' approach of a generic PBMH (that omits social interaction) like the HUI-3 (Feeny et al., 2002) , and would not limit its focus to the measurement of distresses due to disease (diabetes)-related symptoms. The latter approach was adopted with disease specific PBMH like the ASUI (Revicki et al., 1998a) and the RSUI (Revicki et al., 1998b) whose scores were not suitable for calculating QALYs. Polonsky's (2000) guidance in this regard was a starting point in the identification of relevant constructs that ought to be measured using the out-of-skin focus in the assessment of the impact of diabetes. Keeping in mind the nature of diabetes and its management, it was important to build context-specificity to the DUI in order to estimate the true impact of the disease.

The next step involved the identification of an item pool in order to measure these constructs. Structurally, profile-based measures are similar to PBMH, in that they contain items representing various domains that are important in the measurement of the overall construct they measure. Responses of patients with Type 2 Diabetes to the Audit of Diabetes Dependent Quality of Life (ADDQoL), available from our previous research (Sundaram et al., 2006), were explored as a first step towards identifying items. The study utilized a dataset containing responses to the ADDQoL, a profile-based diabetes-specific QoL measure. In addition to the favorable psychometric properties of the measure, the instrument was considered because of its applicability to both Type 1 and Type 2 diabetes patients, as well as because a dataset with previously collected responses to this measure was readily available. In the process of developing the item pool, input from both statistical analysis as well as a diabetes clinical expert panel was obtained. Thus, the diabetes classification system borrowed from subjective input as well and was not based on the content of the ADDQoL alone.

The purpose of using Classical Methods, (including used of Exploratory Factor Analysis) was not limited to selecting (or excluding) ADDQoL items. The intention was to rather explore the possibility of measuring the constructs described by Polonsky in a dataset consisting of responses of patients with diabetes to items that can be relatable to those constructs. The Classical statistical criteria we employed involved item exclusion and forced three factor solution in order to gain a better understanding of the relationships between ADDQoL items and not to propose an alternate structure for the instrument. Two main DUI attributes could be ascertained in the process – 'Relationships' and 'Enjoyment of Diet', with other items potentially describing broad functioning- related aspects in the lives of patients with diabetes.

The Diabetes Clinical Expert Panel was diverse in the area of expertise of its members. Subjective input obtained from the panel helped better understand the nature of the impact of diabetes on aspects of functioning, relationships, and enjoyment of diet, and provided additional areas to consider in the measurement of the impact of diabetes. The attributes of 'Physical Ability & Energy Level', 'Mood & Feelings', and 'Satisfaction with the Management of Diabetes' were identified in the process. The subjective input gathered from the panel formed the basis of explaining the scope of the DUI attributes in the form of actual easy to understand descriptions so as to avoid ambiguity in the minds of the respondents. It was hoped that including these explanations in the DUI would enhance the validity of assessments. Additionally, input from the panel provided initial suggestions on the rating scale for each selected attribute. The intended approach was to avoid generic rating scales with ambiguous interpretation that may be subject to biases.

The five selected attributes were tested in three successive pilot rounds for fit to the Rasch model. The technique is robust to missing data, in comparison to CTT where data have to either be discarded or imputation employed. This held special promise because of the smaller sample sizes expected in the pilot rounds of testing the DUI. Limiting the number of attributes to five helped in the design of health states that could be feasibly presented to the interview participants in Objective Two of this research. In general, people are said to be able to process a limited number of concepts at a time (seven, plus or minus two) and therefore, parsimony in health state classifications systems is desirable (Feeny, 2002a). Fewer attributes also perhaps works well with the over-the-skin nature of an instrument's content since the chances of overlap, and consequentially, the risk of dependence between attributes, is lower.

Indeed, an important consideration in the development of the DUI was the independence between its attributes. The decomposed approach to estimating a preference scoring function essentially stipulates the measurement of the preference of individual attributes of a classifications system and provides a framework for estimating the overall preference of a health state in terms of the preferences of its constituent attributes. Hence, it is important to be able to measure the unique contribution of individual attributes to the overall preferences of health states (Feeny, 2002a). As can be understood from the experience of the development of the HUI-2, dependent attributes complicate the process of preference elicitation when the decomposed approach is the approach chosen (Torrance et al., 1996).

Dependence between items of an instrument can be ascertained by simply understanding the nature of the items, from knowledge of theory, as well as from statistical input. Rasch Analysis fit statistics employed on data from pilot versions of the DUI indicated the problem of overfit of the attribute ‘freedom living life’ because of its dependence on the ‘physical ability’ attribute, which in retrospect could be naturally expected. The attribute was replaced with a related concept of ‘lifestyle freedoms’ in order to tap patients’ perception of the impact diabetes has on account of the major lifestyle changes that the condition imposes, but the problem of dependence persisted. This problem could have potentially led to difficultly in having the interview respondents view a corner state describing full ‘lifestyle freedom’ in the face of severely impaired physical health. While the final set of DUI attributes did not theoretically indicate a definitive threat of dependence, Rasch fit statistics obtained from Pilot 3 as well as the Validation Survey do not suggest otherwise.

Potential Limitations and Sample Size Considerations

While the initial pool of items considered for the description of DUI attributes were based on the structure of existing PBMH, theory on PRO evaluations in diabetes, and an existing diabetes-specific measure of QoL, the diabetes clinical expert panel additionally provided a qualitative and critical review of this pool as well as suggested additional items that were not represented in the pool. This sought to minimize any bias that may have been introduced by the exclusion of relevant and important items. The sample size of the dataset used in the determination of the initial plausible attributes was adequate for performing factor analyses, with more than the suggested 15 subjects per factor. The Rasch pilot samples were smaller in relation to those conventionally reported as required to obtain stable parameters. Due to time and resource constraints, we were unable to pursue larger samples in the pilot rounds, but we were successful in obtaining data on pilot versions of the DUI and in diagnosing issues with the structure of the DUI on the basis of the Rasch model parameters. Our final validation survey had a sufficient sample size to interpret the Rasch output with confidence, and confirmed some of the decisions that were taken by interpreting the output from the pilot rounds. Since Rasch Analysis is considered to be free of sample effects, the results discussed here can be considered to be mostly free of the biases that would normally be associated with the usage of samples that were not geographically diverse- in our case, the use of samples located for the most part in N-Central WV.

5.2 Objective Two

There is no consensus on which approach to choose for the purpose of modeling health state valuations because each has a number of advantages and disadvantages. The scoring function for the DUI was based on the principles of a well-established theory, the MAUT, which permits certain assumptions to be made regarding the relationship between the DUI attributes. The ability of the technique to enable the determination of a scoring function for the DUI using fewer valuations was an important determinant of the choice of the MAUT. The adopted Person-mean approach to summarizing the valuations obtained in the interviews allows the estimation of a MAUF that provides utility estimates for the general diabetes population. The MAUT approach also afforded the ability to estimate single-attribute functions for the five attributes of the DUI.

The preference-elicitation interviews were expected to be a subjective and challenging exercise, given the length of the interviews and the nature of the tasks involved. The participant reward, among others, was designed to evoke a satisfactory level of interest from the participants in completing the interviews in a satisfactory manner. All the interviews were conducted by one trained person who was also technically aware of the principles of preference elicitation and the MAUT. While the possibility of interviewer bias due to the selection of one interviewer alone is not ruled out, the qualifications of the lone interviewer as well as participant feedback seems to allay these concerns - about three-fourths found the FT tasks and the CB tasks to be either easy or very easy to understand. While more participants found the CB to be difficult than those feeling the same about the FT, this was anticipated given the comparatively added cognitive burden of the CB. On the other hand, the interviewer reported that only four percent of the respondents had a great deal of trouble in answering the questions.

The structure of the interview and the interview props were based on detailed technical reports published by the HUI group (Furlong et al., 1990; Furlong et al., 1998), with added consultations available from the group via phone and email. The HUI group is reputed to be one of the pioneers in the design and development of modern health state preference instrumentation techniques. The interview props, including the FT, CB and health state cards were produced by a professional graphic design team using the design suggested in the HUI manuals but making agreeable improvisations to the colors and materials used. The interviews began with the administration of the DUI survey and a prior verbal explanation of the DUI attributes supplemented with a pictorially referenced attribute sheet. This was intended to help reinforce the idea of the constituents of each attribute in the minds of the respondents, and it is expected that this step in the interview process aided the participants in taking part in the FT and CB tasks. Also, completing the survey made participants familiar with the DUI classification system, especially the sentences used to describe the severity levels of the five attributes, as well as provided them an internal reference with which to compare other health states introduced during the course of the interview.

In the interest of reducing the cognitive burden of the interviews, SG tasks were performed only for the Marker States, while all health states including the Marker States were valued on the FT. Though not the best possible approach, this method has been considered to be reasonable, practical (Torrance et al., 2001), and has been used in the determination of preference-scoring functions of other PBMH using the MAUT approach (Torrance et al., 1996; Feeny et al., 2002; Revicki et al., 1998a; Revicki et al., 1998b). Using a combination of VAS and SG tasks in the manner described above, however, necessitates the use of a value-to-utility conversion function. A study-specific conversion function was used, derived separately for the two sub-groups considered for the purpose of analysis, rather than use the functions reported in the literature for generic populations. With adjusted R^2s in excess of 90%, the regression-based conversion functions for both groups showed good ability to predict SG cardinal utilities from the values obtained in FT-based tasks. The conversion functions thus allowed the calculation of SG utilities for all the health states required for the estimation of a MAUF for the DUI, thereby imparting interval properties to the DUI utilities. The study-specific maker states described in order to obtain data for the conversion also functioned as a set of internal references, never removed from the FT board, against which participants could rate the other health states included in the interview.

From the value of the sum of the disutilities of the corner states obtained during the preference interviewers, it was determined that the DUI attributes were complements, exhibiting an antagonistic disutility (or a synergistic utility) interaction, as has been reported with the HUI-1 and HUI-2 (Furlong et al., 1998; Feeny et al., 2002). This means that the disutility of disabilities in two DUI attributes together is less than the sum of the disutility of disability in the two DUI attributes separately. Also, the sum of the utility added by restoring ability in two DUI attributes will be less than the utility added by restoring the abilities in both. Since the DUI attributes are preference complements, they work in tandem; either one alone is not that valuable, but together they are very valuable. This type of preference structure is indicative of multi-attribute risk-seeking (MARS) (Furlong et al., 1998).

There was no strong a priori expectation that the DUI attributes would be better suited to be fitted using multi-linear MAUF models, in which case an econometric approach would be a better modeling approach than the MAUT (Feeny, 2002b). From the data gathered during the preference elicitation interviews, a multiplicative MAUF was fitted for the DUI because the additive model had to be rejected. This result was expected since healthcare applications of the MAUF have seldom found the additive model to be appropriate. This indicates that in process of calculating the utility for a DUI health state, the preference interactions between attributes cannot be ignored. Although the multiplicative model chosen to fit the MAUF allows for interactions between the DUI attributes, it places restrictions in that it only allows for a certain type of preference interaction between all the attributes. While multi-linear MAUF models do not place this restriction, comparative studies show that the multiplicative models not only capture the important interactions, but also fare better in predicting utilities for health states using the directly-measured SG utilities for those heath states as a benchmark (Furlong et al., 1998).

The majority of the interview respondents indicated 'Dead' to be their worst state. This choice is generally study-specific and should hence should be understood in relation to the description of the DUI-defined 'Pits' state. Evidently, most respondents found the DUI Pits state to be preferable to Death when asked to make a choice. Although the respondents were classified into groups based on their choice of the worst health state during both the conduct of the interviews as well as the analysis of the interview data, the data were aggregated on the PH-Pits scale for the purpose of calculating the necessary metrics to fit a MAUF.

The DUI MAUF is capable of predicting utilities for all the health states defined in the geometric space bound by its five attributes. Dead is undefined in the DUI system as is the case with other PBMH, while the 'Pits' states defined by PBMH are not standardized but are specific to classification systems. The aggregated SG utility for Dead, obtained directly for Group A respondents and calculated for Group B respondents facilitated the calculation of the DUI utilities in the more traditional PH-Dead scale, such that these utilities could be used to calculate QALYs. Since the majority of the respondents found Dead to be the worst state, the modified DUI MAUF on a PH=1.00/Dead=0.00 scale calculates the utility of the Pits state to be greater than zero.

The utility scores provided by the DUI thus meet the three criteria – they convey preferences of users of healthcare, they are measured on an interval scale, and are calculated on a PH=1.00/Dead=0.00 scale.

The results of the preference interviews indicate the feasibility of conducting preference interviews for health states composed of ‘out-of-skin’ aspects of life that are influenced by diabetes and its treatments. Additionally, these results suggest the practicality of the decomposed approach in formulating a scoring function for a disease-specific PBMH not based entirely on disease-related symptoms. The degree of co-operation provided by the respondents indicates that the health states, described in accordance with the requirements of fitting a MAUF based on the MAUT, were reasonably conceivable and that it was possible to elicit both VAS values and SG utilities for these health states.

Sample Size Considerations

While it would be preferable to calculate a MAUF from a larger sample, the respondent pool of 100 persons with diabetes seems adequate in relation to other studies reporting the development of PBMH using the MAUT approach. More importantly, all the valuations tasks required under MAUT were conducted for all the participants, rather than allocating a sample of the total number of tasks to each participant. Hence, every health state had 100 valuations, although the valuations were categorized into Groups A and B in the ratio of 31:69 for calculation purposes. This categorization could not be ascertained beforehand since it is dependant on participants’ responses during the interview. At final count, it was not feasible to conduct any further number of interviews to boost the sample size of Group A, if at all the increase was warranted for any reason.

Given the non-acute nature of the description of the Pits health state, it can be reasonably expected that the proportion of participants indicating Death to be a worse outcome in comparison to Pits would always be more.

