

Utility Assessment Based On Individualized Patient Perspectives

by

Duane A. Steward

B. S. Biology
Florida State University, 1976

D. V. M.
College of Veterinary Medicine, University of Florida, 1980

M. S. I. E.
College of Engineering, Florida A&M University - Florida State University, 1994

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 1998

© 1998 Duane A. Steward. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature of Author:

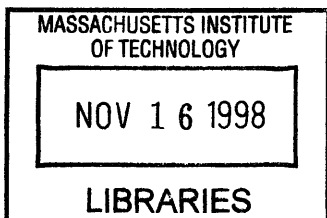
.....
Department Of Electrical Engineering And Computer Science
September 4, 1998

Certified by:

.....
Peter Szolovits
Professor of Computer Science and Engineering
Thesis Supervisor

Accepted by:

.....
Arthur C. Smith
Chairman, Committee on Graduate Students
Department of Electrical Engineering and Computer Science



UTILITY ASSESSMENT BASED ON INDIVIDUAL PATIENT PERSPECTIVES

by

Duane A. Steward

Submitted to the Department of Electrical Engineering
and Computer Science on September 4, 1998 in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

ABSTRACT

The feasibility of using the terminology of the individual patient to capture and express the individual's particular, perhaps unique, value system for healthiness is explored as a means to evaluate the quality of life. The hypothesis is that a systematic approach that treats the patient as a competent expert on a patient's perception of what it means to be healthy is a viable approach to his or her values. It is anticipated that eliciting such knowledge in a form that health care providers can use will enhance medical decision making. A systematic and rigorous protocol is described consisting of a reconstruction of utility assessment using traditional methodological building blocks applied to the descriptors elicited from the individual.

The representation of values is multidimensional. Ordered nominal scales are constructed from the words of the individual's description of familiar people in a structured interview. A complete list of comprehensive scales is composed as indicated by the responses of the individual to hypothetical decisions involving tradeoffs. The result is a scoring system for health state descriptions suitable to represent values for the outcomes in medical decision models constructed by the medical community. The output is a patient preference model referred to as an Individualized Multidimensional Quality of Life (IMQOL) model. This model also provides a means to describe and rank potential outcomes from the same individualized perspective.

Feasibility is explored by empirical evaluation of sixteen interviews of dialysis patients with the IMQOL protocol and applying the resulting model to the patient's own health as well as four other states of health common in dialysis therapy. Comparison is made to results of quality of life assessment with standard gamble and time tradeoff methods in the same patients for the same described outcomes.

Qualitative responses from patients regarding their confidence in the representation of their values are rewarding. Results are quantitatively comparable to traditional utility assessment. A prototypic computer program is used to substantiate the programmability and potential for automation. The protocol expands the information contributing to understanding by both the health care provider and the patient. Future evaluation and extension are discussed.

Thesis Supervisor: Peter Szolovits

Title: Professor of Computer Science and Engineering

Biographical Note

The author began his professional career as a veterinarian in a mixed practice in the islands off South Florida (1980-81) after receiving a B.S. degree from Florida State University (1976) and a D.V.M. degree from the College of Veterinary Medicine at the University of Florida (1980). Before finishing the professional degree, training included the domains of exotic pets and marine mammals (specialty intramural tracts and AquaVet, Woods Hole, Mass. 1979). After a stint of mixed small animal and brood mare practice in Ocala, Florida (1981-82), the author established a house-call veterinary practice for small animals in Tallahassee, Florida (1983-94).

Not only did this architecture better suit the author's desires to attend the needs of his clients and patients on their own terms, it afforded freedom to maintain a higher standard of continuing education. The author's first professional diversion from routine small animal and exotics practice was in the domain of human-animal bond research. He provided literature review and volunteer training as a founding consultant to a local pet-facilitated therapy program organized by a state agency for the aging. Wondering why engineers had so many problem-solving techniques that enabled their professional standards of safety and certainty and why these techniques were not used more in application to medical problem solving, the author's next professional expansion became the pursuit of a Master's degree in Industrial Engineering. The thesis of this research was *The Application of Simulation to the Design & Operation of Veterinary Practice*. This degree was earned from the Florida A&M – Florida State University College of Engineering while still practicing veterinary medicine full time. This was made possible by the emerging technologies of cellular telephone communication and portable computers. In support of his practice, the author developed a paperless medical record system during the mid-1980's that facilitated this dual pursuit. Also emerging at that time was the frontier discipline, called "medical informatics", that proved to be the perfect domain to unite clinical experience and motivation with the new training in industrial engineering and a growing experience with computer science.

Upon completing the Master's degree in 1994, the author left veterinary practice to begin a fellowship in the Harvard/MIT/New-England-Medical-Center medical informatics research training program funded by the National Library of Medicine. As a member of the Clinical Decision Making Group of the Laboratory for Computer Science at Massachusetts Institute of Technology, supervised by Peter Szolovits, this thesis marks the successful completion of a Doctor of Philosophy degree in the field of Computer Science from that institution in September, 1998. While undergoing this training, the author was also informally participating as a fellow in the Clinical Decision Making Division of the New England Medical Center, directed by Steven Pauker and John Wong.

Professional appointments, teaching experience and publications follow.

Professional Appointments:

- Program Committee Chair for Talbot Symposium on Veterinary Informatics at the American Veterinary Medical Association Annual Convention, Baltimore, Md. 1998
- American Academy of Veterinary Informatics (Charter Fellow, Advisory Board) 1993
- Veterinary Informatics Consultant for Veterinary Information Network February 1991 - present
- American Association of Housecall Veterinarians (Founding Board Member) 1990-1994
- Chaired Session at Delta Society Conference at Eastern States Veterinary Conference: "Pets Helping People" 1988
- Co-Chairman Fla. Vet. Med. Assoc. Committee on the Human-animal bond 1988

Teaching Experience

- MIT Undergraduate Research Opportunity Project supervisor for 7 students 1997-8
- Lecturer on "Web Search Engines" and "Applied Simulation In Aquaculture" at annual AquaVet program in Woods Hole
- Speaker at North American Veterinary Conference on "Simple Software Tools that are Useful to the Veterinary Practice", & "Integrating software 'Extras with your Practice Management Software" Jan 1993 Orlando. By request. (proceedings)
- Course Coordinator of "Successful Use of Information Technology in Veterinary Medicine", Online Continuing Education course of Veterinary Information Network Sept 1994, Oct 1995.
- Biology for non-majors - laboratory as a Senior undergraduate at FSU
- Trained Volunteers for Pet-facilitated Therapy (Monthly lecture)
- Lectures on Human/Animal Bond to Big Bend Vet. Med. Soc. and other local (to Tallahassee) groups
- Speaker at Delta Society/ComForT miniconference on Human/Animal Bond May 1991 Tallahassee

Publications

- Steward, Duane and Davis, Mark. Qualification of Discordant Responses in Utility Assessment, submitted to AMIA Fall Symposium.
- Steward, Duane. Seeing the Forest for the Trees: Decision Analysis for Veterinarians, Convention Notes of the AVMA Annual Convention, AVMA, July, 1998.
- Steward, Duane. Utility Assessment Based on Individualized Patient Perspectives, Proceedings of the Spring Symposium on Interactive and Mixed-Initiative Decision-Theoretic Systems, AAAI, March, 1998.
- Steward, Duane. What is Computer-Aided Diagnosis?, in Murtaugh RJ, Ferguson D, Pion P (ed), Seminars In Veterinary Medicine and Surgery (Small Animal), May, 1996.
- Steward, Duane. The Clinician Computer, in Murtaugh RJ, Ferguson D, Pion P (ed), Seminars In Veterinary Medicine and Surgery (Small Animal), May, 1996.
- Steward, Duane and Standridge, Charles R. The Application of Simulation to Veterinary Practice Management, Journal of the American Veterinary Medical Association, Mar. 15, 1996.
- Steward, Duane and Standridge, Charles R. Demonstrations of the use of Simulation in Veterinary Practice Management, Journal of the American Veterinary Medical Association, Mar. 15, 1996.
- Steward, Duane and Standridge, Charles R. A Veterinary Practice Simulator Based on the Integration of Expert System and Process Modeling, Simulation, March, 1996.
- Steward, Duane, What is Your Clinic Staff's Computer Skills Inventory?, Veterinary Forum, Sept. (1994).
- Steward, Duane, It Can Make A 'Horse of a Difference', Veterinary Forum, Oct. (1993).
- Steward, Duane, Information Wars, Veterinary Forum, Sept. (1993).
- Steward, Duane, At Home Veterinary Services bimonthly newsletter to clients, Jan. 1984 - June 1994.

ACKNOWLEDGMENTS

If more people knew the impact they had on others, I can not help but think we would all redefine what is meant by miracles. I have, no doubt, been propelled by the encouragement of those who thought me capable of more. “Plans fail for lack of counsel, but with many advisers they succeed” (Proverbs 15:22).

My first respects must go to two men in memoriam who have left their mark on me. Wyllis Bandler told me so many of my ideas were interesting that I was forced to ask just how interesting. Richard Talbot, considered by many to be the grandfather of veterinary informatics, was instrumental in suggesting there was a place for me to contribute to the same discipline.

Equal in reputation to Richard Talbot, Al Hahn has honored me on many occasions with the respect of a mentor who believes in building leadership by calling it out and then following in support. Al’s faith in me was the thing that caused me to apply to prestigious graduate programs despite previous admissions experience.

I owe a great deal to my Master’s degree supervisor, Charles Standridge, who has remained supportive beyond his due. Since the first decision analysis seminar I attended, John Wong, Steven Pauker and those at the New England Medical Center’s Division of Clinical Decision Making have been a joy and privilege to learn from and work with. This thesis would not have gotten off the ground without my recurrent use of the Division members and its fellows as sounding boards for my out-of-bounds ideas.

Mark Davis and Geoff Rutledge of the Beth Israel Deaconess Hospital Emergency Department contributed repeatedly with support, confidence and critique that proved invaluable in the study of stroke outcomes.

Klemens Meyer, as director of the Dialysis Center, Incorporated, at New England Medical Center, recruited study participants and provided powerful clinical insights. As collaborator, he contributed to the experimental design of the feasibility study with dialysis patients and its ongoing progress. His enthusiasm continues to be energizing.

Many at the Institute have shaped this work. In my first few months as a student of Peter Solovits, I asked what motivated him to benevolence in life. He replied with a statement of hope that people would listen and put to use what he had to offer. I hope he is rewarded seeing me as having done just that. His confidence in this pursuit far afield of others in the training program enabled this trek off the beaten path. Isaac Kohane, Bill Long and Jon Doyle have invested time, contributing question and comment, to improve the presentation of my ideas, adding to their clarity and success. Patrick Winston, graciously accepting my invitation to serve on every exam committee of my graduate student tenure, left his indelible mark on my presentation skills. I will forever seek to reward his personal instruction on presentation by putting it into practice. If anyone is to gain from my research ideas, these educators are to be thanked.

Values assessment cannot be studied without values. I must recognize Thom Miller, Gordon MacDonald and David Johnson for the spiritual discipleship that enables my appreciation of value as a citizen in the Kingdom of Heaven and the importance of the individual. I also gained my appreciation of individuals by serving the clientele of At Home Veterinary Services in Tallahassee, Florida, before this thesis study. These supportive people graciously allowed me to leave a successful practice with their blessing and encouragement after bestowing the intimate honor of allowing me to care for their animal companions. I could not have elicited values without volunteers to interview: the patients of the Dialysis Clinic, Incorporated, at the New England Medical Center and Beth Israel Deaconess Hospital Emergency Department. Undergraduate Research Opportunity Project (UROP) students, Annie Thompson and Richard Chen must be credited with excellent work conducting the interviews of the stroke outcomes study. Richard Chen was also instrumental working with me to develop the S.O.L.O.M.O.N. computer program. He is largely responsible for bringing it to a prototypic functional status from my initial work with the assistance of additional UROP students: Jeffery Sheldon, Jie Zhang, Janelle Prevost and Chris Rhohrs.

As a colleague on the frontier of veterinary informatics, Eric Mills has been a trustworthy friend. His camaraderie as a brother—intellectually, emotionally and spiritually—has kept this labor from being one of isolation and loneliness, preserving my sanity and positive passion in periods of travail.

With unreasonable frequency, Scott Stewart, who married my sister, remedied frustration with computer hardware and software. He and a sealed window have prevented numerous “Chicken Little” injuries on the street below my office. His wife as well as my mother and father have not wavered in their supply of confidence despite any of my misgivings.

My wife, Cheree, has been the most confident that “...He who began a good work in you will carry it on to completion until the day of Christ Jesus” (Philippians 1:6). After 23 years of marriage and this thesis for an anniversary present, we both are still unveiling what friendship can be. I thank my two sons, Shane and Cody, as well, because it is for their future that I wish to uncover and pass on more problem-solving skills.

TABLE OF CONTENTS

ABSTRACT	3
Biographical Note	5
Acknowledgments	7
List of figures	13
List of tables	17
1 Introduction	19
1.1 Motivation	20
1.2 Utility Assessment to Date	21
1.3 The Shortfall	22
1.4 Aims of This Research	24
1.5 Overview of Thesis	25
2 Motivating Examples	27
2.1 Tree Surgeon Facing Otic Neuroma Decision – Individualized Issues	27
2.2 Acute Stroke Treatment And Preference Between Disability And Death – Extreme Tradeoffs	30
2.3 Hemodialysis Versus Peritoneal Dialysis - Complex Tradeoffs	31
3 Background and Significance	39
3.1 Why patient values?	39
3.2 Physicians don’t make decisions based on patient values?	40
3.3 What kept others before me from acquiring and using patient specific values in their own words?	41
3.4 What prior accomplishments do exist in this area?	42
3.5 What is missing?	43
3.6 What stopped others?	43
3.7 How is what this thesis offers different?	44
3.8 How can this method of values clarification be validated?	46
3.9 Why so many outcome measures?	47
3.10 Why dialysis patients?	48
4 Discordant Responses in Classical Assessments	51
4.1 Discord is Endemic	53
4.2 Not all discord is the same	55
4.2.1 Utility assessment for consequences of stroke: Methods	55
4.2.2 Utility assessment for consequences of stroke: Results	58
4.2.3 Utility assessment for consequences of stroke: Discussion	65
4.3 We understand discord poorly	73
4.4 We need a systematic approach to fill in our lack of understanding	74
4.5 How does this proposed values representation and elicitation fill in our understanding?	75
4.5.1 Decomposable model	75
4.5.2 Words of the individual	76
4.5.3 A Complete, Comprehensive and Operational Utility Model	76
4.5.4 Example from early investigation	77
4.6 Summary	79

TABLE OF CONTENTS (continued)

5	Patient Perspective Representation	81
5.1	Multiple Factors Contribute to Values of Health.....	82
5.2	A Complete List of Relevant Issues	85
5.3	Getting the Right Issue	86
5.4	Comprehensive Attributes	87
5.5	Measurable Attributes are Operational	89
5.6	Redundancy?.....	92
5.7	Minimal Dimensions?.....	92
5.8	Recomposing the Whole from the Parts.....	93
5.9	Descriptive Versus Normative or Prescriptive Representation	97
5.10	Labile Values	98
5.11	Summary	98
6	Elicitation of Patient Perspective	99
6.1	Determination of Attributes.....	100
6.2	Determination of intermediate level values for attribute scales	103
6.3	Determination of a combining model for health state scoring.	105
6.4	Rationale for the Protocol.....	108
6.4.1	Alternatives Considered for Eliciting Scaling Coefficients.....	109
6.4.2	Will Difficulties of Utility Assessment Accrue?.....	113
6.4.3	Assuming that the patient-given attributes cover all possible health concerns.....	113
7	Application of Elicited Results.....	117
7.1	Rating the Quality of Life for Outcomes	118
7.2	Illustrated Application of Elicited Values.....	120
7.3	The Health Care Provider Role In IMQOL Application	126
7.4	The Utility Of This Application For Clinicians	130
7.5	Interpreting Coefficients.....	132
8	The Feasibility of IMQOL Elicitation and Application in Practice.....	137
8.1	Study Method.....	138
8.2	Results	139
8.2.1	Cohort Characterized.....	139
8.2.2	Data Collected	142
8.2.3	Withdrawn Participants.....	149
8.2.4	Analysis	151
8.3	Notable Case Studies	153
8.3.1	Absence of difference between states	153
8.3.2	Extreme Scaling Coefficients #1	154
8.3.3	Extreme Scaling Coefficients #2	158
8.3.4	Empirical Requirement For Continuous Attribute Scales?.....	159
8.3.5	Self Discovery in Elicitation	161
8.4	Discussion	162
9	Serendipity	167
9.1	States are not trivial to specify.....	167

TABLE OF CONTENTS (continued)

9.2	Death States Assessed In Two Frames Of Mind	170
9.3	Operational Versus Knee-Jerk Responses	172
9.4	Values Clarification	173
10	Next Steps	177
10.1	SOLOMON Program.....	178
10.1.1	Implementation of IMQOL protocol	178
10.2	Evaluate Utility for Clinicians.....	183
10.3	Scope of Utility	184
10.4	Evaluate Psychometrics.....	184
10.4.1	Reliability – Repeatability Over Time.....	185
10.4.2	Construct Validity	186
10.5	Extensions to the Representation.....	188
10.5.1	Non-linear Multiattribute & Hierarchical Models.....	189
10.5.2	Elaborating the Recomposing Model With Relational Theory.....	189
10.5.3	Conjoint Analysis	190
10.5.4	Lens Model.....	191
11	Conclusion	195
	References	203
	Appendix	213
	Satisfaction With Decision Questionnaire.....	213
	Mini-Mental State Examination.....	215
	Standardized Health State Descriptions Used in the Feasibility Study - 1 of 4	217
	Standardized Health State Descriptions Used in the Feasibility Study - 2 of 4	218
	Standardized Health State Descriptions Used in the Feasibility Study - 3 of 4	219
	Standardized Health State Descriptions Used in the Feasibility Study - 4 of 4	220
	IMQOL Models of Dialysis Patients	221

TABLE OF CONTENTS (continued)

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
 <i>Chapter 2</i>	
<i>Figure 2.1 Decision tree for a tree surgeon with an otic neuroma.....</i>	<i>29</i>
 <i>Chapter 4</i>	
<i>Figure 4.1 Results of standard gamble utility assessment for five outcomes of stroke arranged by severity of outcome according to biomedical perspective.</i>	<i>61</i>
<i>Figure 4.2 Results of standard gamble utility assessment for five outcomes of stroke arranged by patient ranked order of outcome according to biomedical perspective.</i>	<i>62</i>
<i>Figure 4.3 Mean (cross-hair), range (whisker), and 95% confidence interval (box) of the elicited utility values for mild disability, moderate disability, severe disability, death within two days and death within six months following a stroke. Triplets represent (from left to right) the "strictly-concordant" responses (S), "at-least-weakly-concordant" responses (W) and "nearly-concordant" responses (N).....</i>	<i>63</i>
<i>Figure 4.4 Effects of loosening the boundary for discordance in Rankin 0-1 Disability Assessments of Quality of Life.</i>	<i>68</i>
<i>Figure 4.5 Effects of loosening the boundary for discordance in Rankin 2-3 Disability Assessments of Quality of Life.</i>	<i>69</i>
<i>Figure 4.6 Effects of loosening the boundary for discordance in Rankin 4-5 Disability Assessments of Quality of Life.</i>	<i>70</i>
<i>Figure 4.7 Effects of loosening the boundary for discordance in Death within 2 days Assessments of Quality of Life.</i>	<i>71</i>
<i>Figure 4.8 Effects of loosening the boundary for discordance in Death within Six Months Assessments of Quality of Life.</i>	<i>72</i>
<i>Figure 4.9 Results of an interview showing how multiple dimensions explain results. "na" indicates the individual felt there was no relationship between the state of health in that column and the attribute of that row. In that case, the score he assessed his current health was used to compute the overall score.....</i>	<i>77</i>
 <i>Chapter 5</i>	
<i>Figure 5.1 Categories of constructs given in interviews.....</i>	<i>83</i>
<i>Figure 5.2 Categories of "other" constructs given in interviews.....</i>	<i>84</i>

LIST OF FIGURES (continued)

<i>Number</i>	<i>Page</i>
 <i>Chapter 6</i>	
<i>Figure 6.1. Overview of decomposition and re-composition of patient values. The individual's perspective of healthiness is composed of one or more attribute scales, each scale having a number of possible levels. Each level has a corresponding value on a numeric scale from 0 to 1. Knowing the level for a given health state for each attribute scale and the model for recombining the attribute scale scores in a composite score allows the use of the individual's perspective to score that health state.</i>	100
<i>Figure 6.2. Ordered attribute scales of one individual interviewed. All levels of each scale are listed with their corresponding value relative to the scale extremes to the right. The four attribute scales for this patient are arbitrarily numbered on the left. The level which this individual claims best describes his current state of health is marked with two asterisks, "**"</i>	108
<i>Figure 6.3 A Short Sited Rating Scale Method for Relative Weights Between Attributes of Health.....</i>	110
<i>Figure 6.4 A -Fuel gauge when car is rented. B -Fuel gauge after 100 miles.</i>	111
 <i>Chapter 7</i>	
<i>Figure 7.1 An example for accomplishing the second step of application: the rating of Rankin Scale level 3 with an elicited attribute scale involving mental well being versus mental decay. A specified state of health is rated using only the terms of the elicited attribute scale unless the state of health is deemed to bear no relevance or not to change anything concerning the attribute.....</i>	122
<i>Figure 7.2. Blank Scoring Grid for Five Health States Using an Elicited IMQOL model.</i>	123
<i>Figure 7.3 Scorecard with all health states rated before substituting numeric values associated with attribute levels.....</i>	124
<i>Figure 7.4 Scorecard from previous figure filled in with assertions of the interview process. Italicized numbers represent a rating of "not applicable" and hence a substitution of the subject's self assessment on the attribute scale.....</i>	125
 <i>Chapter 8</i>	
<i>Figure 8.1 Dimensionality of Elicited Models.....</i>	144
<i>Figure 8.2 Distribution of Attribute Scale Complexity.....</i>	144
<i>Figure 8.3 Average Complexity of Attribute Scales for Each Participant - Sorted.....</i>	145
<i>Figure 8.4 Mean (cross-hair), range (whisker), and 95% confidence interval (box) of the elicited utility values for hemodialysis(HD), hemodialysis with the complication of a clotted access (HD/c), continuous ambulatory peritoneal dialysis (PD) and continuous ambulatory peritoneal dialysis with complication of peritonitis every 9-18 months (PD/c). Triplets represent (from left to right) results from the IMQOL, time tradeoff (TTO) and standard gamble (SG) methods of assessment.....</i>	146

LIST OF FIGURES (continued)

<i>Number</i>	<i>Page</i>
 <i>Chapter 10</i>	
<i>Figure 10.1</i> Screen shots captured from Gambler Examiner computer program. The top window demonstrates the capacity to program analog rating scale methodology. Bottom window demonstrates the capacity to program the standard gamble methodology of utility assessment. Normally the program is rendered in color. Pills in the bottle are an assortment of red and blue in the same proportion as the outcome probabilities, Well and Dead, respectively, for Alternative B.	180
<i>Figure 10.2</i> Screen shots captured from U-Titer computer program. The top window demonstrates the capacity to program the standard gamble methodology of utility assessment. Bottom window demonstrates the capacity to program the time tradeoff methodology of utility assessment.	181
<i>Figure 10.3</i> Screen shots captured from IMPACT computer program. The top window demonstrates the capacity to program the analog rating scale methodology. Bottom window demonstrates the capacity to program the standard gamble methodology of utility assessment. This program is run on the Internet with any Web browser.	182
<i>Figure 10.4</i> The lens model. A linear regression of actual outcomes is calculated on the left side of the lens and a linear model using the judge's assertion of factor weights is calculated on the right side of the lens. X_1, \dots, X_n represent the variables (cues) of a multivariate judgement; r_{1s}, \dots, r_{ns} are the correlations between the individual variables and the judgement made; r_{1e}, \dots, r_{ne} are the correlations between the individual variables and the actual outcomes; and r_a is the correlation between the judgement and the actual outcomes.	191

LIST OF FIGURES (continued)

Number

Page

LIST OF TABLES

<i>Number</i>	<i>Page</i>
 <i>Chapter 2</i>	
<i>Table 1.1 “A Dialysis Modality Decision Guide Based on the Experience of Six Dialysis Centers.” Source: Adapted for Baxter Healthcare Corporation from: Hamburger, R.J., et al[Hamburger, 1990 #280].</i>	35
 <i>Chapter 4</i>	
<i>Table 4.1 Rankin Scale of Disabilities</i>	56
 <i>Chapter 8</i>	
<i>Table 1.2 Results of Standard Gamble Utility Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c).</i>	145
<i>Table 1.3 Results of Time Tradeoff Utility Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c). Results are expressed in the least number of years of perfect health (numerator) considered equivalent to an expected life span (denominator) in the state of health at the head of the column. For example, participant #01 considered 10 years of perfect health to be equivalent to 20 years of uncomplicated hemodialysis.</i>	145
<i>Table 1.4 Results of IMQOL Method Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c).</i>	146
<i>Table 8.4 IMQOL Model for Participant #01 (Self-Assessment in Bold).</i>	148

Chapter 1

Introduction

Medical decision making does not have a convenient language of discourse. Medical decisions are often complex. Patients are understandably anxious about the risks and benefits of both their condition and its possible treatments. They have concerns but do not know if it is appropriate to express them. They have values but do not necessarily know how to articulate them. Physicians currently lack a reproducible and rigorous technique to describe patient values with accuracy and precision, and to integrate what they have learned about those values through ad hoc methods with biomedical information about the patient. Consequently, readily measurable outcomes, such as mortality and hospitalization rates remain the almost exclusive basis for decision making. For lack of a convenient language, there may be no discourse at all even though traditional decision making by the physician as a proxy for the patient is now often rejected as poor practice. Today's clinical decisions need the patient's preferences. To include those preferences, patient and physician must share a common language in which health states and risks can be described, discussed and evaluated. The professional language of medical diagnoses, treatments, outcomes and tradeoffs is not ordinarily familiar to patients, and even if translated into lay language may not adequately convey the real options and tradeoffs the patient faces or believes to be relevant. Conversely, patients may try to express their views in

terms of their own life experiences and familiar concepts, but physicians may be impatient with these and may fail to understand how to relate them to the clinical concepts in terms of which they analyze the decision. Until a formal methodology for defining the individual's value system is clearly established health care providers will continue to be forced to rely on their intuition about patients' values. The goal of this research is to develop, implement and test a method for easing this problem with a language for discourse and an elicitation process that promotes discourse.