Potential Limitations

Preferences are considered to be heterogeneous. Hence, a potential limitation of the MAUF derived in the study is its generalizability to populations in other geographical regions and cultures. The recruitment efforts were concentrated more at outpatient settings than community settings, though it is expected that a reasonable number of the 100 participants were people with diabetes who came to know about the interviews through our dissemination efforts in the general community. While the sample was more or less balanced across age-groups, more than half of the participants lacked a college degree or came from households earning less than $50,000 a year, and more than two-thirds were women. No clinical information could be collected during the course of the interview to be able to compare with the answers to the interview questions, but it is reported that lower income and less educated populations tend to exhibit poorer control of their diabetes, leading to worse health outcomes.

5.3 Objective Three

Rasch analysis conducted on data from the Validation Survey indicated a lower than expected Person-Separation of 1.55. However, the Person-Reliability, an equivalent of the classical 'test' reliability, for this brief five-item measure was acceptable at 0.71. Larger Rasch Person reliabilities result from respondent samples with a greater range in abilities on the construct being measured, better test targeting, greater number of items on the instrument and longer rating scales. For reasons of maintaining parsimony as discussed earlier, the number of attributes selected for inclusion in the DUI seems adequate. Further increases in the length of the DUI severity rating scales also seem unnecessary at this point since doing so may result in the inclusion of levels that lie insufficiently endorsed or are not sufficiently differentiable by respondents, leading to arbitrary levels being chosen. Even the three severity categories for the Satisfaction attribute seem to provide a theoretically reasonable range of measurement. With the severity levels presented in the form of statements, it was expected that there would be less ambiguity in the minds of respondents when choosing one statement for each attribute. Rasch rating-scale diagnostics indicated that these statements were indeed reasonably well-functioning.

Within the constraints of the number of attributes and severity levels for the attributes discussed above, a higher Person reliability for the DUI may be obtained if it were administered to a diabetes population widely varying in the extent of the severity of diabetes and its complications. While one-third of the respondents to the survey had no complications, another third had only one diabetes-related complication with the mean DSC-R standardized score across all the respondents being 20.7. Another Rasch reliability, the 'item reliability' is independent of the length of the instrument and was calculated to be 0.96 for the DUI. While high Item reliabilities can be obtained in larger sample sizes, they are also indicative of a larger range of item difficulty. The Item-Separation and Item-Reliability values together indicated that the five-item DUI covered a satisfactory range in the measurement of the impact of diabetes.

The scores on relevant patient-reported outcomes measures like the SF-12, W-BQ12, DSC-R and DES were different across the severity categories of the DUI attributes. The results indicate that the rating scale statements were differentiable and also represented validated increases (or decreases) in the abilities on the construct (the attribute) they represent. The DUI utility scores showed moderate correlations with other generic PBMH scores (r=0.6) like the EQ-5D, the SF-6D, and the derived HUI-3 that, in turn, showed moderate to high correlations among themselves. While this indicates the concurrent validity of the DUI, the lack of large correlations with generic PBMH can be expected due to the differences in content arising out of the disease specificity of the DUI.

The DUI scores were largely free of socio-demographic effects since they were not influenced by gender and age. The lower DUI scores among respondents with lower education levels can be explained by the greater diabetes symptom burden observed in this group, as seen from the DSC-R scores. The instrument also performed well in relation to disease severity indicators – the number of complications and DSC-R scores. Differences were observed in DUI scores between those with two or more complications and those with either no or one complication. The differences in DUI utilities between those with none and one complication were non-significant, as was the case with DSC-R scores, leading to the reasonable conclusion that those with one diabetes-related complication were relatively asymptomatic and hence did not report significantly different scores from those with none. While the correlation with average A1C levels in the past one year was higher and significant with the DUI utility scores than with generic PBMH, the magnitude of this correlation itself was small and can be interpreted in light of the suggested small magnitude of association between a biomedical marker such as A1C and humanistic outcome such as QoL (Lau, Qureshi, & Scott, 2004; Sundaram et al., 2006). On the other hand, those respondents with A1Cs below the American Diabetes Association-suggested level of 7.0 were found to have significantly higher DUI utility scores.

The DUI MAUF was used to determine the utility scores for respondents to the Validation Survey, in which a large population of people with diabetes answered the DUI questions. Not all the health states that can be described on the DUI were directly valued using the SG technique, which can be considered to be the closest there is to any available gold-standard for comparison with the MAUF predicted utilities. Since the DUI MAUF will be frequently used to generate utilities for health states that were not directly measured, an assessment of this out-of-sample prediction will influence the degree of confidence placed in the scores calculated from the function. The scores predicted by the DUI MAUF for the three Marker States were compared to the 10% trimmed means of the directly derived SG scores for the respective states. The results of this internal validity check show that while the MAUF slightly under-estimates SG utilities for mild and moderate health states, the differences are very small for relatively severe health states. The mean absolute difference from SG utilities was 0.05, which was equal to the mean differences in scores and is hence not misleading as would be the case due to possible offsets of under-estimated utilities by over-estimated utilities. Unfortunately, only three health states provided data for this comparison.

While interpreting this difference, it is important to understand that the DUI MAUF was derived from Person-Mean preferences and that it was designed to provide utility scores for a broad range of health states that can be defined from its classification system. These preferences were obtained from individuals with diabetes rather than from a general population because of the need to build context specificity within the preference scores. While members of the general population can be considered to provide more objective evaluations without any bias, their preferences, however, may not be able to adequately reflect the unique decisions that patients with diabetes have to permanently make in their every day lives in managing diabetes.

Sample Size Considerations & Potential Limitations

The sample sizes that were available to conduct the various validation analyses were adequate. The validation results presented here are based on the analysis of responses obtained from a diabetes outpatient population in N-Central WV, and hence may not be generalizable. The survey respondents had a relatively good level of glycemic control overall, with an average of A1C in the past year equal to 7.5 (±1.6) and more than 90% self-reporting not having either an ER visit or hospitalization related to diabetes during the previous year. While more than 70% had at least one ICD-9 diagnosis code related to a diabetes complication, the mean DSC-R score (standardized to a high score of 100 that corresponds to the greatest burden of the symptoms of diabetes) was 21, indicating that the respondents of the validation survey had a low diabetes burden overall. On the other hand, about 80% were either overweight or obese. Our respondents were overwhelmingly white, with more than two-thirds not having a college degree. The performance of the DUI in populations with a more diverse racial mix that has better education and greater burden of diabetes symptoms needs to be assessed. Such samples may exhibit a higher probability of selecting the lowest (highest severity) level on the attributes, providing an instructional discussion on the floor and ceiling effects observed with the use of the DUI.

Suggestions for Future Research

Additional validation testing using the DUI in more diverse populations will enhance confidence in the applicability of the DUI to other research settings. Further research is also needed to evaluate the responsiveness of the DUI. Since the DUI was able to discriminate between groups based on disease severity, it can be expected to be clinically responsive, which will enable its use in the assessment of changes in the detection of the effects of disease progression and due to clinical interventions like the Staged Diabetes Management program (Mazze et al., 1994; Mazze, Bergenstal, & Ginsberg, 1995). Also, evaluation of the Minimally Important Difference on the DUI alongside indicators of disease severity will be needed to ascertain the smallest change in DUI utility score associated with a clinically meaningful change in health status. With diabetes being a relatively asymptomatic and chronic disease, these changes will need to be explored at longer time intervals.

Potential Application of the Diabetes Utility Index

The utility scores provided by the DUI MAUF can be useful in summarizing the health of diabetes populations. Hence the DUI will be an appropriate tool to obtain preference-weights to calculate Quality-Adjusted Life Years for use in Cost-Utility Analyses. The DUI also finds application in comparing outcomes between groups of patients with diabetes in clinical trials and clinical studies. Both overall outcomes and attribute-specific outcomes can be compared between groups using the general MAUF and single-attribute functions of the DUI respectively. At the same time, the DUI MAUF utilities should not be used to make clinical decisions for individual patients since preferences for health states are not homogenous. Direct elicitation of utilities, using the SG, is the most appropriate technique in this regard, to use with decision trees and other types of decision analysis.

The content of the DUI also makes it appropriate for use in diabetes intervention and educational programs in which patients' satisfaction with the management of the disease and its treatments and patients' initiative and control with dietary regulation are key goals. The diet-related attribute of the DUI describes the level of burden required for adaptation to and enjoyment of food and drink despite the limitations often imposed by diabetes. Diabetes Intervention Programs are typically designed to lead to a decreased sense of burden and an increased sense of enjoyment of diet that may result from more frequent appropriate food and drink choices. The fifth DUI attribute describes the satisfaction felt with individuals' management of diabetes. The more individuals feel confident to do what is needed to take care of their diabetes and that they are taking the necessary steps in that direction, the more satisfaction should be present. Through the inclusion of these attributes, the DUI not only reflects attitude but confidence or ability as well.

From the above discussion, it can be understood that in addition to being able to evaluate the impact of diabetes on aspects of functioning, the DUI will be able to measure the degree of burden or difficulty associated with prescribed behaviors as perceived by the patient. The adjustments needed to be made in the management of diabetes as discussed above are permanent and can be interpreted as constituting separate domains that independently influence diabetes patients' standing on the construct of QoL. The

DUI single-attribute scores, and hence overall DUI utility scores, can be expected to be positively associated with improved diabetes management. It would seem that the outcomes might be dependent upon the method, content and duration of instruction. Longitudinal research designs involving Diabetes Intervention Programs with known efficacy, like the Staged Diabetes Management program (Philis-Tsimikas & Walker, 2001; Zanetti et al., 2007) and the Chronic Disease Model that incorporates the American Association of Diabetes Educators' 7 Self-Care Behaviors, can also be used to evaluate the responsiveness of the DUI to these outcomes.

CONCLUSIONS

Patient assessed measures of health outcomes are increasingly being used alongside traditional biomedical measures for the evaluation of treatments and the management of diabetes. The Diabetes Utility Index provides this assessment by providing a preference-weighted score that can be used to calculate QALYs. The diabetes health state classification system of the DUI exhibited satisfactory fit to the Rasch uni-dimensional model. The preference scoring function was derived from a well-established theory, the Multi-attribute Utility Theory, and provided utility scores that compared adequately with cardinal Standard Gamble utilites. Results of the concurrent validation, with existing established preference-based measures of health, and construct validation, with indicators of diabetes severity, were both encouraging. The DUI will be appropriate for use in both clinical interventions as well as disease management programs for patients with diabetes.

BIBLIOGRAPHY

American Diabetes Association (1998). Economic consequences of diabetes mellitus in the U.S. in 1997. *Diabetes Care, 21,* 296-309.

Centers for Disease Control and Prevention (2003). *National diabetes fact sheet: general information and national estimates on diabetes in the United States, 2003*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (2006, March 22). BMI-Body Mass Index. Retrieved December 15, 2004, from http://www.cdc.gov/nccdphp/dnpa/bmi/index.htm

DAFNE Study Group (2002). Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: dose adjustment for normal eating (DAFNE) randomised controlled trial. *British Medical Journal, 325,* 746.

National Institute of Diabetes and Digestive and Kidney Diseases (2005). *National Diabetes Statistics fact sheet: general information and national estimates on diabetes in the United States, 2005*. Bethesda, MD: U.S. Department of Health and Human Services, National Institute of Health.

Anderson, R. M., Fitzgerald, J. T., Gruppen, L. D., Funnell, M. M., & Oh, M. S. (2003). The Diabetes Empowerment Scale-Short Form (DES-SF). *Diabetes Care, 26,* 1641-1642.

Anderson, R. M., Funnell, M. M., Fitzgerald, J. T., & Marrero, D. G. (2000). The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy 1. *Diabetes Care, 23,* 739-743.

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics, 11,* 447-456.

Bleichrodt, H. & Johannesson, M. (1997). An experimental test of a theoretical foundation for rating-scale valuations. *Medical Decision Making, 17,* 208-216.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boyle, M. H., Torrance, G. W., Sinclair, J. C., & Horwood, S. P. (1983). Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *New England Journal of Medicine, 308,* 1330-1337.

Bradley, C. (1994). The Well-Being Questionnaire. In C.Bradley (Ed.), *Handbook of Psychology and Diabetes: A Guide to Psychological Measurement in Diabetes Research and Practice.* (pp. 89-109). Switzerland: Harwood Academic Publishers.

Bradley, C. (2000). The 12-Item Well-Being Questionnaire: origins, current stage of development, and availability
38. *Diabetes Care, 23,* 875.

Bradley, C. (2001). Importance of differentiating health status from quality of life. *Lancet, 357,* 7-8.

Bradley, C. & Speight, J. (2002). Patient perceptions of diabetes and diabetes therapy: assessing quality of life. *Diabetes/Metabolism Research and Reviews, 18 Suppl 3,* S64-S69.

Bradley, C., Todd, C., Gorton, T., Symonds, E., Martin, A., & Plowright, R. (1999). The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. *Quality of Life Research, 8,* 79-91.

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment, 3,* i-164.

Brazier, J. & Fitzpatrick, R. (2002). Measures of health-related quality of life in an imperfect world: a comment on Dowie. *Health Economics, 11,* 17-19.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21,* 271-292.

Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology, 51,* 1115-1128.

Brazier, J. E. (2005). *Current state of the art in preference measures of health and avenues for further research* (Rep. No. 05/05 (Discussion Paper Series)). Sheffield, UK: University of Sheffield.

Brazier, J. E., Kolotkin, R. L., Crosby, R. D., & Williams, G. R. (2004a). Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value in Health, 7,* 490-498.

Brazier, J. E. & Roberts, J. (2004b). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42,* 851-859.

Brazier, J. E., Roberts, J., Platts, M., & Zoellner, Y. F. (2005). Estimating a preference-based index for a menopause specific health quality of life questionnaire. *Health and Quality of Life Outcomes 3,* 1. Retrieved December 20, 2005, from http://www.hqlo.com/content/3/1/13

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy, 37,* 53-72.

Bryant, F. B. & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L.G.Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate analysis* (Washington, D.C.: American Psychological Association.

Casey, R., Tarride, J. E., Keresteci, M. A., & Torrance, G. W. (2006). The Erectile Function Visual Analog Scale (EF-VAS): a disease-specific utility instrument for the assessment of erectile function. *Canadian Journal of Urology, 13,* 3016-3025.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases, 40,* 373-383.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2 ed.) Hillsdale, NJ: Erlbaum.

Conrad, K. J. & Smith, E. V., Jr. (2004). International conference on objective measurement: applications of Rasch analysis in health care. *Medical Care, 42,* I1-I6.

Coste, J., Guillemin, F., Pouchot, J., & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50,* 247-252.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, FL: Harcourt Brace Jovanovich College Publishers.

d'Hoore, W., Bouckaert, A., & Tilquin, C. (1996). Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of Clinical Epidemiology, 49,* 1429-1433.

DeVellis, R. F. (2003). Factor Analysis. In *Scale Development* (2 ed., pp. 102-137). Thousand Oaks, CA: Sage Publications.

Dickey, R. A. (2005). ICD-9-CM Diagnosis Codes for Diabetes Mellitus Revised. Retrieved May 15, 2006, from http://www.endo-society.org/news/endocrine_news/2005/codingnewsApr2005.cfm

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35,* 1095-1108.

Dolan, P. (1999). Whose preferences count? *Medical Decision Making, 19,* 482-486.

Dolan, P. (2000). The measurement of health-related quality of life. In A.J.Culyer & J. Newhouse (Eds.), *Handbook of health economics. Vol 1B.* (pp. 1723-1760). Amsterdam, Netherlands: Elsevier.

Dolan, P. (2002). Modelling the relationship between the description and valuation of health states. In C.J.L.Murray, J. A. Salomon, C. D. Mathers, & A. D. Lopez (Eds.), *Summary measures of population health: Concepts, Ethics, Measurement and Applications* (pp. 510-513). Geneva, Switzerland: World Health Organization.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1995). *A social tariff for EuroQoL: Results from an UK general population survey. Discussion paper* (Rep. No. 38). York: Center for health economics, University of York.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). The time trade-off method: results from a general population study. *Health Economics, 5,* 141-154.

Doward, L. C., Meads, D. M., & Thorsen, H. (2004). Requirements for quality of life instruments in clinical research. *Value in Health, 7 Suppl 1,* S13-S16.

Drummond, M. F., O'Brien, B. J., Stoddart, G. L., & Torrance, G. W. (1997a). Cost-utility analysis. In *Methods for the economic evaluation of healthcare programs* (2 ed., pp. 139-204). New York, NY: Oxford University Press.

Drummond, M. F., O'Brien, B. J., Stoddart, G. L., & Torrance, G. W. (1997b). *Methods for the economic evaluation of healthcare programs*. (2 ed.) New York, NY: Oxford University Press.

Eiser, C. & Tooke, J. E. (1993). Quality-of-life evaluation in diabetes 3. *Pharmacoeconomics., 4,* 85-91.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28,* 1-11.

Faul, F. & Erdfelder, E. (1992). *GPOWER: A priori, post hoc, and compromise power analysis for MS DOS (computer program)*. Bonn, Germany: Bonn University, Dept. of Psychology.

Fayers, P. M. & Machin, D. (2000). *Quality of life: assessment, analysis and interpretation*. New York, NY: John Wiley.

Feeny, D. (2002a). Health-status classification systems for summary measures of population health. In C.J.L.Murray, J. A. Salomon, C. D. Mathers, & A. D. Lopez (Eds.), *Summary measures of population health: Concepts, Ethics, Measurement and Applications* (pp. 329-341). Geneva, Switzerland: World Health Organization.

Feeny, D. (2002b). The utility approach to assessing population health. In C.J.L.Murray, J. A. Salomon, C. D. Mathers, & A. D. Lopez (Eds.), *Summary measures of population health: Concepts, Ethics, Measurement and Applications* (pp. 515-528). Geneva, Switzerland: World Health Organization.

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics., 7,* 490-502.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S. et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care, 40,* 113-128.

Fisk, J. D. & Doble, S. E. (2002). Construction and validation of a fatigue impact scale for daily administration (D-FIS). *Quality of Life Research, 11,* 263-272.

Franks, P., Lubetkin, E. I., Gold, M. R., Tancredi, D. J., & Jia, H. (2004). Mapping the SF-12 to the EuroQol EQ-5D Index in a national US sample. *Medical Decision Making, 24,* 247-254.

Froberg, D. G. & Kane, R. L. (1989). Methodology for measuring health-state preferences--II: Scaling methods. *Journal of Clinical Epidemiology, 42,* 459-471.

Fryback, D. G., Lawrence, W. F., Martin, P. A., Klein, R., & Klein, B. E. (1997). Predicting Quality of Well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. *Medical Decision Making, 17,* 1-9.

Furlong, W., Feeny, D. H., Torrance, G. W., Barr, R., & Horsman, J. (1990). *Guide to design and development of health state utility instrumentation*. Hamilton, Ontario: Center for Health Economics and Policy Analysis, McMaster University.

Furlong, W., Feeny, D. H., Torrance, G. W., Goldsmith, C. H., DePauw, S., Zhu, Z. et al. (1998). *Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI 3): A technical report* (Rep. No. 98-11). Hamilton, Ontario: Center for Health Economics and Policy Analysis, McMaster University.

Gabriel, S. E., Kneeland, T. S., Melton, L. J., III, Moncur, M. M., Ettinger, B., & Tosteson, A. N. (1999). Health-related quality of life in economic evaluations for osteoporosis: whose values should we use? *Medical Decision Making, 19,* 141-148.

Garratt, A. M., Schmidt, L., & Fitzpatrick, R. (2002). Patient-assessed health outcome measures for diabetes: a structured review. *Diabetic Medicine, 19,* 1-11.

Gold, M., Franks, P., & Erickson, P. (1996). Assessing the health of the nation. The predictive validity of a preference-based measure and self-rated health. *Medical Care, 34,* 163-177.

Gold, M. R., Patrick, D. L., Torrance, G. W., Fryback, D. G., Hadorn, D. C., Kamlet, M. S. et al. (1996a). Identifying and valuing outcomes. In M.R.Gold, J. E. Siegel, L. B. Russell, & M. C. Weinstein (Eds.), *Cost-effectiveness in Health and Medicine* (pp. 82-134). New York, NY: Oxford University Press.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996b). *Cost-effectiveness in Health and Medicine*. New York, NY: Oxford University Press.

Grootenhuis, P. A., Snoek, F. J., Heine, R. J., & Bouter, L. M. (1994). Development of a type 2 diabetes symptom checklist: a measure of symptom severity 42. *Diabet.Med, 11,* 253-261.

Hareendran, A. (2005). Patient-reported outcomes: When do they add value? *Patient Reported Outcomes Newsletter*.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32,* 1008-1015.

Hogan, P., Dall, T., & Nikolov, P. (2003). Economic costs of diabetes in the US in 2002. *Diabetes Care, 26,* 917-932.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S., & Layte, R. (1997). Evaluation of index and profile measures of health status in a randomized controlled trial. Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures. *Medical Care, 35,* 1109-1118.

Jones, J. & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal, 311,* 376-380.

Juniper, E. F., Guyatt, G. H., & Jaeschke, R. (1996). How to develop and validate a new health-related quality of life instrument. In B.Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2 ed., pp. 49-56). Philadelphia, PA: Lippincott-Raven.

Juniper, E. F., Guyatt, G. H., Streiner, D. L., & King, D. R. (1997). Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology, 50,* 233-238.

Kaplan, R. M. & Anderson, J. P. (1988). A general health policy model: update and applications. *Health Services Research, 23,* 203-235.

Kaplan, R. M. & Anderson, J. P. (1996). The general health policy model: and integrated approach. In B.Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2 ed., pp. 302-322). Philadelphia, PA: Lippincott-Raven.

Keeny, R. L. & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs.* (1 ed.) New York, NY: John Wiley & Sons.

Keeny, R. L. & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs.* (2 ed.) New York, NY: Cambridge University Press.

Kind, P. (1996). The EuroQoL instrument: an index of health-related quality of life. In B.Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2 ed., pp. 191-201). Philadelphia, PA: Lippincott-Raven.

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage.

Kochanek, K. D. & Smith, B. L. (2004). *Deaths: Preliminary data for 2002. National vital statistics reports, 52(13)*. Hyattsville, MD: National Center for Health Statistics.

Krabbe, P. F., Essink-Bot, M. L., & Bonsel, G. J. (1997). The comparability and reliability of five health-state valuation methods. *Social Science and Medicine, 45,* 1641-1652.

Lau, C. Y., Qureshi, A. K., & Scott, S. G. (2004). Association between glycaemic control and quality of life in diabetes mellitus
1. *J Postgrad.Med, 50,* 189-193.

Lenert, L. & Kaplan, R. M. (2000). Validity and interpretation of preference-based measures of health-related quality of life. *Medical Care, 38,* II138-II150.

Lenert, L. A. (2000). The reliability and internal consistency of an Internet-capable computer program for measuring utilities. *Quality of Life Research, 9,* 811-817.

Linacre, J. M. (1999). Investigating rating scale category utility. *J Outcome.Meas, 3,* 103-122.

Linacre, M. (2007). *Winsteps (Version 3.63.2) [Computer Software]*. Chicago: Winsteps.com.

Llewellyn-Thomas, H., Sutherland, H. J., Tibshirani, R., Ciampi, A., Till, J. E., & Boyd, N. F. (1984). Describing health states. Methodologic issues in obtaining values for health states. *Medical Care, 22,* 543-552.

Lloyd, C. E., Matthews, K. A., Wing, R. R., & Orchard, T. J. (1992). Psychosocial factors and complications of IDDM. The Pittsburgh Epidemiology of Diabetes Complications Study. VIII. *Diabetes Care, 15,* 166-172.

Lloyd, C. E., Wing, R. R., Orchard, T. J., & Becker, D. J. (1993). Psychosocial correlates of glycemic control: the Pittsburgh Epidemiology of Diabetes Complications (EDC) Study. *Diabetes Res.Clin Pract., 21,* 187-195.

Louviere, J. J. (1988). *Analyzing decision making: Metric conjoint analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences 07-067.* Newbury Park, CA: Sage Publications.

Mazze, R. S., Bergenstal, R., & Ginsberg, B. (1995). Intensified diabetes management: lessons from the diabetes control and complications trial 12. *Int J Clin Pharmacol.Ther., 33,* 43-51.

Mazze, R. S., Etzwiler, D. D., Strock, E., Peterson, K., McClave, C. R., Meszaros, J. F. et al. (1994). Staged diabetes management. Toward an integrated model of diabetes care. *Diabetes Care, 17 Suppl 1,* 56-66.

McKenna, H. P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of Advanced Nursing, 19,* 1221-1225.

Mrus, J. M., Yi, M. S., Freedberg, K. A., Wu, A. W., Zackin, R., Gorski, H. et al. (2003). Utilities derived from visual analog scale scores in patients with HIV/AIDS. *Medical Decision Making, 23,* 414-421.

Nijsten, T., Unaeze, J., & Stern, R. S. (2006). Refinement and reduction of the Impact of Psoriasis Questionnaire: classical test theory vs. Rasch analysis. *British Journal of Dermatology, 154,* 692-700.

Nijsten, T. E., Sampogna, F., Chren, M. M., & Abeni, D. D. (2006). Testing and reducing skindex-29 using Rasch analysis: Skindex-17. *Journal of Investigative Dermatology, 126,* 1244-1250.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. (3 ed.) New York, NY: Mc-Graw Hill.

O'Boyle, C. A., McGee, H., Hickey, A., O'Malley, K., & Joyce, C. R. (1992). Individual quality of life in patients undergoing hip replacement. *Lancet, 339,* 1088-1091.

O'Connor, R. (2004). *Measuring quality of life in health*. Churchill Livingstone.

O'Leary, J. F., Fairclough, D. L., Jankowski, M. K., & Weeks, J. C. (1995). Comparison of time-tradeoff utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. *Medical Decision Making, 15,* 132-137.

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Philis-Tsimikas, A. & Walker, C. (2001). Improved care for diabetes in underserved populations. *J Ambul.Care Manage., 24,* 39-43.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28,* 206-224.

Poissant, L., Mayo, N. E., Wood-Dauphinee, S., & Clarke, A. E. (2003). The development and preliminary validation of a Preference-Based Stroke Index (PBSI).

Health and Quality of Life Outcomes 1, 1. Retrieved November 20, 2005, from http://www.hqlo.com/content/1/1/43

Polonsky, W. H. (2000). Understanding and assessing diabetes-specific quality of life. *Diabetes Spectrum, 13,* 36-41.

Postulart, D. & Adang, E. M. (2000). Response shift and adaptation in chronically ill patients. *Medical Decision Making, 20,* 186-193.

Prieto, L., Alonso, J., & Lamarca, R. (2003, July 28). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes 1,* 1. Retrieved December 13, 2005, from http://www.hqlo.com/content/1/1/27

Prieto, L., Alonso, J., Lamarca, R., & Wright, B. D. (1998). Rasch measurement for reducing the items of the Nottingham Health Profile. *Journal of Outcome Measurement, 2,* 285-301.

Prieto, L., Thorsen, H., & Juul, K. (2005). Development and validation of a quality of life questionnaire for patients with colostomy or ileostomy. *Health and Quality of Life Outcomes 3,* 1. Retrieved October 20, 2006, from http://www.hqlo.com/content/3/1/62

Raat, H., Bonsel, G. J., Hoogeveen, W. C., & Essink-Bot, M. L. (2004). Feasibility and reliability of a mailed questionnaire to obtain visual analogue scale valuations for health states defined by the Health Utilities Index Mark 3. *Medical Care, 42,* 13-18.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.