1.1 Motivation

The health care community is composed mainly of highly trained individuals who have chosen for their lifetime career to relieve suffering and illness caused by disease. Pledging to uphold the Hippocratic oath, they are held accountable "to do no harm." It could be said that no other profession asserts so boldly the ambition to pursue the individual's welfare above all else. Technological advances leap forward because of the devotion of massive resources both economic and personal. The most promising industry in advanced industrial nations is that of medicine. Consequently, there exists an exponentially growing list of benefits to be had from the practice of medicine and the impact on length and quality of life. Most recipients of this service are alert, cognizant, intelligent, articulate humans. The practice of medicine is dependent upon the role of the patient in articulating signs and symptoms to optimize diagnosis and outcomes. When it comes to preferences and values, no one lives with them more than the patient does. In my experience, the only topic people are more willing to talk about than health is his or her pet animal. Although it is uncertain whether all patients desire to participate in all aspects of medical decision making¹⁻³, there are plenty of willing participants and more who are willing to express opinions when the outcome is in their disfavor. Many patients have the capacity for very articulate contributions. More desire to be at least cooperative, recognizing it is in their best interest. Willing participants exist and resources are poised for the exchange of preferences. Health care providers often want to know the preference of the patient as desperately as the patient wants to know cure.

If a convenient language of discourse did exist, it would be composed of terms commonly used and clearly understood by the decision-maker. To fulfill a role of support in collaborative decision making, the terms would be able to distinguish treatment strategies and their outcomes. The relationships between the terms would have an operational quality. The systematic approach to discovering such a discourse language would be repeatable and not limited in who might employ it to communicate or distinguish the issues. The success of this language, and hence the discourse, would depend upon how complete and comprehensive this language was.

1.2 Utility Assessment to Date

Decision analysis is an elegant technology for making the issues of medical decisions explicit where tradeoffs are involved. It is difficult, however, to apply this technology to individual patients' decisions when the value for each outcome is unclear. This is most often the case because of the diversity, uncertainty and ambiguity of patients' values. It is made worse by a provider's use of unfamiliar terms or a poor understanding of the individual patient's way of viewing the specific medical decision. Consequently, despite a two-decade history of pioneering application to medical problems, decision analysis is absent in common medical practice. For eliciting a language of convenience and promoting discourse, this paper describes a novel approach to capturing patient utility values using attributes for healthiness that are elicited directly from the individual patient.

Von Neumann- and Morganstern-based utility theory ⁴ provides a coherent basis for value elicitation methods. These methods, along with maturing health status measures, have a growing potential impact on health care policy and resource management from a social and corporate perspective. One glaring absence is the means to represent the individual's perspective if it differs significantly from the empirical norm of studied or consulted populations. Traditionally, health status measures have reflected the value system of the biomedical community (e.g., life span was used long before quality of life adjustments were included). Recent recognition of this biased view of life has given rise to the question of what dimensions should be included in health utility assessment. Most recently this has been

addressed through the efforts to elicit the most meaningful parameters from experts and patients gathered in “focus groups” intending to develop a patient sensitive model for what dimensions to include. However, as a matter of practicality, dimensions shared by very few persons are neglected in this approach. This does nothing to accommodate those individuals for which these dimensions serve an unusual importance. It serves the purpose of society but not the individual.

Fundamentally, the practice of value assessment is capable of individual application. The commonly recognized methods (rating scale, standard gamble and time trade-off) have been applied to a lengthy list of medical domains, one patient at a time. In the absence of a satisfying “gold standard”, ambivalence about the results exists on the part of the health care provider. This ambivalence, as well as patient violation of prescriptive conclusions, has prompted modified methods. Repeatedly in this evolutionary process, the evidence describes diversity among individuals. Responses often contain internal inconsistency, which so far eludes a coherent explanation outside the assertion that people are often irrational. The complexity of instruments and results underscore the demand for trained experts to conduct value assessment. The skill required to administer a quality elicitation process is exceeded by the skill required to interpret the answers given. Further expertise is required to keep up with the growing list of questions raised about the methodology and compensatory modifications. Not only must the analyst be highly trained, they must also slow down and take time to actively listen to the often inarticulate person from whom the information must be acquired. Rushing this assessment confuses and confounds.

1.3 The Shortfall

For current methodologies to acquire meaningful information, value assessment must force the person responding to formulate answers to grave potential outcomes for their life. When, due to constraints, value assessment and decision modeling are reserved for only the most serious applications the patient is more likely be asked to repetitiously face their mortality resulting in severe prolonged emotional stress completing the elicitation. Not only is this

emotionally painful but it is embarrassing for both the patient and the analyst, further discouraging the use of such technology except in extreme cases.

Therefore, applied on an individual basis, value assessment is expensive. Consequently, the skills of highly trained experts are not conveniently available. As a case in point, a member of the staff of the Clinical Decision Making Division of the New England Medical Center, famous for implementing decision analytic technology, can only attest to performing utility assessment on an individual patient a few dozen times in half as many years^{*}. When performed, the process takes two hours and usually leaves the patient in tears. The assessment must be performed during regular hours of the professional day most likely in the sterile environment of a referral health care institution. The result of the individualized utility assessment, even when performed with multiple methods and cross validated, is no more than a gross proxy measure. For all the effort, the value for each health-state evaluated is a single overall summary measure for everything involved in that specific health state. Should the context of the decision change, i.e., the potential outcomes shift in their description or context, the procedure must be repeated from scratch. Should the patient change their outlook on life - a goal for most of medical therapy and provider-patient interactions - the results of such former utility assessments are equally obsolete. No parts of the all-encompassing proxy measure are reusable.

There is a great need for a systematic approach to the discovery of what is and what is not important to the individual and an operational understanding of the patients' perspectives if their preferences are to be used in medical decisions. Empirical evidence exists for the inadequacy of the physicians ability to assess patients' preferences⁵. Of more advantage would be a representation of those values that is generic enough to allow use in many scenarios rather than the non-reusable values acquired with current approaches. The protocol developed in this thesis will accommodate the individual, with a systematic approach. The result is a representation of the individual's values that is multidimensional and reusable.

^{*} Personal Communication, Mark H Eckman, MD, Boston, MA, 1996.

1.4 Aims of This Research

My aim is to explore the feasibility of using the terminology of the individual patient to capture and express the individual's particular, perhaps unique, value system for healthiness, to establish a convenient language for discourse. The research question is whether it is possible to use individualized patient terminology to represent values for the outcomes in medical decision models constructed by the medical community such that the utility of each outcome may be determined from the patient preference model. Furthermore, I propose to address whether it is possible to assess values in the terms and concepts of the patient such that these assessed values facilitate health care providers as they guide patients in medical decision making.

In order to integrate such values in treatment decision making, it is necessary to devise a rigorous, reproducible method for determining those values in a way that others can use them. A substantial body of decision analytic technology exists, but is not frequently used. Inadequacies of current assessment methods upon which decision analysis depends are major contributors to the infrequent use. These methods are expensive, laborious, demanding, and embarrassing tasks with results that remain dubious. This study will attempt to implement a rigorous method of determining the values of dialysis patients such that physicians may use their patient's value system in scoring health states and comparing the outcome of treatment strategies. The intent is to devise a protocol that would enable automating as much of the interview as possible. Conducting the interview with a computer program would enable values assessment in more comfortable and private settings than hospitals or physician's offices in the presence of analysts.

Should we provide a means to grasp the values of a patient in proficient manner, we enable the health care community to better understand and accommodate the uniqueness of individuals. If this understanding enables the health care provider to make choices clearer to the patient in his or her own terms, that patient is better prepared to participate in resolving any decision involving tradeoffs and uncertainty. Should the patient be rendered unable to represent his/herself, the externalized representation of his or her values may still be adhered to. Any rigorous and reproducible approach to assessing the values of the patient for described outcomes reduces the uncertainty of the overall decision.

1.5 Overview of Thesis

The fundamental thesis of this research is that it is feasible to use the patient as an expert on the values he or she uses in evaluating quality of life. I propose a representation of values consisting of multiple dimensions defined in the patient's own terms that can be applied to the assessment of quality of life for arbitrary states of health. I will refer to this representation as the Individualized Multidimensional Quality Of Life (IMQOL) model and a protocol for eliciting all the necessary components. The IMQOL protocol is promising because it is based upon the fundamental assumption that traditional frameworks for utility assessment can be applied to patient-elicited attributes of healthiness as well as it can be applied to physician-described clinical health states. Furthermore, I hypothesize that both the health care provider and the patient can associate descriptors from the patient's unique IMQOL model to clinical outcomes enabling the collaborative evaluation of quality of life. Studies using this paradigm for evaluating states of health will be described revealing the feasibility of applying such models in a way consistent with traditional utility assessment.

After motivating the thesis with three examples of medical decisions illustrating different aspects that warrant individualized decision support, a foundation of previous work in the area of quality of life assessment will be laid. This is a natural introduction because the IMQOL model and elicitation protocol is a reconstruction of quality of life assessment from components used already. A taxonomy of discordant responses will be described prior to the IMQOL model as further motivation based on interesting findings uncovered by pilot studies in pursuit of the IMQOL model. The representation itself will then be described in detail and justified before explaining the elicitation process. Illustrations based on the examples given up front will be used as the IMQOL protocol is rolled out. Values accomplish little without application and so the application of the IMQOL model for assessing the quality of life will be specified.

To evaluate the feasibility of the protocol in practice, a group of dialysis patients were interviewed with the IMQOL protocol and asked to assess the quality of life in four specified states of health common to dialysis therapy and their own health. The results of this first step toward establishing the feasibility of the protocol are presented. The character of the sample

population, collective assessment results, and a few difficulties before the review of three notable case studies are reviewed. An encouraging participant response is shared and indications of operational quality discussed. A chapter ensues containing serendipitous lessons learned followed by a detailed treatise of the next steps to be taken in programming and evaluating the protocol. The thesis ends with the conclusion that it is feasible to elicit patient values in the patient's own terms and that doing so can be a process of discovery in which the health care providers should play a role. You will see that this protocol shows promise of augmenting traditional quality of life assessment and individualized decision support.

Chapter 2

Motivating Examples

The need to understand patient preferences can be illustrated in three contexts. Not all tradeoffs are alike. These three examples will characterize different contexts in which patient preferences contribute to resolve questions regarding optimal treatment strategies. First, the patient may have peculiar concerns, needs or issues that make him or her atypical. Judgements based on the majority of people will be inappropriate for such persons. Second, extreme tradeoffs may be so disparate that it is simply difficult to anticipate what the majority of persons would want. Finally, the tradeoffs between treatment strategies may be so complex that though the choice is simple; the basis for choosing is not so simple. More often, both the choice and the basis for choosing are complex and that complexity is made worse by representation in unfamiliar terms.

2.1 Tree Surgeon Facing Otic Neuroma Decision – Individualized Issues

For the first example, consider the case of a 45-year-old male who develops an otic neuroma and faces the decision between surgical treatment and either medical treatment or no treatment at all for this slow developing growth on his otic nerve. An otic neuroma is a benign growth on the nerve that communicates between the ear and the brain – the hearing circuit. As a benign tumor, this growth would be no more consequential than a wart if it were not for the location. Between the inner ear and the brain, there is a rigid skull without much room for things that do not belong. Small and slow growing an otic neuroma will not be evident until the space occupied by it compromises the space other normal structures count on. For example, the otic nerve itself would get along just fine with the benign growth attached were it not for the

pressure on the nerve created by the tumor inside the confined space both the nerve and tumor must occupy. The slow growth of the tumor even allows the cells of the nerve to adapt somewhat to the pressure. Were the tumor's growth more rapid, the sudden change in pressure would be more consequential to nerve function and therefore to hearing. With slow growth, the nerve tissue is able to accommodate the subtle changes in the environment and postpone the arrival of significantly detrimental effects. By the same token, progress of symptoms will be drawn out over time, usually years. Consequently these types of tumors are not usually found early and their discovery does not represent an emergency. One may rationally consider a decision of no treatment for this tumor or delay treatment until intervention is more warranted.