Revicki, D. A. (1996). Relationship of pharmacoeconomics and health-related quality of life. In B.Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2 ed., pp. 1077-1084). Philadelphia, PA: Lippincott-Raven.

Revicki, D. A., Leidy, N. K., Brennan-Diemer, F., Sorensen, S., & Togias, A. (1998a). Integrating patient preferences into health outcomes assessment: the multiattribute Asthma Symptom Utility Index. *Chest, 114,* 998-1007.

Revicki, D. A., Leidy, N. K., Brennan-Diemer, F., Thompson, C., & Togias, A. (1998b). Development and preliminary validation of the multiattribute Rhinitis Symptom Utility Index. *Quality of Life Research, 7,* 693-702.

Riazi, A., Bradley, C., Barendse, S., & Ishii, H. (2006). Development of the Well-being questionnaire short-form in Japanese: the W-BQ12. *Health and Quality of Life Outcomes, 4,* 40.

Rizzo, J. A., Pashko, S., Friedkin, R., Mullahy, J., & Sindelar, J. L. (1998). Linking the health utilities index to National Medical Expenditure Survey data. *Pharmacoeconomics., 13,* 531-541.

Rizzo, J. A. & Sindelar, J. L. (1999). Linking health-related quality-of-life indicators to large national data sets. *Pharmacoeconomics., 16,* 473-482.

Robinson, A., Loomes, G., & Jones-Lee, M. (2001). Visual analog scales, standard gambles, and relative risk aversion. *Medical Decision Making, 21,* 17-27.

Rosser, R. & Kind, P. (1978). A scale of valuations of states of illness: is there a social consensus? *International Journal of Epidemiology, 7,* 347-358.

Rubin, R. J., Altman, W. M., & Mendelson, D. N. (1994). Health care expenditures for people with diabetes mellitus, 1992. *Journal of Clinical Endocrinology and Metabolism, 78,* 809A-809F.

Rubin, R. R. (2000). Diabetes and quality of life. *Diabetes Spectrum, 13,* 21-23.

Schackman, B. R., Goldie, S. J., Freedberg, K. A., Losina, E., Brazier, J., & Weinstein, M. C. (2002). Comparison of health state utilities using community and patient preference weights derived from a survey of patients with HIV/AIDS. *Medical Decision Making, 22,* 27-38.

Schwartz, A. (1998). Rating scales in context. *Medical Decision Making, 18,* 236.

Sengupta, N., Nichol, M. B., Wu, J., & Globe, D. (2004). Mapping the SF-12 to the HUI3 and VAS in a managed care population
1. *Med Care, 42,* 927-937.

Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical Care, 43,* 203-220.

Smith, R. M., Schumacker, R. E., & Bush, M. J. (1998). Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement, 2,* 66-78.

SPSS (2007). *SPSS for Windows (Rel. 15.0.1.) [Computer Software]*. Chicago: SPSS Inc.

Stiggelbout, A. M., Eijkemans, M. J., Kiebert, G. M., Kievit, J., Leer, J. W., & De Haes, H. J. (1996). The 'utility' of the visual analog scale in medical decision making and technology assessment. Is it an alternative to the time trade-off? *International Journal of Technology Assessment in Health Care, 12,* 291-298.

Stiggelbout, A. M., Kiebert, G. M., Kievit, J., Leer, J. W., Habbema, J. D., & De Haes, J. C. (1995). The "utility" of the Time Trade-Off method in cancer patients: feasibility and proportional Trade-Off. *Journal of Clinical Epidemiology, 48,* 1207-1214.

Streiner, D. L. & Norman, G. R. (1989). *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press.

Sullivan, P. W. & Ghushchyan, V. (2006). Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Medical Decision Making, 26,* 401-409.

Sundaram, M. (2005). *Quality of Life and Clinical Outcomes in Type 2 Diabetes Patients at the Primary Care Clinics of The West Virginia University Hospital.* Master's Thesis, West Virginia University, 2005.

Sundaram, M., Kavookjian, J., Patrick, J. H., Miller, L. A., Madhavan, S. S., & Scott, V. G. (2006). Quality of life, health status and clinical outcomes in Type 2 diabetes patients. *Quality of Life Research*.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics*. (4 ed.) New York, NY: Harper Collins.

Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health, 7 Suppl 1,* S22-S26.

Torrance, G. W. (1976a). Social preference for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Science, 10,* 129-136.

Torrance, G. W. (1976b). Toward a utility theory foundation for health status index models. *Health Services Research, 11,* 349-369.

Torrance, G. W., Boyle, M. H., & Horwood, S. P. (1982). Application of multi-attribute utility theory to measure social preferences for health states. *Operations Research, 30,* 1043-1069.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical Decision Making, 21,* 329-334.

Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y., & Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Medical Care, 34,* 702-722.

Torrance, G. W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics., 7,* 503-520.

Torrance, G. W., Keresteci, M. A., Casey, R. W., Rosner, A. J., Ryan, N., & Breton, M. C. (2004). Development and initial validation of a new preference-based disease-specific health-related quality of life instrument for erectile function. *Quality of Life Research, 13,* 349-359.

Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research, 7,* 118-133.

Tsuchiya, A., Brazier, J. E., McColl, E., & Parkin, D. (2002). *Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices.* (Rep. No. 02/01 (Discussion Paper Series)). Sheffield, UK: University of Sheffield.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

van Osch, S. M., Wakker, P. P., van den Hout, W. B., & Stiggelbout, A. M. (2004). Correcting biases in standard gamble and time tradeoff utilities. *Medical Decision Making, 24,* 511-517.

von Neumann, J., Morgenstern, O., & Rubinstein, A. (2004). *Theory of games and economic behavior*. (Commerative ed.) Princeton, NJ: Princeton University Press.

Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity
6. *Med Care, 34,* 220-233.

Ware, J., Jr., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). *How to Score Version 2 of the SF-12 Health Survey (With a Supplement Documenting Version 1).* Lincoln, RI: QualityMetric Inc.

Wildes, K. R., Greisinger, A., & O'Malley, K. J. (2003). Critical review of quality of life measures for patients with diabetes: Report of the measurement excellence initiative. Retrieved January 17, 2004, from http://www.measurementexperts.org/criticalReview.htm.

Wright, B. D. (1999). Common sense for measurement. *Rasch Measurement Transactions, 13,* 704-705.

Wright, B. D. & Linacre, M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8,* 370.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Young, T. A., Yang, Y., Brazier, J. E., & Tsuchiya, A. (2005). *Using Rasch analysis to aid the construction of preference-based measures from existing quality of life instruments*. Poster presented at the 12th Annual Conference of the International Society for Quality of Life Research, San Francisco, CA

Zanetti, M. L., Otero, L. M., Biaggi, M. V., dos Santos, M. A., Peres, D. S., & Guimaraes, F. P. (2007). Satisfaction of diabetes patients under follow-up in a diabetes education program. *Rev.Lat.Am Enfermagem., 15,* 583-589.

APPENDICES

Appendix 1: Expectations from Expert Panel

Appendix 2: Expert Panel Input

Appendix 3: Suggested Version One

Appendix 4: Version One

Appendix 5: Suggested Version Two

Appendix 6: Version Two

Appendix 7: Final Version

Appendix 1: Expectations from Expert Panel

Dear member of my expert panel,

Thank you for agreeing to be a part of my PhD dissertation project. From my discussions with you in the past, you may remember that the purpose of my project is to develop a questionnaire that will ultimately enable the calculation of a score that describes the 'utility' that a patient with diabetes places on his or her present heath condition.

In case you may want to refresh your memory on the concept of 'utility', I have provided you with links to two websites:
1. http://www.bjmath.com/faq/utility/faqmanager.cgi?file=utility&toc=faq
This link offers a general overview of the theory of utility, by using a simple economic example.
2. http://symptomresearch.nih.gov/chapter_24/sec4/cmgs4pg1.htm
This link offers a healthcare perspective to measuring utility, or health state utility.

I'm requesting your input in two main areas:
1. Selecting attributes that influence health state utility in people with diabetes, and
2. Describing the selected attributes and designating attribute levels.

1. Selecting attributes that influence health state utility in people with diabetes:

One of the major selection criteria will be the importance that is accorded to attributes in terms of how they affect utility. This exercise ideally involves an extensive amount of primary research on these attributes in subjects with diabetes. Researchers have suggested that this step can be supplemented by the use of previously designed questionnaires that measure a similar concept and that have been validated in similar populations.

In my Master's thesis, I collected data on the quality of life of people with Type 2 diabetes using an instrument called the 'Audit of Diabetes Dependent Quality of Life', or ADDQoL. The ADDQoL asks respondents to rate how the following aspects of life would be influenced if that individual did not have diabetes:

1. Working life
2. Family life
3. Social life
4. Sex life
5. Physical Appearance
6. Physical activities
7. Holidays/ leisure
8. Travel
9. Confidence in ability
10. Motivation
11. Society reaction
12. Worries about the future
13. Finances
14. Dependence
15. Living Conditions
16. Freedom to eat
17. Enjoyment of food
18. Freedom to drink

A copy of the ADDQoL is attached towards the end of this document (Appendix A). I will be using statistical analyses (using ADDQoL scores from a dataset I compiled previously in my Master's research) to guide in the selection of 5 broad attributes from the 18 issues addressed in the ADDQoL. While this method does use indirect patient input, I've been advised by my dissertation committee to take a more balanced approach by obtaining the input of an expert panel as well.

1. Please advise if these attributes in all, in your opinion, provide adequate coverage of the MAIN areas that would influence the utility that a person with diabetes would place on his or her current condition. Please keep in mind that there may be several issues to consider, but the goal is to conceive broad areas or attributes (4-5 in number) that may be representative of those issues.

2. Please do not feel limited by the items in the ADDQoL on which the statistical analyses are based. Based on your experience in this clinical field, you may be aware of important areas not represented in the instrument that you think patients may consider very important, or that, according to you, is important in the assessment of utility.
   For example, the instrument does not provide coverage on symptoms of the condition (perhaps because primary research preceding the development of the instrument noted that there was lack of overall significantly distressing symptoms in diabetes that impaired patients' QoL; you may or may not agree to this). If you have in mind attributes that you think are crucial to patients' utility but are unrepresented in the list of items from the ADDQoL, please feel free to indicate those.

The underlying approach is to develop a brief questionnaire that does not severely compromise on the coverage of issues necessary to capture patient's health state utility.

Let me provide you with some examples of attributes included in instruments that assess health state utility. You may notice that these instruments differ in the nature of attributes included. While the Health Utilities Index Mark 3 (HUI-3) and the Short Form 6-D (SF-6D) are instruments that offer a generic perspective to the measurement of health state utility, the menopause health state classification system focuses on aspects unique to the condition.

The attributes included in the **HUI-3** are:
1. Vision
2. Hearing
3. Speech
4. Ambulation
5. Dexterity
6. Emotion
7. Cognition
8. Pain

You may notice that this instrument employs an 'under the skin' approach in choosing attributes.

The attributes included in the **SF-6D** are:

1. Physical functioning
2. Role Limitations
3. Social Functioning
4. Pain
5. Mental Health, and
6. Vitality

You may notice that this instrument incorporates aspects of life and existence beyond those related to the state of the body.

The attributes included in the **menopause health state classification system** are:

1. hot flushes
2. aching joints or muscles
3. anxious or frightened feelings
4. breast tenderness
5. bleeding
6. undesirable cosmetic signs (facial or body hair growth, greasy skin or acne)
7. vaginal dryness

You may notice that this classification system focuses on symptoms.

For our own purpose, it is useful to consider the broad areas listed above. I gave you these examples to provide you with a glimpse into the state of the art in the development of an instrument assessing health state utility.

2. Describing the selected attributes and designating attribute levels

Attribute levels need to be defined to cover the full range of possible abilities/disabilities and to be clearly distinguishable from one another. I will attempt to provide initial thought on the desirable range on the basis of statistical input, which you may or may not agree to.

3. Please suggest a one line description of the attribute; this will be a way to guide the respondent to provide an answer about the attribute with better clarity. This may be contrasted with the use of words or phrases to describe attributes in the examples provided above.

4. Please suggest descriptions for the attribute levels. We may consider avoiding designating attribute levels such as ‘highly affected, moderately affected’, and so on. While these types of scales have been used widely, the interpretation of the relative differences between adjoining scale levels is left to the respondent. We should address that by describing attribute levels briefly, and if at all possible, keeping in mind scenarios that are more reflective of what a person with diabetes may undergo with regards to that particular attribute.

Again, I’ve provided you with examples of the way attribute levels have been described for the instruments discussed before, in Appendices B, C, and D.

Where do we go from here?

I will be collating input from a panel of experts, of which you are a part, using the Delphi technique. I will soon be emailing you a document with a summary of the results of the statistical analyses. I will make an attempt to suggest 4-5 attributes from available ADDQoL data as well as their severity levels. I will request you to make your own individual assessments, using the suggestions in this document as a guideline if you agree to them in principle, and return them in 10-15 days. This will be Delphi Component One.

I will then compile individual contributions and email a document which, this time, contains a draft of the questionnaire that may be influenced by the opinions of several panel members. The expert panel will be approached again - this time to fine-tune the questionnaire with regards to the wording, and if necessary, to alter the item content, and return it to me by email in 7-10 days. This will be Delphi Component Two.

At this point I should have enough input from the panel to be able to finalize the instrument. I will have a better idea as to the scheduling of these components once the statistical analyses are ready. Component Two would depend on the timeliness of completion of Component One.

Thank you for going through this document. I look forward to working with you.

Appendix A: The ADDQoL Questionnaire

ADDQoL

This questionnaire asks about your quality of life and the effects of your diabetes on your quality of life. Your quality of life is how good or bad you feel your life to be.

Please shade the circle which best indicates your response on each scale.