For the purposes of illustration, let us assume that there is one medical alternative to surgery. Let's assume there is a single chemotherapy protocol that could be used to treat this tumor. For simplicity sake, let's say the treatment has no risks or long term side effects other than it shortens the life span of the patient. Without belaboring the means of determination, let us assume that all parties involved agree that the value of the outcome of chemotherapy is worth 80 on a scale from 0 (equivalent to death) and 100 (equivalent to perfect health for the patient's normal life span). This assertion takes into account the temporary set back of chemotherapy and the foreshortened life span. There are no other side effects.

The neurosurgeons suggest that the surgery available consists of removing the tumor surgically by curettage or, in effect, scraping it off the otic nerve. They further reveal that there are three possible outcomes for surgery, assuming anesthetic risk is non-existent. The ideal outcome is no side effects with hearing and balance preserved. If the surgery traumatizes the otic nerve unavoidably, it might effect hearing. In addition, the nerve that runs from the vestibular organ, i.e., the gyroscope necessary for a human to maintain his or her balance, to the brain is immediately next to the otic nerve. Although the tumor threatens loss of hearing more than loss of balance, there is also a risk of traumatic injury to the vestibular nerve in curettage procedure. Let us assume that the surgeons have told the patient that the probability distribution for these outcomes are a 0.2 probability of no side effects, a 0.4 probability of hearing loss and a 0.4 probability of both hearing loss and vertigo following surgery.

If the decision faced by this patient is modeled as a choice between chemotherapy and surgery, the decision can be expressed in decision analytic terms as found in Figure 2.1. Again, without detailing how it is measured, let us assume that the value of each of the surgical outcomes are 85 for no side effects, 85 for hearing loss in one ear, and 50 for loss of balance and hearing in one ear on a scale from 0 to 100 for this person. Analysis of this decision tree reveals that, in terms of the value of the outcome with hearing loss, the threshold for the change from surgery to chemotherapy as the optimal choice lies just below 50—i.e., if the value is any lower than 50, the optimal choice is medical therapy rather than surgical.

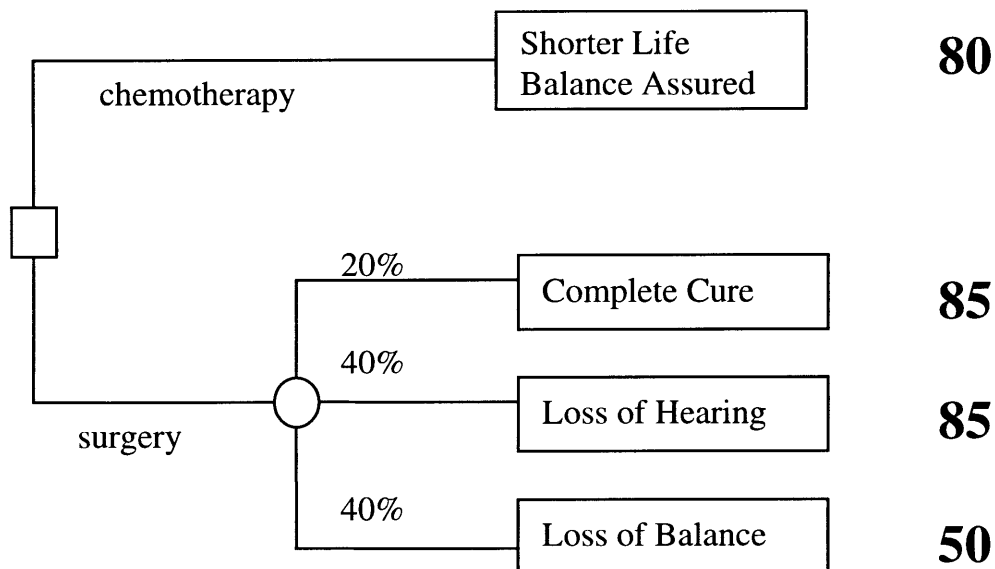


Figure 2.1 Decision tree for a tree surgeon with an otic neuroma.

Now consider the case where this patient is a tree surgeon whose livelihood depends upon climbing into trees and cutting off branches while delicately perched on other limbs in awkward body positions. Even when with extensive use of safety ropes, this person is dependent upon the sense of balance to maintain a personal sense of confidence necessary to perform many occupational tasks. It is easy to imagine that this individual is very likely to have less appreciation for life with vertigo than most other individuals. Peculiar circumstances for this individual add up to an entirely different recommendation based on the assumptions of the model and rational behavior.

It is difficult to capture the variety of unique considerations that make up the diversity of humanity through focus groups. Their aim is to discover the relevant issues for target populations, not individuals. While this approach to the discovery of salient issues is unsurpassable in many ways, it is impractical for capturing particular concerns of individuals. Nonetheless, when it comes to the decisions of those individuals, the impact of those concerns can be significant. This illustration demonstrates the significant impact of an individual's peculiar concerns on medical decisions motivating the demand for individualized assessment of outcome values.

2.2 Acute Stroke Treatment And Preference Between Disability And Death – Extreme Tradeoffs

Recent research investigating the use of thrombolytic or “clot busting” drugs for stroke suggests another characteristic example of a decision needing the preferences of individual patients. This case is one of extreme outcomes of very different nature. It is the disparity of the nature of the outcomes that makes it difficult to reason about which is optimal. This is the epitome of tradeoffs.

Stroke is the occasion of something blocking the flow of blood to some region of the central nervous system. This is usually in the form of a blood clot that blocks the flow of blood. There is some suggestion that administration of anti-clotting drugs to the patient in the first few hours following the onset of signs will result in a decreased chance of disability measured months after the stroke. There are multiple studies investigating multiple drugs. Only a few studies suggest any significant change in the frequency or severity of disability, but many suggest that an increased risk of death in the first few days is associated with the use of such drugs. Apart from the details of the issue, the fundamental question regarding patient preferences is that of disability versus sudden death.

There is no other known side effect of using some of these drugs, e.g., tissue plasminogen activator. The drug causes no detectable sensation when administered. It can be injected into the veins using the routine intravascular access that in all probability is already in place in any such patient, so it does not even come with a needle prick. The drug is produced

by the body naturally and so readily metabolized without consequence. Thus, it turns out that the decision involved is not complicated with “ifs” and “buts” but boils down to a straightforward issue of the patient’s preference for avoiding disability versus avoiding death. It turns out that some people would rather avoid severe disability than avoid death. Judging from a study I performed interviewing nearly 40 emergency department patients during their visit to a Boston hospital, most people interviewed think severe disability following stroke is a quality of life that is worse than death. This is, however, not a universally held view. There are those who consider sudden death worse than such disability. Others think death after several months of post-stroke disability is worse than severe disability for a normal life span. The question emergency department physicians would like to know the answer to is which group the patient they are working on is in. They have a patient who has only shown signs of stroke for a few hours and they have the opportunity to effect the chance of an outcome with severe disability. Should they give tissue plasminogen activator to this patient at the slightly increased risk of death? Or should they avoid that risk and increase the probability of a vegetative existence for the remainder of life? The choice is clear-cut but the meaning of death versus severe disability is more than just a proverbial “apples to oranges” comparison. This comparison pits grave outcome against grave outcome. It involves issues that most people want to avoid thinking about and so little is available to suggest what to expect of patient behavior when faced with a real choice. The basis on which the choice is made is personal and difficult to predict. What might contribute to a person’s preferences in such a context is an immense universe of possibilities. Although the options faced might be regarded as an uncomplicated tradeoff, the extremely grave context and complexity of the basis for decision may be formidable.

2.3 Hemodialysis Versus Peritoneal Dialysis - Complex Tradeoffs

Not all tradeoffs are as simple as the previous examples. Patients who suffer from end stage renal disease (ESRD) face a growing choice of therapeutic modalities. Hemodialysis therapy has more variability than most people suspect. The frequency and duration of dialysis treatments are varied by the clinician as needed to effect a life sustaining control of byproducts normally excreted by the kidney. The patient schedule is largely constrained by the dictates of

dialysis center logistics and schedules but there is still room for patients to express preferences. Increasing the dose of dialysis (frequency, duration, and rate of extracorporeal blood flow) will improve the values of the parameters used to monitor the patient's status, but it is not clear that the change in perceived health warrants the increase in time spent attached to the hemodialysis machine. Outside the weekly details of the schedule there is the issue of when in the life of the ESRD patient to initialize dialysis and sacrifice the recurrent time slots that therapy will consume. The result of dialysis is not an all or none phenomenon. It involves a trade off between the improved health that results from time spent attached to the machine and freedom from such an anchor.

The most dramatic and profound decision facing hemodialysis patients is how long to continue artificial organ therapy. Although chronic dialysis has become an accepted feature of the modern medical environment, at least in developed countries, it remains an extraordinary prolongation of life, and the prerogative of competent patients to withdraw from dialysis therapy is accepted. But as high as the general level of technical care may be, it is not at all clear that either patients or providers are particularly well equipped to deal with this issue. It represents an ultimate need for values clarification.

The dialysis faces many other choices in regard to diet and fluid ingestion. Smaller interdialytic weight gains are associated with lower treatment-related morbidity and with improved blood pressure control; improved blood pressure control may in turn improve long-term outcomes. Diminished phosphorus ingestion and attention to phosphorus binder dosing will reduce itching in the short term and the complications of hyperparathyroidism in the long term.

In patients with chronic renal disease, treatment with angiotensin converting enzyme inhibitors clearly retards progression of renal failure, and can defer the need for renal replacement therapy. However, as renal failure progresses, particularly in diabetic patients, treatment with these drugs is frequently complicated by acute renal insufficiency and by hyperkalemia, and can require more frequent blood testing. Co-administration of sodium polystyrene sulfonate can prevent hyperkalemia, but may require increased diuretic treatment because of the salt load, and certainly causes constipation unless given with sorbitol or other

laxative agents. Is the prolongation of time to renal failure great enough to justify this more complicated and messier treatment to an individual patient?

There is evidence, though not conclusive, that dietary protein restriction retards the progression of renal disease. In the context of this uncertainty, is the magnitude of the benefit great enough to justify the dietary changes and monitoring of nutritional status?

In patients approaching end-stage renal disease, early creation of vascular access can reduce morbidity and save resources. The current standard of care for nephrologists is to achieve permanent vascular access placement before dialysis is initiated. The earlier one starts working on this problem, the more likely it is that the patient will have a working native vein access at the time of renal failure. Logically, all diabetics with even early macroalbuminuria might be encouraged to protect their veins from venipuncture, and attempts at fistula creation might begin much earlier than is current practice. However, many clinicians would hesitate to confront patients with the prospect of renal failure this early.

The hemodialysis patient whose vascular access fails faces the choice between dialysis by a catheter, with the prospect of eventual infection, repeated attempts at fistula creation or graft placement, and conversion to peritoneal dialysis. This decision depends on considerations of risk, discomfort, inconvenience and body image. The hemodialysis patient whose vascular access shows signs of failing faces the choice whether to anticipate the problem, or to wait for the failure to occur.

Without explaining the details of the dosage units, it can be pointed out that there is a threshold of dialysis dosage that is considered the standard of care for ESRD patients. These recommendations are based on observations about mortality. They presuppose that the patient's goal is to maximize life expectancy. But abundant evidence from formal studies of decision making shows that maximizing life expectancy is not a universal or exclusive goal. Furthermore, the mathematical relationship between dialysis intensity and survival is incompletely understood. For the individual patient, therefore, the choice of dialysis treatment duration might be seen not as a simple question of compliance with doctor's orders, but as a complex tradeoff under conditions of uncertainty.

It is clear that hemodialysis is a complexity of tradeoffs, but it does not stop there for the patient. The option for peritoneal dialysis extends the options confounding the tradeoffs even more. Peritoneal dialysis involves placing a soft plastic tubing in the wall of the abdomen to allow dialysate to be alternatively drained into and out of the abdominal cavity. The fluid is not immediately drained to allow normally excreted byproducts to migrate from the body into the dialysate. This mode of treatment does not involve the use of needles and venipuncture. It can be done by the patient and does not require trips to the hospital. Special care must be taken to keep the end of the tube sterile so as to avoid getting an infection in the wall of the belly, but the patient can go anywhere if they can take the bags of dialysate along. Treatments must be performed daily rather than every other day as in hemodialysis. The belly is distended with fluid between fluid exchanges, which may be of concern to some patients concerned with appearance.

Many nephrologists would like to see more ESRD patients treated with peritoneal dialysis instead of hemodialysis. There is no conclusive evidence that patients do better on this mode of therapy as measured by recognized parameters. It is more a matter of opinion that more patients would be well suited to the alternative with potentially more cost effectiveness. Using focus groups and traditional utility assessment methodologies, the CHOICE group is investigating the question of which mode should be considered a better quality of life⁶.

Peritoneal dialysis can be performed in different ways. The patient can manually perform the fluid exchanges or a machine can be programmed to do it while they sleep. To maintain the standard of care for dialysis dosage, the patient must be connected to this machine a few hours more than most people sleep. Consequently, a cycle or two can be performed in the evening with other activities interposed before going to bed attached to the machine for the rest of the night. Each issue of dialysis treatment choice is compounded by the alternative of peritoneal dialysis. The dose of dialysis, start of therapy, protection of vascular access, nutritional supplements, use of gastrointestinal binders and cessation of therapy are all relevant to the peritoneal dialysis mode of therapy as they were to hemodialysis. So, the tradeoffs involved with each are complicated by the plurality of modes of therapy available.

Table 2.1 A Dialysis Modality Decision Guide Based on the Experience of Six Dialysis Centers. Source: Adapted for Baxter Healthcare Corporation from: Hambuger, R.J., et al.

Usually do better on hemodialysis (HD)	Do well on either therapy			Usually do better on peritoneal dialysis (PD)
	Tending toward HD	Tending toward PD (CAPD, APD)	Tending toward APD	
Severe inflammatory bowel disease				Unstable vascular disease
Active, acute diverticulitis/schematic bowel disease	Dependent lifestyle	Independent lifestyle	Pressure-related complications	Difficulty in vascular access
Marked intellectual disability with no helper	Dementia	Cardiovascular disease	Lifestyle--free days	Children under 5
Severe, active psychotic disorder	Chronic poor hygiene	Residual renal function	More prescription flexibility	Younger diabetics
Homeless	Multiple abdominal adhesions/ostomies	Transmissible disease	Social support billed by helper at home	Strong patient need for independence, autonomy or control
Unresolvable PD complications	Severe, recurrent hernias	Variable schedule		Distance from center
	Severe diabetic gastroparesis	Travel		Unresolvable HD Complications
	Frequent and substantial therapy changes			
<i>Center</i>	<i>Center/Home</i>			<i>Home</i>

The complexities of the tradeoffs that have been mentioned are mostly described here in biomedical terminology, as is most of the communication of these issues to the patient. If the patient does not think and reason in those terms, there will be more complexity added by the

translation from biomedical idioms to patient taxonomies. For example, what does “protein restriction” translate into for the patient’s actual diet? What does “angiotensin converting enzyme inhibitors” or “sodium polystyrene sulfonate” medication imply for the life style of the patient? Retarded progression of renal failure is similar to deteriorating ozone layers or global melting icecaps until it is translated into everyday terms that have meaning for each patient.

Where tradeoffs are simple, it is easier to reason about what is better for the patient. Complex tradeoffs render even the experts at a loss for reasoning about what is optimal from a strictly biomedical perspective. The result is health care providers who openly wish they reliably knew what the patient valued most in hopes that it would resolve the choice of treatment strategy. This is clearly the case in dialysis therapy. Table 2.1 shows five lists of issues in a chart adapted by Baxter to characterize the factors contributing to the choice of dialysis treatment between hemodialysis and peritoneal dialysis⁷. Ambiguity abounds in this guide making a clear choice difficult to find even for the healthcare professional. The complexity of the decision can only increase when non-biomedical issues of the patient are added into the decision. Breckenridge attempts to classify the factors that influence the decision regarding the type of dialysis treatment modality as physiologic, psychologic, sociocultural, developmental, and spiritual⁸. Groome, et al, asked which factors are important in the decision finding that peritonitis ranked most important, but life style considerations ranked higher than medical consequences of specific therapies. Nurses, doctors and patients were found to agree on some issues yet disagree on others⁹. The more recent CHOICE study, using focus group elicitation methods, similarly finds a different set of classifications and disparity between providers and patients quantitatively if not qualitatively⁶.

It is hard to imagine anyone being certain that what is good for the current patient is equally applicable to the next. This is just one example of a specific decision in medicine representative of a population of decisions involving complex tradeoffs. It would be more appropriate to develop individual preference models for patients applied to specific disease contexts than to develop ambiguous disease decision models for universal application to all patients. Disproportionate estimation of value can lead to erroneous recommendations as well as qualitative misunderstanding and omission. Without the explanatory contribution of a

multidimensional view, it can be difficult to make sense of alternate viewpoints. The more factors involved in a decision, the more opportunity for disproportionate estimations of value. Reduction of this dimensionality for quality of life measurement into unidimensional proxy measures would seem perilous and indeed our unidimensional measures leave analysts wondering about the results.

Chapter 3

Background and Significance

In this chapter, I give the background, significance, and rationale for this thesis. There are multiple disciplines—each with many motivating elements—which have an interest in the elicitation and use of patient preferences. The diversity of interests make it difficult to devise a single coherent thread which ties all of them together in a neat package, easy to read and digest. Consequently, this chapter is organized as a series of answers to questions that arise when considering the vast diversity of interest that may bear on the domain of the thesis.

3.1 Why patient values?

Most people consider it the patient's right to have his or her value judgments included in the medical decision process. O'Conner and O'Brien-Pallas maintain that an effective decision is informed, consistent with decision-maker's values and behaviorally implemented¹⁰.

Tradeoffs between treatment options often make it unclear which choice is best. Patient preferences can be valuable for resolving the selection of a treatment strategy in such cases. Most patients do want to participate in treatment choice decisions and outcome evaluation^{1, 2}. Patients may exercise their values—even at increased expense or detriment to health—by withdrawal from therapy or exchange of providers when they perceive the lack of adequate consideration. “Harm” in the Hippocratic oath might be interpreted to include the omission of patient values consideration. Fundamentally, most medical schools teach physicians that they should use patient terminology at the bedside for the best patient care. More importantly, psychological research has shown that when the patient's own terminology is employed, rather than the therapists' terminology, the potency of the sentiments expressed is stronger and the therapeutic results are more consequential¹¹⁻¹³.

3.2 Physicians don't make decisions based on patient values?

Despite growing interest in outcomes measures and quality of life assessment¹⁴ the inclusion of patient judgment in treatment decisions is limited^{15, 14}. Studies at the New England Medical Center dialysis unit indicate that the Medical Outcomes Study SF-36 questionnaire¹⁶ frequently uncovers pertinent information not otherwise known or recorded in the medical record¹⁷. Health status measures are largely biomedical attributes and, more recently, functional capabilities. There is a distinction to be made between these and the patient's values (“utility values” in the terminology of decision analysts). Even where patient “experience” has been measured by psychometrics, we still lack the patient's description needed to get a true representation and clear understanding of his or her values. Attention to the individual patient is not a novel concept nor does it escape the efforts of health care providers^{18, 19}. What is missing is a feasible, systematic way to discover and use patient values in his or her own words particularly where the optimal treatment strategy is not clear. Thus far either the concepts used to evaluate health are those of the medical community (i.e., biomedical perspectives or functional status) or the terminology of someone other than the patient. While the patient may

be represented in some form (e.g., utility value for an outcome in a decision model), the value system by which the individual patient ascribes meaning to outcomes is not explicitly represented²⁰. The nephrologist who wishes to involve the patient's value perspective in dialysis treatment planning is left to his or her own adhoc devices.

Efforts to empirically validate the necessity of patient input in determining the issues in medical decisions have had mixed results. Groome, et al²¹, found no significant difference between the frequency and composition of items deemed relevant by health care professionals and patients to the decisions regarding end stage renal disease (ESRD) treatment by hemodialysis or continuous ambulatory peritoneal dialysis. This study compared the frequency between these two groups. It also compared the frequencies with the relative importance indicated by direct measures. The results failed to show additional contributions by the patients to the issues identified by professionals. The professionals did mention some items with more frequency than patients. However, simply comparing what professionals identified more than once with items patients identified more than once, the study found no significant difference. The CHOICE study by Bass, et al⁶ differed in that each group was left uninformed of any issues it had omitted that the other had raised. They found that "although health professionals independently identified most of the quality of life issues that were important to patients, they seemed to assign less importance to some of the quality of life issues than did the patients." Further, in addition to domains included in generic instruments that have been used to assess quality of life in ESRD patients, the focus groups with patients identified a few additional domains that "have received relatively little attention."

3.3 What kept others before me from acquiring and using patient specific values in their own words?

Nothing but time constraints keeps physicians from learning and incorporating patient specific values in familiar terms by ad hoc devices. However, physicians currently lack a reproducible and rigorous technique to describe patient values with accuracy and precision.

Even if they did, there is no reproducible and rigorous technique to integrate what they have learned about those values with biomedical information about the patient. Consequently, other, better-defined outcomes, such as mortality and hospitalization rates remain the almost exclusive basis for decision-making. Until a means for detecting a formal methodology for defining the individual's value system is clearly established, health care providers will continue to be forced to rely on their intuition about patients' values.

3.4 What prior accomplishments do exist in this area?

What prior accomplishments do exist in this area? A twenty-year history of attempts to incorporate personal preferences in medical decisions^{22, 23} has resulted in a significant list of health status measures^{24, 25}, established decision analysis methodologies, and utility assessment tools developed to support decision analysis. Despite this heritage, it is still not common practice to use this existing technology, or to systematically represent individual patient preferences in evaluation of treatment options. Medical decision analysts do apply what is referred to as "utility assessment" tools to directly determine a numeric value to represent the patient's overall value of a specific decision's potential outcomes. However, these tools are limited by being labor intensive, time consuming and constrained by the availability of trained analysts. They are also conceptually difficult for the patient, leaving uncertainty as to the patient's understanding of the instrument and even the meaning of his or her own answers²⁶. When applied to individual patient cases, decision analytic models use the individual's "utility value" to score each potential outcome of a medical decision, e.g., the choice between continued immunosuppression therapy or not in a transplant patient with superficial spreading malignant melanomas²⁷. The current forms of utility assessment are constructed from disease and decision specific contexts. As such they do not explicitly distinguish any value systems of the patient, rather they directly determine a numeric value, called the "utility value", for the specific health state under consideration (one notable exception is the work of Pierce²⁸). In health care policy and protocol the concerns of the patient are in fact represented; however, the

mechanisms exercised to elucidate those concerns to date are limited to statistics on survey results and focus groups for determination of the issues that matter to patients²⁹⁻³¹. The result is attention to concerns that are shared among some threshold quantity of representatives, but the impact of all solitary individual concerns is still beyond the capacity of such a system.

3.5 What is missing?

We need a systematic method to determine individualized patient values and express those values in that patient's terms. This mechanism must be complete, operational, and comprehensive. To be complete, the elicitation must supply all issues of concern to the individual. Let us call these "attributes" of healthiness in the perspective of the individual patient. Each attribute must be measurable in order to be useful or contributing to the discriminating power of the value system. The attribute then could be called a scale with some order to its elements correlating to the scale from healthy to unhealthy in some way. The scope of each attribute scale must be comprehensive, i.e., cover every possible level between the highest known extreme and the lowest known extreme. In other words, if we know all the measurable dimensions by which an individual values health and have a working scale that spans the spectrum of each dimension between its extremes, we will have a pretty good handle on the way that individual values health.

3.6 What stopped others?

The exhaustive nature of the completeness requisite is what threatens to make individualized values impractical for routine health care. Cost containment and capitation are driving the contact time with the patient down and the elicitation of patient values can only demand more time consumption if acquired from friendly conversation or personalized face-to-face interview. Survey instruments alleviate the necessity of staff to collect data, more so if the process can be automated. Printed surveys, however, are rigid and disallow adaptive questions

specific to the patient or circumstance. Inadequacies of utility assessment have already been discussed. Studies indicate that current utility assessment results contain a wide variation between individuals and can be poorly correlated with symptom severity³². This makes their usefulness more difficult to understand.

3.7 How is what this thesis offers different?

The values elicitation presented here anticipates the use of computer programs to emulate the natural way humans might approach the task. To learn about a person's interests, one might start with conversation, learning what vocabulary this person uses. A good listener would hear out the individual without putting words in the speaker's mouth. Next the good listener might make an effort to use the words of the other individual in novel combination to test the understanding. This would likely reward both parties with the fruit of better understanding. Conscientious physicians would certainly do this to gain a better understanding of their patients if they had the time. A computer program that emulates this manner of communication could be expected to facilitate understanding in similar fashion.

My proposal for values determination then is to engage the individual in a structured interview in which he or she reveal the vocabulary and dimensions they use in discriminating healthy from unhealthy friends and family. Attention is given to deciphering the scales used to evaluate each dimension in his or her own terms. An understanding of those scales is acquired individually and patient approval sought on the functionality of that understanding. Numerical values are derived to be associated with each level of these scales. The means by which the individual patient combines these scales is then determined so that health care providers may employ the same mechanism of evaluating health states.

The model by which the patient composes the individual attributes into an overall judgment of healthiness is a critical step in value system determination. Psychometric research has a long history of using linear models. If we follow suit, we need nothing more than to determine the scaling coefficients for each attribute in the equation. The opportunity exists to

explore the consideration of utility independence and preference independence at this juncture³³, but the value of that complexity is not yet clear. We do have a feasible starting point in the linear model since the scaling coefficients can be determined by using the descriptors authored by the patient to describe hypothetical yet "familiar" extremes in combination for standard utility assessment methods. Still no specific health state is required but those generated by the patient's attribute scales. The formalism I present is systematic enough to suggest successful implementation with computer aided self administration. This raises the possibility of more familiar and leisurely environments for administration than disconcerting physician offices and hospitals. It also allows for repetition that fosters familiarity with the instrument. All of this increases reliability, judging by more traditional applications of utility assessment. If the individualized patient values elicitation involves significant investment of time, it pays back that investment in reusability. The absence of dependency on specific health states frees the elicited perspective to application across decisions without limitation as long as the patient remains stable in his or her perspective. The freedom of the administration also allows more frequent administration in the event of suspected change of perspective (e.g., patient education, change in health state) hence greater ease for longitudinal study.