There are no right or wrong answers; we just want to know how you feel about your life now.

I) In general, my present quality of life is:

O	O	O	O	O	O	O
excellent	very good	good	neither good nor bad	bad	very bad	extremely bad

For the next statement please consider the effects of your diabetes, its management and any complications you may have.

II) If I did not have diabetes, my quality of life would be:

O	O	O	O	O	O	O
very much better	much better	a little better	the same	a little worse	much worse	very much worse

Please respond to the 18 more specific statements on the pages that follow.

For each statement, please consider the effects of your diabetes, its management and any complications you may have on the aspect of life described by the statement.

In each of the following boxes:

a) shade a circle to show how diabetes affects this aspect of your life;

b) shade a circle to show how important this aspect of your life is to your quality of life.

Some statements have a "not applicable" option. Please shade this "not applicable" circle if that aspect of life does not apply to you.

1a)	**If I did not have diabetes, my working life and work-related opportunities would be:**							
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse	O not applicable
1b)	**This aspect of my life is:**							
		O very important	O important	O somewhat important	O not at all important			

2a)	**If I did not have diabetes, my family life would be:**							
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse	O not applicable
2b)	**This aspect of my life is:**							
		O very important	O important	O somewhat important	O not at all important			

3a)	**If I did not have diabetes, my friendships and social life would be:**						
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse
3b)	**This aspect of my life is:**						
		O very important	O important	O somewhat important	O not at all important		

4a)	**If I did not have diabetes, my sex life would be:**							
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse	O not applicable
4b)	**This aspect of my life is:**							
		O very important	O important	O somewhat important	O not at all important			

5a)	If I did not have diabetes, my physical appearance would be:						
	○ very much better	○ much better	○ a little better	○ the same	○ a little worse	○ much worse	○ very much worse
5b)	This aspect of my life is:						
		○ very important	○ important	○ somewhat important	○ not at all important		

6a)	If I did not have diabetes, the things I could do physically would be:						
	○ very much increased	○ much increased	○ a little increased	○ the same	○ a little decreased	○ much decreased	○ very much decreased
6b)	This aspect of my life is:						
		○ very important	○ important	○ somewhat important	○ not at all important		

7a)	If I did not have diabetes, my vacations or leisure activities would be:						
	○ very much better	○ much better	○ a little better	○ the same	○ a little worse	○ much worse	○ very much worse
7b)	This aspect of my life is:						
		○ very important	○ important	○ somewhat important	○ not at all important		

8a)	If I did not have diabetes, ease of travelling (local or long distance) would be:						
	○ very much better	○ much better	○ a little better	○ the same	○ a little worse	○ much worse	○ very much worse
8b)	This aspect of my life is:						
		○ very important	○ important	○ somewhat important	○ not at all important		

9a)	**If I did not have diabetes, my confidence in my ability to do things would be:**					
O	O	O	O	O	O	O
very much increased	much increased	a little increased	the same	a little decreased	much decreased	very much decreased

9b)	**This aspect of my life is:**		
O	O	O	O
very important	important	somewhat important	not at all important

10a)	**If I did not have diabetes, my motivation to achieve things would be:**					
O	O	O	O	O	O	O
very much increased	much increased	a little increased	the same	a little decreased	much decreased	very much decreased

10b)	**This aspect of my life is:**		
O	O	O	O
very important	important	somewhat important	not at all important

11a)	**If I did not have diabetes, the way society at large reacts to me would be:**					
O	O	O	O	O	O	O
very much better	much better	a little better	the same	a little worse	much worse	very much worse

11b)	**This aspect of my life is:**		
O	O	O	O
very important	important	somewhat important	not at all important

12a)	**If I did not have diabetes, my worries about the future would be:**					
O	O	O	O	O	O	O
very much decreased	much decreased	a little decreased	the same	a little increased	much increased	very much increased

12b)	**This aspect of my life is:**		
O	O	O	O
very important	important	somewhat important	not at all important

13a)	**If I did not have diabetes, my finances would be:**						
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse
13b)	**This aspect of my life is:**						
	O very important	O important	O somewhat important	O not at all important			

14a)	**If I did not have diabetes, my need to depend on others for things I would like to do for myself would be:**						
	O very much decreased	O much decreased	O a little decreased	O the same	O a little increased	O much increased	O very much increased
14b)	**This aspect of my life is:**						
	O very important	O important	O somewhat important	O not at all important			

15a)	**If I did not have diabetes, my living conditions would be:**						
	O very much better	O much better	O a little better	O the same	O a little worse	O much worse	O very much worse
15b)	**This aspect of my life is:**						
	O very important	O important	O somewhat important	O not at all important			

16a)	**If I did not have diabetes, my freedom to eat as I wish would be:**						
	O very much increased	O much increased	O a little increased	O the same	O a little decreased	O much decreased	O very much decreased
16b)	**This aspect of my life is:**						
	O very important	O important	O somewhat important	O not at all important			

17a)	If I did not have diabetes, my enjoyment of food would be:					
O very much increased	O much increased	O a little increased	O the same	O a little decreased	O much decreased	O very much decreased

17b)	This aspect of my life is:		
O very important	O important	O somewhat important	O not at all important

18a)	If I did not have diabetes, my freedom to drink as I wish (e.g. sweetened hot and cold drinks, fruit juice, alcohol) would be:					
O very much increased	O much increased	O a little increased	O the same	O a little decreased	O much decreased	O very much decreased

18b)	This aspect of my life is:		
O very important	O important	O somewhat important	O not at all important

If there are any other ways in which diabetes, its management and any complications affect your quality of life, please say what they are below:

Appendix B: The Menopause health state classification

Please consult the following paper for more details on this instrument and for potential use:

Brazier JE, Roberts J, Platts M, Zoellner YF. Estimating a preference-based index for a menopause specific health quality of life questionnaire. Health Qual Life Outcomes. 2005 Mar 15;3:13.

1. Hot flushes
    1) You have no hot flushes
    2) You get 1–3 hot flushes per day
    3) You get 4 or more hot flushes per day

2. Aching joints or muscles
    1) You have no aching joints or muscles at all.
    2) You have 1–3 episodes of aching joints or muscles per week.
    3) You have 4 or more episodes of aching joints or muscles per week.
    4) You have mild to moderate constant pain in your joints or muscles.
    5) You have severe constant pain in your joints or muscles.

3. Anxious or frightened feelings
    1) You do not have anxious or frightened feelings.
    2) You have anxious or frightened feelings 1–3 times per week.
    3) You have anxious or frightened feelings 4 or more times per week.

4. Breast tenderness
    1) You have no breast tenderness.
    2) You have mild to moderate breast tenderness.
    3) You have severe breast tenderness

5. Bleeding
    1) You have no bleeding
    2) You have mild regular (monthly) bleeding
    3) You have mild irregular bleeding
    4) You have intense regular (monthly) bleeding
    5) You have intense irregular bleeding

6. Undesirable cosmetic signs (facial or body hair growth, greasy skin or acne)
    1) You have no undesirable cosmetic signs.
    2) You have mild to moderate undesirable cosmetic signs
    3) You have severe undesirable cosmetic signs.

7. Vaginal dryness
    1) You have no vaginal dryness.
    2) You have mild to moderate vaginal dryness.
    3) You have severe vaginal dryness.

Appendix C: The HUI-3 Classification System

1. Vision

1) Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.
2) Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, but with glasses.
3) Able to read ordinary newsprint with or without glasses but unable to recognize a friend on the other side of the street, even with glasses.
4) Able to recognize a friend on the other side of the street with or without glasses but unable to read ordinary newsprint, even with glasses.
5) Unable to read ordinary newsprint and unable to recognize a friend on the other side of the street, even with glasses.
6) Unable to see at all.

2. Hearing

1) Able to hear what is said in a group conversation with at least three other people, without a hearing aid.
2) Able to hear what is said in a conversation with one other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least three other people.
3) Able to hear what is said in a conversation with one other person in a quiet room with a hearing aid, and able to hear what is said in a group conversation with at least three other people, with a hearing aid.
4) Able to hear what is said in a conversation with one other person in a quiet room, without a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid.
5) Able to hear what is said in a conversation with one other person in a quiet room with a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid.
6) Unable to hear at all.

3. Speech

1) Able to be understood completely when speaking with strangers or friends.
2) Able to be understood partially when speaking with strangers but able to be understood completely when speaking with people who know me well.
3) Able to be understood partially when speaking with strangers or people who know me well.
4) Unable to be understood when speaking with strangers but able to be understood partially by people who know me well.
5) Unable to be understood when speaking to other people (or unable to speak at all).

The Health Utilities Index (Mark 3) is copyrighted. Please contact Bill Furlong for enquiries.
E-mail: furlongb@mcmaster.ca

4. Ambulation

1) Able to walk around the neighbourhood without difficulty, and without walking equipment.
2) Able to walk around the neighbourhood with difficulty; but does not require walking equipment or the help of another person.
3) Able to walk around the neighbourhood with walking equipment, but without the help of another person.
4) Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood.
5) Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood.
6) Cannot walk at all.

5. Dexterity

1) Full use of two hands and ten fingers.
2) Limitations in the use of hands or fingers, but does not require special tools or help of another person.
3) Limitations in the use of hands or fingers, is independent with use of special tools (does not require the help of another person).
4) Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with use of special tools).
5) Limitations in use of hands or fingers, requires the help of another person for most tasks (not independent even with use of special tools).
6) Limitations in use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools).

6. Emotion

1) Happy and interested in life.
2) Somewhat happy.
3) Somewhat unhappy.
4) Very unhappy.
5) So unhappy that life is not worthwhile.

7. Cognition

1) Able to remember most things, think clearly and solve day to day problems.
2) Able to remember most things, but have a little difficulty when trying to think and solve day to day problems.
3) Somewhat forgetful, but able to think clearly and solve day to day problems.
4) Somewhat forgetful, and have a little difficulty when trying to think or solve day to day problems.
5) Very forgetful, and have great difficulty when trying to think or solve day to day problems.
6) Unable to remember anything at all, and unable to think or solve day to day problems.

8. Pain

1) Free of pain and discomfort.
2) Mild to moderate pain that prevents no activities.
3) Moderate pain that prevents a few activities.
4) Moderate to severe pain that prevents some activities.
5) Severe pain that prevents most activities.

William (Bill) Furlong, MSc
Department of Clinical Epidemiology & Biostatistics
Faculty of Health Sciences, McMaster University
Health Utilities Inc.
88 Sydenham Street
Dundas, ON, Canada L9H 2V3
Telephone: (905) 525-9140 extension 22389
FAX: (905) 627-7914
E-mail: furlongb@mcmaster.ca

Appendix D: The SF-6D Classification System

Table 1
The short form-6D[a]

Level	Physical functioning	Role limitations	Social functioning	Pain	Mental health	Vitality
1	Your health does not limit you in *vigorous activities*	You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems	Your health limits your social activities *none of the time*	You have *no* pain	You feel tense or downhearted and low *none of the time*	You have a lot of energy *all of the time*
2	Your health limits you a little in *vigorous activities*	You are limited in the kind of work or other activities as a result of your physical health	Your health limits your social activities *a little of the time*	You have pain but it does not interfere with your normal work (both outside the home and housework)	You feel tense or downhearted and low *a little of the time*	You have a lot of energy *most of the time*
3	Your health limits you a little in *moderate activities*	You accomplish less than you would like as a result of emotional problems	Your health limits your social activities *some of the time*	You have pain that interferes with your normal work (both outside the home and housework) *a little bit*	You feel tense or downhearted and low *some of the time*	You have a lot of energy *some of the time*
4	Your health limits you a lot in *moderate activities*	You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems	Your health limits your social activities *most of the time*	You have pain that interferes with your normal work (both outside the home and housework) *moderately*	You feel tense or downhearted and low *most of the time*	You have a lot of energy *a little of the time*
5	Your health limits you *a little in bathing and dressing*		Your health limits your social activities *all of the time*	You have pain that interferes with your normal work (both outside the home and housework) *quite a bit*	You feel tense or downhearted and low *all of the time*	You have a lot of energy *none of the time*
6	Your health limits you *a lot in bathing and dressing*			You have pain that interferes with your normal work (both outside the home and housework) *extremely*		

Appendix 2: Expert Panel Input

Dear member of my expert panel,

Some days ago, I sent you a document explaining the specific inputs I will be requesting from you as a member of an expert panel on issues perceived as important to the health state utility of persons with diabetes.

This document contains my initial suggestions pertaining to the important attributes that may influence the health state utility of persons with diabetes. Your inputs will constitute the first round of inquiry of what is known as the Delphi technique on this topic. As a follow up to this round I will compile each panel member's inputs, at the end of which I will email you a draft of the questionnaire (the health-state utility questionnaire, also called the 'diabetes health-state classification system') that incorporates the opinions of all the participating panel members.

This will initiate another round of inputs from the panel members with the objective of fine-tuning the questionnaire with regards to its wording, or even making modifications to the attributes listed. This will constitute the second round of inquiry of the Delphi technique, following which I will further refine the instrument based on each panel member's responses.

A final round will involve panel members receiving a draft of the revised questionnaire utilizing the inputs obtained in the first two rounds of the Delphi, and will be finalized upon receiving consent from each member.

Thank you for going through this document. I look forward to receiving your responses **BY APRIL 9th, 2007**. If you have any questions, please do not hesitate to contact me.