The structure required to administer a values clarification exercise with a computer program causes a formalization of the process. At the same time consistency is provided by the non-human element. Computers can provide increased consistency while expanding the flexibility of data collection. An interview may be programmed such that every question of a survey is unique for the individual and context, yet consistently satisfies specified rules of composition or content. Paper surveys must be reprinted in order to change the questions in any way. Personal interviews afford such flexibility and allow more intelligence to be employed in the administration of the data collection. However, artificial intelligence techniques can provide much of the same benefit without the mandate of trained professional attendance restricting time and location of the exercise. They can employ terminology familiar to the patient as discovered mid-stream in the interview. Thus more demanding data collection tasks can be

accomplished at the convenience of the patient if the exercise and expertise of administration can be characterized adequately for knowledge based computer programming technology. Such an approach enables the treatment of the patient as an expert on his or her perspectives and the task becomes knowledge elicitation from an expert who does not necessarily articulate his or her expertise well without structured prompting. This approach has been applied to urban planning and the elicitation of values from individual urban residents³⁴.

The details necessary for the computer program to administer the values clarification and elicitation depends upon the early experience with the IMQOL methodology. Building such expert system programs is dependent upon refinements necessitated by discoveries made during early encounters with domain experts, which in this case are the clinical patients. While the methodology of the interviews is predetermined, the mechanisms by which the knowledge based program may organize and manipulate the interview responses to expedite the interview is based upon the detailed nature of the knowledge to be reasoned with, in this case the values of the dialysis patients. Thus, until I began to ask the questions presented in the IMQOL protocol for values clarification, I knew little about what kind of answers to expect. Therefore, the values elicitation was initially conducted manually until the clinical requirements of such a program were clarified. Subsequent development then proceeded on the basis of actual clinical interviews and will continue to do so.

3.8 How can this method of values clarification be validated?

In the absence of a gold standard in values elicitation for comparison, the validation of computer programmed elicitation of patient values is not easy. Psychometric disciplines provide guidelines, however, regarding the validity and reliability of a measurement mechanism that can be applied here. The dynamic nature of the values clarification tool which is unique for every patient makes reliability measures difficult to find. The elicitation may be repeated with the same patient but it is easy to imagine the patient will naturally remember his or her reply from one time to the next on such thought provoking topics. The requirement for

completeness in the list of attribute scales makes the likelihood slim for using multiple attribute scales to measure the same thing in order to monitor internal consistency. Nonetheless, repetitious evaluation is feasible, even encouraged owing to the improved reliability found in utility assessment in general when frequent administration allows increased familiarity with the instrument. Construct validity may be inferred when the systematic variation of scores based on these patient values correlate as expected with health status, compliance and satisfaction with decision. For example, no significant change in the value system would be anticipated during stable health status and the absence of additional medical information (as might occur on a visit with the health provider or receipt of a lab report). In contrast, when an effort has been made to educate the patient, success should be marked by a demonstrable change in the subsequently elicited value system. For example, if you convince them of the benefit of non-smoking, smoking versus non-smoking should show up in the next value system elicited. To this end, the research design includes validating steps in which health status measures are collected simultaneously with patient values over time to track the changes or stability in value systems as traditional status measures change or remain stable. Content validity is made easier by the nature of the method. Values expressed in the patient's words are more likely to be accurate semantics than the words supplied by others. Additionally, steps involved in testing the functionality of the attribute scales during the process allow the patient to recognize and correct misunderstanding.

3.9 Why so many outcome measures?

Although it is anticipated that a better understanding of the patient's values will lead to better outcomes¹⁵ the nature of what is better about the outcome has not been empirically evaluated. There is evidence that a patient's perception of control or choice may affect their biomedical status through immune mediated mechanisms³⁵⁻³⁷. It may be that the patient who perceives himself or herself to be better understood is more willing to comply with provider instructions and hence measuring compliance may support this hypothesis. Decreased

compliance in the presence of improving health status might be misinterpreted if compliance data alone is monitored. It may be that although there is no demonstrable change in biomedical or functional status, the patient is more satisfied with the decision and that is worth something in itself. The possibility remains that these categories may interact, e.g., increased compliance may cause improved biomedical or functional status and hence knowing the outcomes in terms of all three would be useful. In an effort to characterize the nature of the benefit of improving the outcome of patients, multiple outcome measures should be employed. Until validated, the IMQOL protocol can be employed in tandem with other measures. Not only is it important to understand how the IMQOL models relate to traditional utility assessment, but how they relate to other measures. I will propose then, as future work to understand the values captured in IMQOL models, to measure health status, compliance and satisfaction with decision in parallel with quality of life. Changes in the output of IMQOL quality of life assessment can be compared to changes in health status measurement with the MOS SF-36 questionnaire. To compare with compliance, prescriptions filled, instructions followed and quantification of appointments made and kept can be compared to changes in IMQOL models. Alteration in satisfaction with decision measured as described Holmes-Rovner, et al³⁸ can also be compared to changes in the IMQOL models. Just how these measures correlate with the changes in values captured should provide insight. This is what psychometrists do to evaluate the “construct validity” of an measurement instrument—they investigate the collaboration between alternate means of measuring the same or related constructs. Until a gold standard is found, we are better off looking for a convergence of indications from multiple sources.

3.10 Why dialysis patients?

The duration of dialysis provides a large block of time that the patient is immobile and often looking for relief from the boredom of inactivity. They are cogent and not suffering during the procedure. This provides an opportunity to engage them in conversation or structured interview. They are engaged in a chronic confrontation with their tenuous state of

health. The issues they face are substantial and are worthy of values clarification if we can provide it. The risk this research poses is only that of embracing the reality of their circumstance; the values clarification exercise is easily refused or discontinued if so desired. The possibility of reviving suppressed memories exists. However, the risk thereof is assumed minimal and pale in comparison to the possibility of improved compliance generally anticipated by better understanding one's own infirmities and means of control or that of understanding more thoroughly how everything possible is being done.

This population of patients have responded well to the administration of SF-36 Quality of Life assessment¹⁶ and a similar compliance could be anticipated in this study. The regularity of treatment provides a convenience for longitudinal aspects of this study. The issues that concern the ESRD patient perplex nephrologists and yet provide some diversity in scope. These patients are present long enough to accommodate the length of manually conducted interviews required in the early development of the values elicitation tools. The temporal properties of renal failure are insidious enough to allow some stabilization of these patients' values. Finally, and most important, knowing more about the patient's value of the stages of renal failure and alternative courses of therapy should shed light on the comparison of specific treatments and the means of matching suitable therapy plans with individuals.

Values clarification is of particular potential importance in the treatment of renal disease. Many decisions implicit in the care of patients with renal disease are never made explicit because patients, physicians and other providers lack an intellectual framework in which to apply general guidelines and principles to the individual. For example, the value of tight glucose control in preventing long term complications of type I diabetes was established by the Diabetes Control and Complications Trial. However, there is an implicit trade-off between the immediate burden of tight control, and its long-term benefits. In type II diabetes, there is not only a trade-off, but also greater uncertainty about the magnitude of the benefit. As pointed out in the previous chapter, there are many decisions involving tradeoffs for dialysis patients, sometimes very complex ones.

Chapter 4

Discordant Responses in Classical Assessments

Decision analysis is a rigorous technology that provides a model of a decision to help identify the critical issues that face the decision-maker in choosing treatment strategies. The values of outcomes represented in the model are critical to the validity of the model's implications. Recognized methods used to determine the value of an outcome from a patient's perspective exist and are known as "utility assessments". However, different methods often result in conflicting responses for the same individual, which may be referred to as *discordant* responses. Such inconsistency either questions the validity of one or more methods employed or indicates a poor understanding by the patient and consequently suggests an invalid result. Erroneous measurement of values can lead to inappropriate medical recommendations when those value measurements are used to support decisions involving tradeoffs. Inconsistent responses are therefore an important issue in the use of decision analytic approaches to medical decision support, health care protocol development and resource allocation where utility assessment is employed.

When answers to different questions lead to different conclusions rather than consistently lead to the same conclusion, one is left doubting his or her understanding. This is particularly true in learning the values of others with utility assessment. Questions are asked of the patient and conclusions derived from the responses. The responses that lead to mutually exclusive conclusions can be called *discordant* responses as opposed to *concordant* responses. Responses that differ by leading to different conclusions that are not mutually exclusive are helpful leading to a broader understanding, but mutually exclusive conclusions suggest the presence of logical inconsistency. Some responses may appear mutually exclusive, and hence, logically inconsistent at first, but with further thought or explanation don't turn out to be. Others defy logical explanation.

Discordant responses may occur for a number of reasons. We can systematically explore the possibilities by considering the elements of the interaction individually. The interaction consists of questions asked, responses given and inferences made about the responses. The respondent may not have understood the questions. Perhaps they were not asked correctly or were confusing in delivery. The respondent might be making erroneous responses. Perhaps the environment is distracting. Perhaps the respondent is tired or out-of-sorts. The respondent may, in fact, be irrational in his or her thinking. Perhaps the patient is just being inconsistent in his or her responses. Perhaps the patient is miss-articulating his or her intended response or having a hard time finding any words that capture his or her sentiments. Perhaps none of the above is true and it is the understanding of the responses that is erroneous. Perhaps concluding that the responses are discordant is hasty. We must be careful not to confuse inadequate evidence with inconsistency. Perhaps the respondent's line of reasoning is not fully evident and what appears to be mutually exclusive is not so.

If discordant responses are truly representing inconsistency, are all inconsistencies alike? Is consistency a binary phenomenon – either one is or one is not consistent – rational or irrational? For most people it is considered all or none. However, when it comes to people expressing their values, some people are more consistent than others. This suggests that

consistency might be a matter of degree. If so, it is worth thinking about how one should regard people who express themselves with a small degree of inconsistency versus those who express vast inconsistencies. How much inconsistency is significant inconsistency? This chapter will show that discordant responses provide a metric by which to explore the issue.

The substance of the chapter is the product of investigating the proposed protocol for values assessment. Early in the investigation comparing results of traditional methods with the newly proposed method, discordant responses were notable in traditional methods. At the same time, a result of the multiattribute preference elicitation clearly provided more explanation than traditional methods. What first appeared to be a case of irrational response turned out to be explainable as rational and made the notion of mild inconsistency more tenable. This only served to further drive home the need to systematically uncover rational explanations for seemingly inconsistent responses. In this chapter I will discuss the nature of discord and propose, on the basis of traditional utility assessment results, that there exists a taxonomy of discord types that begs explanation and motivates the desire to better understand the respondent. The motivation will be substantiated by the illustrative example. This motivation will then be rewarded in subsequent chapters with the means to elicit better understanding for how quality of life is measured by an individual.

4.1 Discord is Endemic

Utility assessment is a required part of decision analytic approaches to medical decision support, health care protocol development and resource allocation. The gold standard for utility measurement eludes us, but desperate for an understanding of the patient's preference, analysts, clinicians and policy makers wish to make the most of what technology has been developed for assessing utilities. Utilities assessment is an expensive process. It requires trained analysts to interview the individual or patient population to be represented. If taken seriously, these interviews usually involve some consideration of grave outcomes, which are at least sobering if not traumatic for the person interviewed. Questions of reliability and stability of responses

plague the widespread acceptance of the results as useful information. In an effort to standardize administration of the utility assessment, computer programs have been developed. These programs ease the demand for trained experts to perform utility assessment. A natural progression to Internet administration with web pages has ensued, relaxing the constraints on when and where the assessments can be performed.

As the employment of utility assessment proliferates what happens to the rate of discordant responses? The literature³⁹⁻⁴² suggests that discordant responses are rarely absent and perhaps occur with a regular frequency (17-30% or more). Reporting on the use of web page administration of utility assessment using standard gamble and visual rating scale methodologies, Lenert, et al⁴³, observe a 20% and 36% rate, respectively, of discordant responses between these assessments and the ordering produced by pair-wise comparison. Furthermore, they found that neither self-assessment of understanding nor of confusion predicted the consistent or inconsistent responses. Little has been done to investigate this significant portion of the population of responses and the meaning or significance of the discord. It represents a significant portion of the population of patients. If the methods of utility assessment are only considered valid in some portion of the population, a rigorous method must be employed to qualify which responses are valid for use in decision support and which are not.