Sincerely,

Murali Sundaram
PhD Candidate
PSP Dept., WVU School of Pharmacy
P.O. Box 9510
Morgantown, WV - 26506-9510
muralisundaram@hsc.wvu.edu
(304) 685 3106

Michael J Smith
Assistant Professor & PhD Advisor
PSP Dept., WVU School of Pharmacy
P.O. Box 9510
Morgantown, WV - 26506-9510
msmith2@hsc.wvu.edu
(304) 293 1832

A. INFORMATION SOURCE

My input comes from a statistical analysis of items in the Audit of Diabetes Dependent Quality of Life (ADDQoL), an existing instrument assessing the impact of diabetes on quality of life of people with the condition. The ADDQoL asks respondents to rate how the following aspects of life would be influenced if that individual did not have diabetes:

1. Working life
2. Family life
3. Friendship & Social life
4. Sex life
5. Physical appearance
6. Physical activities
7. Vacations & leisure activities
8. Ease of travel
9. Confidence in ability to do things
10. Motivation
11. Way society reacts
12. Worries about the future
13. Finances
14. Dependence
15. Living conditions
16. Freedom to eat
17. Enjoyment of food
18. Freedom to drink

Please consult the document 'Expert panel role.doc' (sent to you in my previous email) for a copy of the instrument and for the exact wording of the above items.

B. PROPOSED ARRANGEMENT OF THE ABOVE ITEMS INTO BROAD ATTRIBUTES

In interpreting the results of the statistical analysis of the ADDQoL items, I also employed my own logic regarding the attributes these items represent, as well as consulted the components of existing instruments assessing health-state utility. Please bear in mind that

- the attributes suggested below are plausible,
- the phrases describing them are my interpretation, and
- the arrangement of items to plausible attributes is not indisputable.

Items		*Plausible attribute*
Freedom to eat Enjoyment of food Freedom to drink	⇨	**Diet-related**

Items		*Plausible attribute*
Work life Family life Friendship & social life Sex Life	⇨	**Relationships**

Items		*Plausible attribute*
Living conditions Finances Way society reacts Dependence Appearance	⇨	**Life situations**

Items	*Plausible attribute*
Ease of travel Physical activities Holidays & leisure activities - - - - - - - - - - Confidence in ability to do things Motivation Worries about the future	**Physical & Mental Functioning ***

***** *Statistical analysis of the ADDQoL items in my dataset did not strongly support separating these set of five items into individual attributes of Physical and Mental Functioning. Theory, however,* <u>*strongly supports*</u> *such a separation. In such a case, the two items below the jagged line in the box to the left would intuitively be regarded as describing mental functioning, while the three items on top of the jagged line would be regarded as describing physical functioning.*

In summary, the initial plausible attributes based on statistical analysis and my own interpretation and intuition are:

Diet-related
Social functioning
Life situations
Physical & Mental functioning
Are there other important BROAD attributes NOT REPRESENTED?

Please continue reading

C. EXPERT'S RECOMMENDATION ON COVERAGE OF INITIAL ATTRIBUTES – TASK 1

TASK 1. The purpose of this task is to ascertain your opinion about whether the attributes listed on page 3 provide adequate coverage of the MAIN areas that would influence the utility that a person with diabetes would place on his or her current health condition. Towards that end, please revisit the items and assess how they are influenced by diabetes, using the format provided.

NOTES:

You are requested to provide your answers in three steps:

- First, rate how important you think each aspect of life, represented by that item, may influence the utility that a person with diabetes places on his or her health condition.
- Second, briefly describe how, in your experience dealing with persons with diabetes, this aspect of life may be influenced.
- Third, please suggest an attribute, from the list on page 3 **OR** any another that you may think of as appropriate, that the item under consideration best fits under.

Item: Working life and work-related opportunities

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced by diabetes.

c. Which attribute do you think this item best fits under?
*(Please suggest one or more from the list on page 3 **OR** suggest another attribute)*

Item: Family life

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Friendships and social life

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Physical appearance

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(<u>Please bold the appropriate option</u>)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(<u>Please suggest one or more from the list on page 3 **OR** suggest another attribute</u>)*

Item: Physical activities

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(<u>Please bold the appropriate option</u>)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(<u>Please suggest one or more from the list on page 3 **OR** suggest another attribute</u>)*

Item: Vacations or leisure activities

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(<u>Please bold the appropriate option</u>)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(<u>Please suggest one or more from the list on page 3 **OR** suggest another attribute</u>)*

Item: Ease of travel (local or long distance)

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(<u>Please bold the appropriate option</u>)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(<u>Please suggest one or more from the list on page 3 **OR** suggest another attribute</u>)*

Item: Confidence in ability to do things

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Motivation to achieve things

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: The way society at large reacts

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Finances

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Dependence on others

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Living conditions

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Freedom to eat

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Enjoyment of food

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Freedom to drink

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Sex life

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

Item: Worries about the future

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
(Please suggest one or more from the list on page 3 ***OR*** *suggest another attribute)*

C. EXPERT'S RECOMMENDATION ON COVERAGE OF INITIAL ATTRIBUTES (contd.) – TASK 2

TASK 2. Based on your clinical experience with this disease, you may be aware of important areas not represented in the ADDQoL that you feel persons with diabetes may consider very important, or that, according to you, is important in the assessment of the utility that persons with diabetes place on their health condition.

NOTES:

- *For example, the items you have considered up until this point do not provide coverage on symptoms of the condition (perhaps because primary research preceding the development of the instrument noted that overall, there aren't distressing symptoms of diabetes that significantly impair patients' QoL; you may or may not agree with this).*
- *If you have in mind attributes that you think are important but are unrepresented in the list of items from the ADDQoL, please feel free to indicate those.*
- *Please indicate AS MANY ITEMS as you think represent aspects that are presently not covered in the items listed in Task 1, and that would impact the utility that persons with diabetes place on their health condition.*

PLEASE LIST THE ITEM HERE: (   )

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition?
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(Please suggest one or more from the list on page 3 **OR** suggest another attribute)*

PLEASE LIST THE ITEM HERE: ( )

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(Please suggest one or more from the list on page 3 **OR** suggest another attribute)*

PLEASE LIST THE ITEM HERE: ( )

a. Using the scale below, rate how important you think this may influence the utility that a person with diabetes places on his or her health condition.
(Please bold the appropriate option)

0	1	2	3
Not at all important	Somewhat important	Important	Very important

b. Please briefly describe how this aspect of life may be influenced in diabetes.

c. Which attribute do you think this item best fits under?
*(Please suggest one or more from the list on page 3 **OR** suggest another attribute)*

Add more items if necessary, 'copy-pasting' the above format

C. EXPERT'S RECOMMENDATION ON COVERAGE OF INITIAL ATTRIBUTES (contd.) – TASK 3

TASK 3: Now, based on your recommendations of the items completed in Task 1 and Task 2, please UPDATE or MODIFY the initial list of attributes indicated on page 3.

In summary, the attributes I consider important in influencing the perceived utility of a person's health condition are:

D. EXPERT'S RECOMMENDATION ON DESCRIPTION OF ATTRIBUTES – TASK 4

TASK 4. Please suggest a description for each attribute that you listed on the previous page. This could be a phrase, a sentence, or a short paragraph. This description will be directly incorporated into the questionnaire being developed.

<u>NOTES</u>:

- *Each attribute can be said to represent an underlying common thread among items. In Task 1 and Task 2, you indicated the impact of diabetes on aspects of life represented by individual ADDQoL items as well as those you additionally listed. You may consider these items in generating your description for each attribute.*

- *Please use a common format (a phrase, a sentence, or a short paragraph) for the descriptions of all attributes.*

- *You can modify the following table to suitably meet your needs.*

Please go to the next page

Name of Attribute	*Description of Attribute*

Add more rows if necessary

E. EXPERT'S RECOMMENDATION ON RATING SCALE FOR ATTRIBUTES – TASK 5

TASK 5. Designate SEPARATE CUSTOMIZED rating scales for the attributes that you listed and described in Task 4. These scales will be reflective of the levels of impairment on each attribute. Please suggest FOUR LEVELS FOR EACH ATTRIBUTE. These rating scales will be directly incorporated into the questionnaire being developed.

NOTES:

- *The type of rating scale to be used for this task is a 'severity scale' where the scale levels for an attribute represent a varying level of impairment on that particular attribute.*

- *Please avoid designating each attribute with generic severity levels such as 'Highly Affected, Moderately Affected, Slightly Affected and Not at all Affected'. While these types of rating scales have been used widely, the interpretation of the relative differences between adjoining scale levels is left to the discretion of the respondent.*

- *Please design a separate 4-level rating scale for each attribute. Please consider using diabetes-specific scenarios (as far as possible) representative of the respective rating scale level rather than using generic severity levels.*

- *To aid you in this task, use your responses in Task 1 and Task 2 and also consult the examples (of the HUI-3, the SF-6D, or the Menopause classification system) provided in the document, 'Expert panel role.doc', emailed or sent to you earlier.*

Please go to the next page

<u>Attribute 1 & Description of Attribute 1:</u>
(Copy and paste Attribute 1 from Task 4)

<u>Rating Scale for Attribute 1:</u>
(Please modify the following table as necessary)

*Severity Level***	*Description of Rating Scale Level*
1	
2	
3	
4	

***** 1 (no impairment) to 4 (highest level of impairment)

<u>Attribute 2 & Description of Attribute 2:</u>
(Copy and paste Attribute 2 from Task 4)

<u>Rating Scale for Attribute 2:</u>
(Please modify the following table as necessary)

*Severity Level**	*Description of Rating Scale Level*
1	
2	
3	
4	

***** 1 (no impairment) to 4 (highest level of impairment)

Attribute 3 & Description of Attribute 3:
(Copy and paste Attribute 3 from Task 4)

Rating Scale for Attribute 3:
(Please modify the following table as necessary)

*Severity Level**	*Description of Rating Scale Level*
1	
2	
3	
4	

***** 1 (no impairment) to 4 (highest level of impairment)

Attribute 4 & Description of Attribute 4:
(Copy and paste Attribute 4 from Task 4)

Rating Scale for Attribute 4:
(Please modify the following table as necessary)

*Severity Level**	*Description of Rating Scale Level*
1	
2	
3	
4	

***** 1 (no impairment) to 4 (highest level of impairment)

Attribute 5 & Description of Attribute 5:
(Copy and paste Attribute 5 from Task 4)

Rating Scale for Attribute 5:
(Please modify the following table as necessary)

*Severity Level***	*Description of Rating Scale Level*
1	
2	
3	
4	

***** 1 (no impairment) to 4 (highest level of impairment)

Please add more tables if necessary

END OF DOCUMENT

Appendix 3: Suggested Version 1

The Current State of Your Health and Well-being

The following five questions assess how diabetes may influence the quality of your health and well-being. There are no correct or wrong answers to these questions since we know that your situation is unique. In answering all these questions, please think about your diabetes, including:

Symptoms	For instance, *when your sugar is low*, you may feel weak or tired, dizzy, irritable, sweaty, weak or tired, hungry, have blurry vision or a headache. On the other hand, *when you sugar is high*, you may feel tired, have extreme thirst or hunger, need to urinate often, have dry skin, blurry vision, or slow-healing wounds.
Complications	For instance, you may think about problems connected with your diabetes, such as, wounds that don't heal, numbness or tingling in your feet, feeling tired, or blurry vision.
Treatment	For instance, you may think about taking medications, checking blood sugar, and watching what you eat every day.

1. First, please tell us how diabetes depletes your energy or restricts you physical ability to perform activities of your choice as well as those required towards maintaining your health.

For example, you may think about

- physical health problems you may experience as a person with diabetes
- the pain associated with your diabetes and its treatment
- physical activities you need to do on a daily basis, your mobility, and travel needs
- activities you need to do to take care of your health, like exercising and seeking medical care

Please read all these statements below and **check ONE box** that most closely describes you:

- ☐ **I have no problems with my physical health nor do I face any difficulty in doing activities as a result of my diabetes.**
- ☐ **I have no no problems with my physical health but I accomplish less in physical activities because of the fatigue that is a result of my diabetes.**
- ☐ **I am quite limited in physical activities because of problems with my physical health and lack of energy due to my diabetes.**
- ☐ **I am unable to perform most of the physical activities I would like to as a result of my diabetes.**

2. Next, please share with us your ability to function and enjoy various interpersonal roles to the extent that you would like to, while you manage your diabetes.

For example, you may think about whether your diabetes
- interferes with you family life and your duties towards your family
- limits your social relationships
- allows you to manage your responsibilities and relationships at work

Please read all these statements below and **check ONE box** that most closely describes you:

- ☐ **I manage my diabetes everyday, and yet am always able to perform and enjoy interpersonal roles to my satisfaction.**
- ☐ **I am not always able to perform and enjoy interpersonal roles to my satisfaction as I manage my diabetes.**
- ☐ **I am quite limited in the extent to which I can participate and enjoy interpersonal roles to my satisfaction because of the demands of my diabetes.**
- ☐ **I am unable to perform and enjoy interpersonal roles to my satisfaction because of the demands of my diabetes.**

3. Living with your diabetes may require you to make some changes to your lifestyle.

For example, you may think about
- your ability to undertake and enjoy leisure time activities
- your ability to perform acts of self-care
- your dependence on others for your every day and diabetes management needs
- your need to balance your finances and living conditions to meet the needs of your situation

Please read all these statements below and **check ONE box** that most closely describes you:

- ☐ **I have complete freedom in living my life just the way I want to even though I have diabetes.**
- ☐ **I am not always able to live my life on my terms because of the need to manage my diabetes on an everyday basis.**
- ☐ **I am quite limited in my ability to live my life the way I would like to because of my diabetes, and it makes me dependent on others.**
- ☐ **I find it very difficult to live my life the way I would like to because of my diabetes, and it makes me very dependent on others.**

4. Now, please tell us about your ability to enjoy food and drink while adapting to different social settings.

For example, you may think about
- the restrictions imposed your diet, the need for you to follow meal plans
- situations when you to make educated food choices

Please read all these statements below and **check ONE box** that most closely describes you:

- ☐ **I always make smart food choices, quantity adjustments, or medication adjustment and always enjoy what I eat and drink anywhere.**
- ☐ **I do not always make educated choices or stick to meal plans but I manage to enjoy food and drink without worrying too much.**
- ☐ **I frequently worry about what the food and drink I consume will do to me but am unsure of what to do to make my food intake better.**
- ☐ **I never enjoy food and drink in all situations..** (*need to expand this?*)

5. Finally, we'd like to understand how the burden/stress of daily living combined with managing your diabetes influences your emotions and health behavior.