So how do studies of utility assessment deal with this significantly high discordant response rate? The most common method of dealing with these responses is to exclude them from the sample statistics as incoherent or confused respondents. If responses for ordered health states are not consistent with the axioms of von Neumann Morgenstern utility theory^{44, 4}, then the conclusion is drawn that the individual is not rational. Alternatively, the methodology employed to assess the individual's utility values is questioned. I propose yet another perspective on discordant responses. Perhaps they are not all discordant. In a study I performed eliciting utility values for states of health which follow stroke, a high rate of discordant responses was observed. This study employed the standard gamble method of utility assessment along with rank order on one mild, one moderate and three grave states of health;

two of the grave states included death. Close examination of these discordant responses reveals that not all discordant responses are alike. Simple qualitative and quantitative views of these differences suggest that there may be information outside the concordant population of responses, which is lost by excluding discordant responses. In an effort to understand the elevated discordant response rate, the effect of relaxing the defining boundaries of a discordant response was explored. A refined definition of discord is presented as a systematic approach to extending the population of responses that may be considered rational and hence included in descriptive statistics used to characterize the sampled population. Some characterizations are offered for the observed impact on descriptive statistics when a progressively loosened definition is permitted.

4.2 Not all discord is the same

To validate an assertion that not all discord is the same, I include a description of a study conducted to better understand the preference visitors to an Emergency Department have for long-term disability versus sudden death.

4.2.1 Utility assessment for consequences of stroke: Methods

To address the question whether patients prefer to avoid death or disability because of stroke, rank order and standard gamble utility assessment were used to interview visitors to an emergency department. This pursuit was motivated by the interest in deciding the value of thrombolytic therapies for acute stroke⁴⁵⁻⁵³. This decision was viewed in the context of the onset of signs of stroke within a few hours. Based on a simplified decision model, the motivating scenario involves a tradeoff of decreased disability in patients treated with thrombolytic therapy (e.g., tissue plasminogen activator - tissue plasminogen activator) for a slight increase in the risk of death within a few days. In a model that compares treatment with tissue plasminogen activator to treatment without tissue plasminogen activator the outcomes to both would be the possibility of varying degrees of disability, death within days and, distinctly,

death at a more distant interval. Five post-stroke states of health were identified and descriptions chosen based on disjunctive combinations of the Rankin Scale of Disabilities⁵⁴. These are listed in the appendix. Three states represented grave outcomes: death within 2 days, death within 6 months and severe disability without a shortened life span. The remaining two states were moderate and mild disabilities.

Table 4.1. Rankin Scale of Disabilities.

Health State 1: Some symptoms, e.g., slurred speech, numbness in your face, or reduced strength in an arm or leg, but...Still able to carry out all your usual duties and activities.

Health State 2: You are unable to carry out activities you could participate in prior to the stroke;... You require some help looking after your own affairs, but... You are able to walk without assistance

Health State 3: You are unable to walk without assistance and unable to attend to own bodily needs, or bedridden, incontinent, requiring constant nursing care

Health State 4: Sudden painless death within two days

Health State 5: Death within six months following the stroke

The inclusion criteria for the study population were that the volunteer must be English-speaking, 18 years or older, coherent, not experiencing signs of stroke, a non-psychiatric admission to the emergency department with a triage rating of 3 or 4 out of four and present long enough to conduct the interview. Two undergraduates of a highly ranked institute of technology were specifically trained to conduct these interviews. Candidates were approached with an introduction and invitation to participate. Consenting patients were assured that the interview questions had absolutely nothing to do with their actual admission and in no way would their responses affect the treatment for their current condition. They were then asked to rank the five states of health evaluated. The order of presentation of the states was systematically randomized. The state ranked worst was used for the negative side of the lottery in all standard gamble assessments that followed. The patient was asked to hold an 8x11 “scoreboard” visual aide, on which was placed preprinted stick-up notes containing health state labels reflecting the descending order of the patient’s rank assignment. After a brief explanation of the general frame of the standard gamble questions, each of the four states of health not regarded as worst were assessed. The order of assessment for these states was systematically randomized according to assigned rank. For each health state the patients were asked to imagine they had been admitted to the emergency department and told that they were showing signs of stroke and that they were offered a choice of treatments: (1) “traditional” treatment for which the results are certain and the outcome described by the health state description or (2) a new treatment in which most people are restored to perfect health for their age and gender, but that in some specified number out of 100 people the result is an outcome described by the state they regarded as worst of the five investigated. To minimize anchoring, patients were asked to offer any number of people out of 100 which would be required to experience the beneficial outcome before they would desire the new treatment. After their commitment to any number, a visual aid called the “pie tool” was used with a bracketing algorithm to refine the assessment of their point of indifference or change in decision. This tool was a simple pie chart that could by

changed by hand to represent any proportion in red (unfavorable) and white (favorable) pie shaped regions. Placards, 8x11” with a clear bold font, containing the description of the outcome of “traditional” treatment (the health state currently being evaluated) were the third visual aid present during the elicitation. As a point of indifference or change in decision was elicited and verified by visually aided query just above and below the apparent utility value, that number between 0 and 100 was written on the scoreboard next to the state of health being evaluated. In this way, the patient could easily see discordant responses as they occurred. It was anticipated that this would minimize the oversight of such discord on the part of the patient. If the patient indicated a utility value that was out of order with their rank order of health states, it was pointed out to them. The interviewers were trained to do their best to make sure the patient understood why others would see that something was out of order with brief explanation. They were to ask the patients if they wished to change their response, but to permit the patient to leave it unchanged if that is what they desired.

4.2.2 Utility assessment for consequences of stroke: Results

Of the 63 patients invited, there were 36 consenting participants (52.4%) ranging in age from 19-79 (average 42.5; 20 female; 16 male). Only one (Male) did not complete the interview because he could not understand the questions. Two patients were unable to complete the interview because the hospital staff became available for services interrupting the interview. Twenty interviews for which time required to complete was recorded took from 10 to 60 minutes (mean 28.2 minutes, median 25 minutes). Nineteen out of thirty-three (57.6%) responses were discordant by one or both of the following criteria:

- One of the three grave states scored equal (6 instances) or higher (12 instances) than either of the mild or moderate disability states.
- Score of the moderate disability state higher than the score of the mild disability state. (2 times + 1 time combined with an equal valued grave state + 1 time combined with a greater valued grave state).

Fourteen out of thirty-three (42.4%) responses did not meet either criterion, a sub-population which we shall refer to as *severity-concordant*. I will use the term *strict concordance* to refer to any response that gives no evidence of discord no matter which way you look at it. *Severity-concordant* responses then are those that are *strictly concordant* with respect to the severity of diminished health in the eyes of health care providers. There is no doubt that one could find something in this domain upon which some providers would not agree on the rank of severity, but in principle there is a consensus of opinion with which few would contend. Quality of life ratings that are consistent with states ranked by severity in this common view would represent *strictly concordant* responses in utility assessment.

An alternative definition of discord based on rank order yielded different results. Ignoring the identity of the health state and looking only at the consistency between elicited values and rank order suggests that 18/33 (54.5%) were concordant. Two of the responses had equal utility values for the best-ranked state and two or more states ranked below the best. In the context of only two states not regarded as grave by the investigators, these might be questioned if regarded as concordant. With these removed from the tally, that leaves 16/33 (48.5%) concordant responses from this perspective. We shall hereafter refer to these as *rank-concordant* standard gamble responses. Four (12.1%) of the responses not meeting the criteria for strict concordance were rank-concordant. Two responses that did meet the criteria for strict concordance were *not* rank-concordant. So, it is apparent that it is possible to have responses that are *severity-concordant*, but not *rank-concordant*. Furthermore, it is apparent that it is possible to have responses that are *rank-concordant*, but not *severity-concordant*. We can explain *strictly concordant* responses in this context as those that are both *severity-concordant* and *rank-concordant*. We can apply the “consistency across multiple preference assessments” (CAMPA) criteria of Lenert, et al, to these interviews resulting in 12/33 (36.4%) responses which were both *severity-concordant* and *rank-concordant* which can serve as the operational definition of *strictly concordant* in this context.

In 25 of the 33 (75.8%) completed interviews, the severe level disability was regarded as the worst or equivalent to the worst possible state of health. Ten (30.3%) of these were strictly concordant. Only 8 (24.2%) regarded the event of death within 2 days of stroke symptom onset as the worst possible state or its equivalent. Five (15.2%) of these were strictly concordant responses. Four (12.1%) interviewed subjects regarded death within six months of stroke symptom onset as worse than either of the other two grave outcomes. Three (9.1%) of these were strictly concordant responses. One concordant response indicated all three grave outcomes were of the same value. Two concordant responses indicated the severe disability and death within 2 days were of equal value as the worst possible outcome. Most patients, *but not all* (29/33, 87.9%), indicated mild disability was the best outcome.

Further characterizing the phenomenon of discord, it was noticed that 8 (24.2%) responses not meeting the criteria for strict concordance involve an “erroneous” margin of 10 percentile points or less. The choice of 10 for this suggestion of a tolerance value is arbitrary. It appears to the naked eye to be a clustering boundary. Cluster analysis could form the basis for a refined choice for a tolerance value, but that is beyond the scope of the treatment given in this thesis. It will suffice to introduce the concept of a tolerance parameter and I will leave the appropriate choice for that value for later investigation.

Pat.ID	Rankin 0-1	Rankin 2-3	Rankin 4-5	Acute Death	Chronic Death	Time	Age	Sex	Educatin
102	99	80	0	60	60	25	27	M	13+
103	40	20	0	40	40	15	37	F	13+
104	40	5	0	40	45	20	47	F	13+
106	70	70	0	70	63	15	41	M	13+
107	98	98	0	98	98	15	65	M	0-8
108	90	70	80	0	50	50	31	M	9-12
109	93-95	83	0	80	80	20	26	M	13+
110	68	90	0	70	82	35	55	F	13+
111	75	60	0	80	40	30	50	M	13+
112	84	25	0	80	1	20	56	F	13+
113	50	60	0	50	50	25	37	F	9-12
114	100	98	0	1/1000	1/million		51	F	9-13
201	97-99	45-50	30-40	45?	45?	20	36	F	13+
202	100	100	0	99	100	60	50	M	
203	100	100	0	0-5	?	41	?	F	
204	99-100	5-10	0	80-85	40-50	22	?	M	
205	99	95	35	0	60	20	37	F	13+
206	98	77	0	78	75		37	F	13+
207	95	90	40	0	30		40	M	13+
208	70	60	0	10-15	10-15		62	M	13+
210	40-45	40-45	0	0	30		20	F	13+
211	100-95?	85?	0	0	0		46	M	9-12
212	88	70	40	0	75		79	F	13+
213	58	80	5	0	15		26	F	9-12
214	75	75	0	90	75		19	M	13+
215	55	60	0	50	<1		27	M	13+
216	95	95	70	77	0		77	F	9-12
218	80	80	0	40	60-70	35	62	F	9-12
219	96	95	0	95	95	25	22	F	9-12
220	90	75	50	10	0		23	F	13+
221	75	50	50	0	50	10	64	M	9-12
222	75-80	50-60	0	20	20		53	F	13+
223	90	20	0	50	10		46	M	13+

Legend:

- Severity-Concordant response
- Grave state rated equal to a mild or moderate disability
- Grave state rated higher than a mild or moderate disability
- Moderate disability rated higher than a mild disability

Figure 4.1 Results of standard gamble utility assessment for five outcomes of stroke arranged by severity of outcome according to biomedical perspective.

Pat. ID	First	Second	Third	Fourth	Fifth	Discord Type	Summary Counts	Legend
102	99	80	60	60	0	strong		
103	40	40	40	20	0			
104	40	5	45	40	0	weak δ	Strictly Concordant (all criteria) 12 (36%)	
106	70	70	70	63	0			
107	98	98	98	98	0	strong δ	w Severity-discordant/rank concordant 16 (48%)	
108	90	70	80	50	0			
109	94	83	80	80	0	strong	w only weak discord by rank order 18 (55%)	
110	68	70	90	82	0			
111	75	60	80	40	0	strong δ	Rank-Concordant 16 (48%)	
112	84	25	80	1	0			
113	50	60	50	50	0	strong	w only weak discord by rank order 18 (55%)	
114	100	98	1E-04	1E-06	0			
201	98.5	47.5	35	45	0	strong? δ	Severity-concordant but discordant with rank order 2 (6%)	
202	100	100	99	100	0			
203	100	100	2.5	?	0	strong	Discordant unless rank adjusted 4 (12%) w weak discord allowed 6 (18%)	
204	99.5	7.5	82.5	45	0			
205	99	95	60	35	0	strong	Discord involving a margin of 10 or less 9 (27%)	
206	98	78	77	75	0			
207	95	90	40	30	0	strong? δ		
208	70	60	12.5	12.5	0			
210	42.5	42.5	30	0	0	strong δ		
211	97.5	85	0	0	0			
212	70	88	40	75	0	strong		
213	58	80	5	15	0			
214	75	90	75	75	0	strong δ		
215	55	60	50	0.5	0			
216	95	95	70	77	0	strong δ		
218	80	80	65	40	0			
219	96	95	95	95	0	strong δ		
220	90	75	50	10	0			
221	75	50	50	50	0	strong δ		
222	77.5	55	20	20	0			
223	90	50	10	20	0	strong δ		δ

Figure 4.2 Results of standard gamble utility assessment for five outcomes of stroke arranged by patient ranked order of outcome according to biomedical perspective.