For example, you may think about whether you
- feel anxious about the symptoms and management of your diabetes
- worry that you will pass out, or get complications, or in general about the future
- have trouble sleeping, lack in concentration
- feel confident or motivated in doing necessary acts of self-care

Please read all these statements below and **check ONE box** that most closely describes you:

- ☐ **I know that managing my diabetes is a life-long effort. I work on it myself daily, and am pleased to see the progress I have made.**
- ☐ **I know that I can take care of my diabetes, but just need something to jump-start me. I plan ahead and ask for help when I need it.**
- ☐ **I am usually worried that I don't pay attention to my health care behaviors, but I just don't feel confident in putting all the effort that is required.**
- ☐ **I feel very overwhelmed with all the things I need to do to manage my diabetes and worry a lot. Sometimes, I just take a vacation from diabetes.**

Thank you for your participation. Your response is highly appreciated!

Appendix 4: Version One

The Current State of Your Health and Well-being

The following five questions assess how diabetes may influence the quality of your health and well-being. There are no right or wrong answers to these questions since everyone's experience is different. When answering these questions, please think about your diabetes, including:

Symptoms	For instance, what you go through when your blood sugar is low and when your sugar is high.
Complications	For instance, you may think about health problems that are connected with your diabetes.
Treatment	For instance, you may think about taking medications, checking blood sugar, and watching what you eat every day.

1. First, we would like to ask you if diabetes affects your energy or restricts your physical ability to perform everyday activities.

For example, you may think about
- everyday activities like your ability to move around, your travel needs, bathing and dressing
- activities you need to do to take care of your diabetes, like exercising and seeking medical care
- any pain you may have due to your diabetes and in managing it, when you use finger-sticks, or insulin.

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I don't feel tired nor do I have difficulty doing everyday activities even though I have diabetes.**
☐	**I feel tired some of the time and am able to do some everyday activities even though I have diabetes.**
☐	**I feel tired most of the time and am quite limited in everyday activities due to my diabetes.**
☐	**I feel tired all the time and am unable to perform most everyday activities due to my diabetes.**

2. Next, please share with us whether diabetes interferes with your participation in and enjoyment of social activities and relationships.

For example, you may think about whether your diabetes
- interferes with you family life and your role within your family
- limits your social relationships
- allows you to manage your responsibilities and relationships at work
- affects your ability to be intimate with your partner

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I am always able to participate in and enjoy social activities and relationships even though I have diabetes.**
☐	**I am able to participate in and enjoy social activities and relationships some of time even though I have diabetes.**
☐	**I am quite limited in the extent to which I can participate in and enjoy social activities and relationships due to my diabetes.**
☐	**I am very limited in the extent to which I can participate in and enjoy social activities and relationships due to my diabetes.**

Please turn to the back of this page

3. Please tell us about your ability to enjoy food and drink.

For example, you may think about the need to
- make food choices at home, at parties, during the holidays, or at work
- take or adjust your diabetes medication at meal times

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I always enjoy what I eat and drink.**
☐	**I enjoy food and drink some of the time without worrying too much about what I consume will do to me.**
☐	**I am quite limited in my enjoyment of food and drink and frequently worry about what I consume will do to me.**
☐	**I have lost the ability to enjoy food and drink.**

4. Now, kindly share with us about your freedom to live life the way you would like to.

For example, you may think about
- your dependence on others for your every day and diabetes management needs
- your need to balance your finances and living conditions to meet the needs of your situation

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I have complete freedom in living my life just the way I want to even though I have diabetes.**
☐	**I am able to live my life the way I want to some of the time even though I have diabetes.**
☐	**I am quite limited in my ability to live my life the way I want to to because of the need to manage my diabetes on an everyday basis.**
☐	**I find it very difficult to live my life the way I want to because of the need to manage my diabetes on an everyday basis.**

5. Finally, we'd like to understand how the demands of daily living combined with managing your diabetes influences your mood and feelings.

For example, you may think about whether you
- feel confident and motivated or tensed and anxious about managing your diabetes
- feel generally low, have trouble sleeping, lack in concentration
- worry that you will pass out or get complications or worry in general about the future

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I feel confident all of the time and never worry about what will happen to me even though I have diabetes.**
☐	**I feel confident some of the time but do not worry too much about about what will happen to me even though I have diabetes.**
☐	**I just don't feel confident most of the time and am frequently worried about what will happen to me due to my diabetes.**
☐	**I never feel confident, feel very overwhelmed and worry a lot.**

Thank you for your participation. Your response is highly appreciated!

Appendix 5: Suggested Version Two

The Current State of Your Health and Well-being

These **five** questions ask you how diabetes may affect the quality of your health and well-being. Since everyone experiences diabetes in their own way, there are no right or wrong answers to these questions. When you answer these questions, please think about your diabetes and how it affects your health, including:

Symptoms	-	for example, what you may feel when your blood sugar is low or high.
Complications	-	for example, the health problems you may have that are connected with your diabetes, whether it is with your eyes, your hands, your feet, your heart, or your kidneys.
Treatment	-	for example, taking medicines (including insulin), checking blood sugar (including finger sticks), checking your feet, and watching what you eat every day.

Answer the five questions after reading the examples provided below:

1. First, we would like to ask you if diabetes has limited your **physical ability** or caused changes in your **energy level** to do everyday tasks.

For example, you may think about:
- how able you are to do things around the house, move about or travel
- how able you are to take part in relaxing activities, hobbies or exercise
- how much energy you have to do normal everyday tasks
- how any pain you have due to your diabetes affects what you can do

Please read **all** these statements below **and then check ONE box** that most closely describes you:

☐	**I am *always able* to do my everyday tasks even though I have diabetes.**
☐	**I am *usually able* to do my everyday tasks even though I have diabetes.**
☐	**I *usually find it hard* to do my everyday tasks because I have diabetes.**
☐	**I *always find it hard* to do my everyday tasks because I have diabetes.**

2. Next, please tell us whether diabetes gets in the way of your taking part in and enjoying **social activities** and **relationships**.

For example, you may think about:
- your family life and being able to play your role within the family
- your relationships at work and being able to do your job
- your friendships and being able to socialize
- your sex life and ~~being able to be intimate with your partner~~

Please read **all** these statements below **and then check ONE box** that most closely describes you:

☐	**I am *always able* to take part in and enjoy social activities and relationships even though I have diabetes.**
☐	**I am *usually able to* take part in and enjoy social activities and relationships even though I have diabetes.**
☐	**I *usually find it hard* to take part in and enjoy social activities and relationships because I have diabetes.**
☐	**I *always find it hard* to take part in and enjoy social activities and relationships because I have diabetes.**

Please turn to the back of this page

3. Please tell us about your ability to enjoy **food** and **drink**.

> **For example, you may think about:**
> - making food and drink choices at home, at parties, during the holidays, or at work
> - taking or adjusting your diabetes medicines at meal times

Please read **all** these statements below and **then check ONE box** that most closely describes you:

☐	**I am *always able* to enjoy what I eat and drink even though I have diabetes.**
☐	**I am *usually able* to enjoy what I eat and drink even though I have diabetes.**
☐	**I *usually find it hard* to enjoy what I eat and drink because I have diabetes.**
☐	**I *always find it hard* to enjoy what I eat and drink because I have diabetes.**

4. Now, tell us how **the way you live your life** has been affected by your diabetes.

> **For example, you may think about:**
> - how able you are to take care of yourself, like bathe, dress or feed yourself
> - how able you are to do things to care for your diabetes on a daily basis
> - Any other things I should mention to clarify the concept? (It is anchored by how life changes due to diabetes)

Please read all these statements below and **check ONE box** that most closely describes you:

☐	**I am *always* able to live my life just the way I want to even though I have diabetes.**
☐	**I am *usually* able to live my life the way I want to even though I have diabetes.**
☐	**I *usually find it hard* to live my life the way I want to because I have diabetes.**
☐	**I always *find it hard* to live my life the way I want to because I have diabetes.**

5. Finally, we'd like to understand how living with diabetes influences your **mood** and **feelings**.

> **For example, you may think about:**
> - how happy or how sad you feel in general, and if you are moody or irritable at times
> - whether you find to hard to pay attention
> - whether you feel confident about yourself or worry about what will happen because of your diabetes

Please read **all** the statements below and **then check ONE box** that most closely describes you:

☐	**I *always feel confident* about myself and *never worry* about what will happen to me even though I have diabetes.**
☐	**I *usually feel confident* about myself and *worry little* about what will happen to me even though I have diabetes.**
☐	**I *usually don't feel confident* about myself and *often worry* about what will happen to me because I have diabetes.**
☐	**I *never feel confident* about myself and *always* worry about what will happen to me because I have diabetes.**

Thank you for your participation. We appreciate your time and efforts!

Appendix 6: Version Two

The Current State of Your Health and Well-being

These **five** questions ask you how diabetes may affect the quality of your health and well-being. Since everyone experiences diabetes in their own way, there are no right or wrong answers to these questions. When you answer these questions, please think about your diabetes, including:

Symptoms	-	for example, what you may feel when your blood sugar is low or high.
Complications	-	for example, the health problems you may have that are connected with your diabetes, whether it is with your eyes, your hands, your feet, your heart, or your kidneys.
Treatment	-	for example, taking medicines (including insulin), checking blood sugar (including finger sticks), checking your feet, and watching what you eat every day.

Answer the five questions after reading the examples provided below:

1. First, we would like to ask you if diabetes has limited your **physical ability** or caused changes in your **energy level** to do everyday tasks.

For example, you may think about:

- how able you are to do things around the house, move about or travel
- how able you are to take part in relaxing activities, hobbies or exercise
- how much energy you have to do these normal everyday tasks
- how any pain you have due to your diabetes affects what you can do

Think of these examples and **then check ONE box** below that most closely describes your physical ability and energy level:

☐	**I am *always able* to do my everyday tasks even though I have diabetes.**
☐	**I am *usually able* to do my everyday tasks even though I have diabetes.**
☐	**I *usually find it hard* to do my everyday tasks because I have diabetes.**
☐	**I *always find it hard* to do my everyday tasks because I have diabetes.**

2. Next, please tell us whether diabetes gets in the way of your taking part in and enjoying **social activities** and **relationships**.

For example, you may think about:

- your family life and being able to play your role within the family
- your relationships at work and being able to do your job
- your friendships and being able to socialize
- your sex life and being able to be intimate with your partner

Think of these examples and **then check ONE box** below that most closely describes your relationships:

☐	**I am *always able* to take part in and enjoy social activities and relationships even though I have diabetes.**
☐	**I am *usually able to* take part in and enjoy social activities and relationships even though I have diabetes.**
☐	**I *usually find it hard* to take part in and enjoy social activities and relationships because I have diabetes.**
☐	**I *always find it hard* to take part in and enjoy social activities and relationships because I have diabetes.**

Please turn to the back of this page

3. Please tell us about your ability to enjoy **food** and **drink**.

For example, you may think about:
- making food and drink choices at home, at parties, during the holidays, or at work
- taking or adjusting your diabetes medicines at meal times

Think of these examples and **then check ONE box** below that most closely describes how you enjoy food and drink:

☐	**I am *always able* to enjoy what I eat and drink even though I have diabetes.**
☐	**I am *usually able* to enjoy what I eat and drink even though I have diabetes.**
☐	**I *usually find it hard* to enjoy what I eat and drink because I have diabetes.**
☐	**I *always find it hard* to enjoy what I eat and drink because I have diabetes.**

4. Now, tell us how your freedom **to live life the way you like to** has been affected by your diabetes.

For example, you may think about:
- how able you are to take care of yourself, like bathe, dress or feed yourself
- how able you are to do things to care for your diabetes on your own everyday

Think of these examples and **then check ONE box** below that most closely describes you:

☐	**I am *always* able to live my life just the way I want to even though I have diabetes.**
☐	**I am *usually* able to live my life the way I want to even though I have diabetes.**
☐	**I *usually find it hard* to live my life the way I want to because I have diabetes.**
☐	**I always *find it hard* to live my life the way I want to because I have diabetes.**

5. Finally, we'd like to understand how living with diabetes influences your **mood** and **feelings**.

For example, you may think about:
- how happy and calm, or how low and depressed you feel
- whether you feel moody or irritable, find it hard to pay attention to what you do
- whether you feel confident in general or worry about what will happen because of your diabetes

Think of these examples and **then check ONE box** below that most closely describes your feelings:

☐	**I *always feel confident* about myself and *never worry* about what will happen even though I have diabetes.**
☐	**I *usually feel confident* about myself and *worry little* about what will happen even though I have diabetes.**
☐	**I *usually don't feel confident* about myself and *often worry* about what will happen because I have diabetes.**
☐	**I *never feel confident* about myself and *always* worry about what will happen because I have diabetes.**

Thank you for your participation. We appreciate your time and efforts!

Appendix 7: Final Version:

The Current State of Your Health and Well-being

- These **five** questions ask you about how **diabetes** may affect the quality of your health and well-being.
- There are no right or wrong answers to these questions.

When you answer these questions, please think about your diabetes, in terms of:

Symptoms - for example, what you may feel when your blood sugar is low or high

Complications - for example, the health problems (such as with your eyes, hands, feet, heart, or kidneys) and pain you may have because of your diabetes

Daily care - for example, taking medicines (including insulin), checking blood sugar (including doing finger sticks), checking your feet, and watching what you eat every day.

1. First, we would like to ask you if diabetes has limited your **physical ability** or caused changes in your **energy level** to do everyday tasks.

Think of the following as they apply to you:

- how able you are to move about, do things around the house, take part in hobbies, exercise, and travel
- how able you are to take care of yourself, like bathe, dress or feed yourself
- how much energy you have to do these normal everyday tasks
- how any pain you may have due to your diabetes affects what you physically can do

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your **physical ability** and **energy level**:

☐ I am ***always able*** to do my everyday tasks even though I have diabetes.

☐ I am ***usually able*** to do my everyday tasks even though I have diabetes.

☐ I ***usually find it hard*** to do my everyday tasks because I have diabetes.

☐ I ***always find it hard*** to do my everyday tasks because I have diabetes.

---

2. Next, please tell us whether diabetes gets in the way of your **social interactions** or **relationships.**

Think of the following as they apply to you:

- your family life and being able to play your role within the family
- your relationships with co-workers and your roles in the workplace
- your friendships and being able to socialize with friends and other people
- your sex life and being able to be intimate with your partner

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your **social interactions** or **relationships** with family, friends and co-workers:

☐ My social interactions or relationships are ***never affected*** even though I have diabetes.

☐ My social interactions or relationships are ***usually not affected*** even though I have diabetes.

☐ My social interactions or relationships are ***usually affected*** because I have diabetes.

☐ My social interactions or relationships are ***always affected*** because I have diabetes.

Please turn to the next page

3. We'd like to know whether living with diabetes affects your **mood** and **feelings**.

Think of the following as they apply to you:

- how happy or how sad you feel in general, and how moody you may be
- whether you feel calm in general or feel irritable at times
- how much you worry about what will happen to your health and future because of your diabetes

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your **feelings**:

☐ I ***always feel happy*** in general and ***never worry*** about what will happen to my health and future even though I have diabetes.

☐ I ***usually feel happy*** in general *and* ***worry little*** about what will happen to my health and future even though I have diabetes.

☐ I ***usually don't feel happy*** in general and ***often worry*** about what will happen to my health and future because I have diabetes.

☐ I ***never feel happy*** in general and ***always worry*** about what will happen to my health and future because I have diabetes.

---

4. Now, tell us whether diabetes affects your ability to enjoy **food** and **drink**.

Think of the following as they apply to you:

- having to make food and drink choices at home, at parties, at restaurants, during the holidays, or at work
- taking or adjusting your diabetes medicines at meal times

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your enjoyment of **food** and **drink**:

☐ I am ***always able*** to enjoy what I eat and drink even though diabetes limits my choices.

☐ I am ***usually able*** to enjoy what I eat and drink even though diabetes limits my choices.

☐ I ***usually find it hard*** to enjoy what I eat and drink because diabetes limits my choices.

☐ I ***always find it hard*** to enjoy what I eat and drink because diabetes limits my choices.

---

5. Finally, tell us how **satisfied** you are with how you **manage** the **changes in your life** because of your diabetes.

Think of the following as they apply to you:

- how much you have accepted the changes in your life because of your diabetes
- how much you do what is necessary and important to care for your diabetes
- how confident and satisfied you are in your ability to care for your diabetes on a daily basis

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes how **satisfied** you are with how you **manage** your diabetes:

☐ I am ***very satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

☐ I am ***somewhat satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

☐ I am ***not satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

Thank you for your participation. We appreciate your time and efforts!

Appendix 8: Preference Elicitation Interview Booklet

Living With Diabetes

RESPONDENT I.D.

MALE / FEMALE

Group A / B

INTERVIEWER

START TIME: _________ a.m. / p.m.

INTERVIEW DATE:

END TIME: _________ a.m. / p.m.

SECTION 1: The Current State of Your Health and Well-being

- These **five** questions ask you about how **diabetes** may affect the quality of your health and well-being.
- There are no right or wrong answers to these questions.

When you answer these questions, please think about your diabetes, in terms of:

Symptoms - for example, what you may feel when your blood sugar is low or high

Complications - for example, the health problems (such as with your eyes, hands, feet, heart, or kidneys) and pain you may have because of your diabetes

Daily care - for example, taking medicines (including insulin), checking blood sugar (including doing finger sticks), checking your feet, and watching what you eat every day.

1. First, we would like to ask you if diabetes has limited your **physical ability** or caused changes in your **energy level** to do everyday tasks.

Think of the following as they apply to you:

- how able you are to move about, do things around the house, take part in hobbies, exercise, and travel
- how able you are to take care of yourself, like bathe, dress or feed yourself
- how much energy you have to do these normal everyday tasks
- how any pain you may have due to your diabetes affects what you physically can do

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your **physical ability** and **energy level**:

- ☐ I am ***always able*** to do my everyday tasks even though I have diabetes.
- ☐ I am ***usually able*** to do my everyday tasks even though I have diabetes.
- ☐ I ***usually find it hard*** to do my everyday tasks because I have diabetes.
- ☐ I ***always find it hard*** to do my everyday tasks because I have diabetes.

---

2. Next, please tell us whether diabetes gets in the way of your **social interactions** or **relationships.**

Think of the following as they apply to you:

- your family life and being able to play your role within the family
- your relationships with co-workers and your roles in the workplace
- your friendships and being able to socialize with friends and other people
- your sex life and being able to be intimate with your partner

Think ⇦ about

Based on the above, **check ONE box** below that most closely describes your **social interactions** or **relationships** with family, friends and co-workers:

- ☐ My social interactions or relationships are ***never affected*** even though I have diabetes.
- ☐ My social interactions or relationships are ***usually not affected*** even though I have diabetes.
- ☐ My social interactions or relationships are ***usually affected*** because I have diabetes.
- ☐ My social interactions or relationships are ***always affected*** because I have diabetes.

---

3. We'd like to know whether living with diabetes affects your **mood** and **feelings**.

Think of the following as they apply to you:

- how happy or how sad you feel in general, and how moody you may be
- whether you feel calm in general or feel irritable at times
- how much you worry about what will happen to your health and future because of your diabetes

Think about

Based on the above, **check ONE box** below that most closely describes your **feelings**:

☐ I ***always feel happy*** in general and ***never worry*** about what will happen to my health and future even though I have diabetes.

☐ I ***usually feel happy*** in general *and* ***worry little*** about what will happen to my health and future even though I have diabetes.

☐ I ***usually don't feel happy*** in general and ***often worry*** about what will happen to my health and future because I have diabetes.

☐ I ***never feel happy*** in general and ***always worry*** about what will happen to my health and future because I have diabetes.

---

4. Now, tell us whether diabetes affects your ability to enjoy **food** and **drink**.

Think of the following as they apply to you:

- having to make food and drink choices at home, at parties, at restaurants, during the holidays, or at work
- taking or adjusting your diabetes medicines at meal times

Think about

Based on the above, **check ONE box** below that most closely describes your enjoyment of **food** and **drink**:

☐ I am ***always able*** to enjoy what I eat and drink even though diabetes limits my choices.

☐ I am ***usually able*** to enjoy what I eat and drink even though diabetes limits my choices.

☐ I ***usually find it hard*** to enjoy what I eat and drink because diabetes limits my choices.

☐ I ***always find it hard*** to enjoy what I eat and drink because diabetes limits my choices.

---

5. Finally, tell us how **satisfied** you are with how you **manage** the **changes in your life** because of your diabetes.

Think of the following as they apply to you:

- how much you have accepted the changes in your life because of your diabetes
- how much you do what is necessary and important to care for your diabetes
- how confident and satisfied you are in your ability to care for your diabetes on a daily basis

Think about

Based on the above, **check ONE box** below that most closely describes how **satisfied** you are with how you **manage** your diabetes:

☐ I am ***very satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

☐ I am ***somewhat satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

☐ I am ***not satisfied*** with what I do in order to take care of my diabetes the way I am supposed to.

SECTION 2: Table 1

Card #	Section	Card Code	Card Technical Name	Thermometer Score
1	VI	PH	Perfect Health	
2	VI	PIT	Pits	
3	VI	D	Death	
4	VII	MA	Marker State 1	
5	VII	MB	Marker State 2	
6	VII	MC	Marker State 3	
7	I	SPC	Physical Corner State	
8	I	SP2	Physical Single-Attribute Level State 2	
9	I	SP3	Physical Single-Attribute Level State 3	
10	II	SSC	Social Corner State	
11	II	SS2	Social Single-Attribute Level State 2	
12	II	SS3	Social Single-Attribute Level State 3	
13	III	SFC	Feelings Corner State	
14	III	SF2	Feelings Single-Attribute Level State 2	
15	III	SF3	Feelings Single-Attribute Level State 3	
16	IV	SDC	Diet Corner State	
17	IV	SD2	Diet Single-Attribute Level State 2	
18	IV	SD3	Diet Single-Attribute Level State 3	
19	V	SMC	Satisfaction-Management Corner State	
20	V	SM2	Satisfaction-Management Single-Attribute Level State 2	

Table 1 finished at _____ a.m. / p.m.

SECTION 2: Table 2

Card MA

RECORD SCORE INDICATED ON BOARD = __.__ __

? – PROMPT “Why did you choose a 100% chance of health B rather than a 100% chance of perfect health?
RECORD VERBATIM RESPONSE**:**

?? – ☐ B more preferable than Death - 0.05
☐ B less preferable than Death - less than 0.00
☐ B and Death equally preferable - 0.00

Card MB

RECORD SCORE INDICATED ON BOARD = __.__ __

? – PROMPT “Why did you choose a 100% chance of health B rather than a 100% chance of perfect health?
RECORD VERBATIM RESPONSE**:**

?? – ☐ B more preferable than Death - 0.05
☐ B less preferable than Death - less than 0.00
☐ B and Death equally preferable - 0.00

Card MC

RECORD SCORE INDICATED ON BOARD = __.__ __

? – PROMPT “Why did you choose a 100% chance of health B rather than a 100% chance of perfect health?
RECORD VERBATIM RESPONSE**:**

?? – ☐ B more preferable than Death - 0.05
☐ B less preferable than Death - less than 0.00
☐ B and Death equally preferable - 0.00

Card Death / Pits *(circle appropriate option)*

RECORD SCORE INDICATED ON BOARD = __.__ __

? – PROMPT “Why did you choose a 100% chance of health B rather than a 100% chance of perfect health?
RECORD VERBATIM RESPONSE**:**

?? – ☐ B more preferable than Death - 0.05
☐ B less preferable than Death - less than 0.00
☐ B and Death equally preferable - 0.00

Table 2 finished at _____ a.m. / p.m.

SECTION 3: About You and About the Study

In the next stage of the interview, I will collect demographic information and your opinions on the interview.

<u>Socio-demographics</u>:

1. Gender: CHECK Male or Female
   ( ) Male [1]
   ( ) Female [2]

2. How old are you? ____________

3. Which of the following would you say is your race?

   ( ) White/ Caucasian [1]
   ( ) Black/African American [2]
   ( ) Hispanic/Latino [3]
   ( ) Pacific Islander [4]
   ( ) Asian [5]
   ( ) American Indian/ Native Indian/ Alaskan Native [6]
   ( ) Other (please specify) _______________ [7]
   ( ) Refused [0]

4. What is your current marital status?

   ( ) Single [1]
   ( ) Married [2]
   ( ) Divorced or Separated [3]
   ( ) Widowed [4]
   ( ) Not married, living with partner [5]
   ( ) Other (please specify) _________________ [6]
   ( ) Refused [0]

5. What is the highest level of education you have completed?

   ( ) Some high school or less [1]
   ( ) High school graduate or GED [2]
   ( ) Vocational college or some college [3]
   ( ) College degree [4]
   ( ) Professional or graduate degree [5]
   ( ) Other (please specify) ___________________ [6]
   ( ) Refused [0]

6. Which of the following best describes your main activity?

( ) Employed (including self employment) [1]
( ) Employed part-time [2]
( ) Retired [3]
( ) Keeping house/ home-maker [4]
( ) Student [5]
( ) Seeking work [6]
( ) Other (please specify) ________________ [7]
( ) Refused [0]

7. What is your annual household income from all sources? (Please check one)

( ) $25,000 or less [1]
( ) $25,001 - $50,000 [2]
( ) $50,001 - $75,000 [3]
( ) More than $75,000 [4]
( ) Refused [0]
( ) Don't know --- PROBE: If you had to guess, what would you say?

<u>Opinions about the interview</u>:
The final questions ask for your opinions of this interview. We constantly try to improve our method of collecting information. Any assistance you may provide would be appreciated.

8. How would you rate the Feeling Thermometer and questions?

( ) Very easy to understand [1]
( ) Easy to understand [2]
( ) Neither easy nor difficult [3]
( ) Difficult [4]
( ) Very difficult [5]
( ) Refused [0]
( ) Don't Know [99]

9. How would you rate the Chance Board and its questions?

( ) Very easy to understand [1]
( ) Easy to understand [2]
( ) Neither easy nor difficult [3]
( ) Difficult [4]
( ) Very difficult [5]
( ) Refused [0]
( ) Don't Know [99]

10. Thank you very much. Is there anything else you would like· to say or add about the interview?

THIS PART BELOW TO BE COMPLETED BY INTERVIEWER

11. Did anyone other than the respondent contribute information?

Please explain

( ) 1. YES
( ) 2. NO

12. Rate degree of co-operation from the respondent in the following categories:

( ) 1. Complete co-operation
( ) 2. General co-operation – not fully open on all questions
( ) 3. Substantial lack of co-operation

13. How much thought did the respondent put into his/ her answers?

( ) 1. A great deal
( ) 2. Some
( ) 3. Very little
( ) 4. None at all

14. How well did the respondent understand the questions?

( ) 1. Totally
( ) 2. For the most part
( ) 3. Somewhat
( ) 4. Only a little
( ) 5. Not at all

15. How much trouble did the respondent have in answering these questions?

( ) 1. A great deal
( ) 2. Some
( ) 3. Very little
( ) 4. None at all

16. Record impression of the interview (i.e., the quality of response):

( ) 1. Very good
( ) 2. Good
( ) 3. Average
( ) 4. Poor
( ) 5. Very Poor

Comments