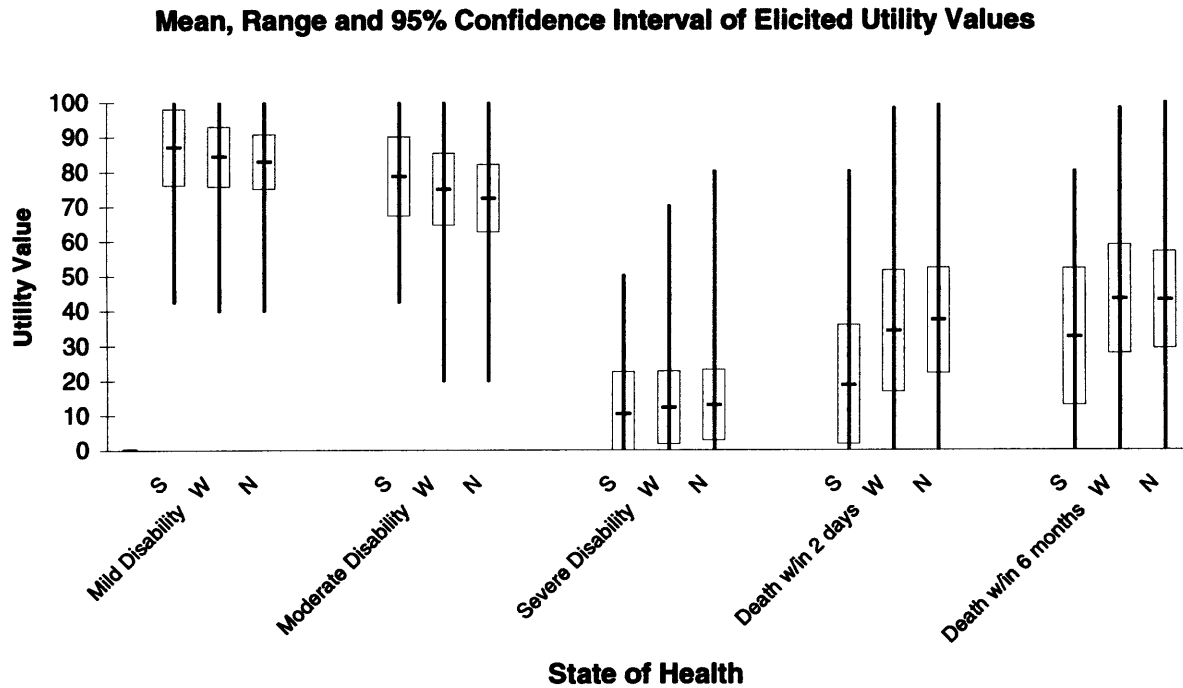


Figure 4.3 Mean (cross-hair), range (whisker), and 95% confidence interval (box) of the elicited utility values for mild disability, moderate disability, severe disability, death within two days and death within six months following a stroke. Triplets represent (from left to right) the "strictly-concordant" responses (S), "at-least-weakly-concordant" responses (W) and "nearly-concordant" responses (N).

Noticing the inclusion of equality of utility values for disparate health states as part of the first criteria for severity-concordance, I propose further to divide the group of responses meeting the first criteria of discord into "weak" versus "strong" discordance. *Weak* discord is

used to refer to disparate health states assessed as equal in utility value (i.e., equal values are asserted for any grave state and for the mild or moderate state) and *strong* discord is used to refer to the case of unequal but disordered utility. Using three definitions for discord, descriptive statistics were compared using (1) strict concordance as the most restricted case, (2) a modified version of strict concordance which included weakly discordant responses and (3) the consideration of a response as discordant only if the margin of discord exceeded 10 percentile points. These three cases are referred to as *strictly concordant*, *weakly concordant*, and *nearly concordant* respectively. It has already been pointed out that 12/33 (36.4%) responses were strictly concordant, 20/33 (60.6%) were weakly concordant and 25/33 (75.8%) were nearly concordant.

For the mild disability state the mean utility value given by the strictly concordant cohort was 87 with a 95% confidence level of 10.95 and range of 57.5 (42.5-100). For the weakly concordant cohort the mean was 84.3 with a 95% confidence level of 8.62 and range of 60 (40-100). The nearly concordant cohort had a mean of 82.9 with a 95% confidence level of 7.89 and range of 60 (40-100).

For the moderate disability state the mean utility value given by the strictly concordant cohort was 78.6 with a 95% confidence level of 11.36 and range of 57.5 (42.5-100). For the weakly concordant cohort the mean was 74.9 with a 95% confidence level of 10.36 and range of 80 (20-100). The nearly concordant cohort had a mean of 72.3 with a 95% confidence level of 9.7 and range of 80 (20-100).

For the severe disability state the mean utility value given by the strictly concordant cohort was 10.41 with a 95% confidence level of 12.15 and range of 50 (0-50). For the weakly concordant cohort the mean was 12.25 with a 95% confidence level of 10.59 and range of 70 (0-70). The nearly concordant cohort had a mean of 13.0 with a 95% confidence level of 10.29 and range of 80 (0-80).

For death within 2 days the mean utility value given by the strictly concordant cohort was 18.75 with a 95% confidence level of 17.18 and range of 80 (0-80). For the weakly

concordant cohort the mean was 34.15 with a 95% confidence level of 17.34 and range of 98 (0-98). The nearly concordant cohort had a mean of 37.28 with a 95% confidence level of 15.1 and range of 99 (0-99).

For death within six months the mean utility value given by the strictly concordant cohort was 32.5 with a 95% confidence level of 19.66 and range of 80 (0-80). For the weakly con-cordant cohort the mean was 43.34 with a 95% confidence level of 15.59 and range of 98 (0-98). The nearly concordant cohort had a mean of 43.1 with a 95% confidence level of 13.94 and range of 100 (0-100).

4.2.3 Utility assessment for consequences of stroke: Discussion

These results show that the definition of concordance affects the descriptive statistics for collective utility assessments. The impact can be characterized to motivate a systematic approach to considering a set of extended definitions that avoid complete loss of information outside the obviously concordant responses.

It should be noted that Torrance⁴⁴ suggests that standard gamble is ill suited for utility assessment at the extremes of the range for utilities arguing that it lacks the power to distinguish between fates close to death and close to perfect health. This would lead us to expect an indistinguishable convergence of values for states of health near death. Such was only the case for some individuals in this study. It was sometimes true of intermediate health states, i.e., mid-range values given for both grave and non-grave states (e.g., moderate disability, severe disability and death in six months all assessed as 50). From a collective viewpoint, graphs of utility value distributions for strictly concordant, weakly concordant and nearly concordant responses are not easy to distinguish (Figure 4.3). Patrick, et al^{55, 56}, found the category scaling, rank order, time tradeoff and standard gamble method for utility assessment suitable for states considered worse than death. They also reported only 18% agreement between rank orders of the four methods in well adults and no successful agreement in nursing home residents. The difficulty appears most severe for states near or worse than

death. They conclude only that the cognitive burden of these methods must be reduced if these assessments are to be used in frail older institutionalized adults. No mention is made of the degree of discord or any patient explanations offered for such responses.

Certain properties can be observed which should characterize the expected effects of relaxing the definition of discord. For states near perfect health, many patients would be expected to give utility values close to 100. Since there can be no values greater than 100 on a scale from 0-100, the only range of values which can be more populated by unusual responses would be those further below the strictly concordant population which clusters beneath the upper bound of 100. This effect is observed for the mild health state evaluated in this study (Figure 4.4) where the range shifts little but the mean drops as the boundary for concordant responses is relaxed. Conversely, for states near the worst extreme, strictly concordant respondents will give values clustering at the bound of zero. The only effect possible by including other respondents is to populate the data set with higher values. This effect is observed for the severe disability state evaluated in this study (Figure 4.6). The mean increases as the bounds for concordant responses are relaxed. The effect is observed to a progressively lesser degree for death within 2 days (Figure 4.7) and death within 6 months (Figure 4.8).

For an intermediate state of health, one might expect a more distributed range of utility values making it unpredictable which direction the mean would change with the relaxation of bounds for concordance. However, it would make sense that the range should be increased by expansion in either direction. In this study, the utility values for moderate disability exhibit such characteristics (Figure 4.5). There is a change of range from 57.5 to 80 percentile points with more range added to the lower end of the scale than the upper.

It is interesting to note that the utility values for death within six months look more like those for moderate disability than those for severe disability. Distributions of the values given populate the center of the range more than the zero extreme in this state (Figure 4.8). I regard this as a stronger argument that severe disability is worse than death for most people than that made by comparison of mean utility values alone.

A weakly discordant response could be considered a rational response if one is willing to accept that the patient actually considers neither state better than the other. Alternatively stated, the states may not differ in a significant dimension in the patient's perspective. For example, if a patient is terminally ill with AIDS, and also has a compensatory high regard for mental health over physical incapacity, he or she might score any health state that does not affect mental health with relatively equal value. He or she is simply using an adaptive multidimensional model for health that heavily weighs the mental health dimension. Details of an interview in which this was the case will be given at the end of this chapter as an illustrative example.

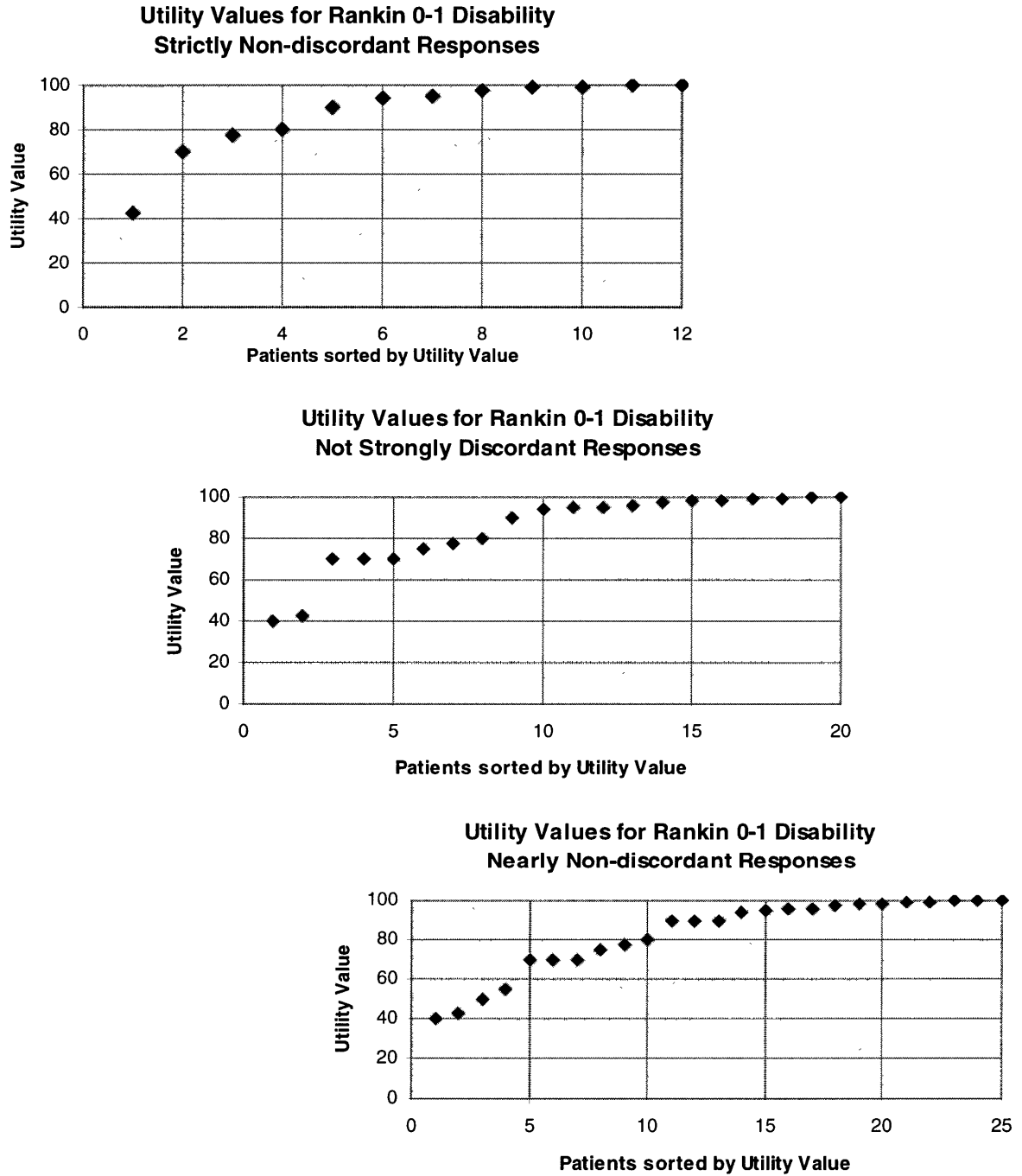


Figure 4.4 Effects of loosening the boundary for discordance in Rankin 0-1 Disability Assessments of Quality of Life.

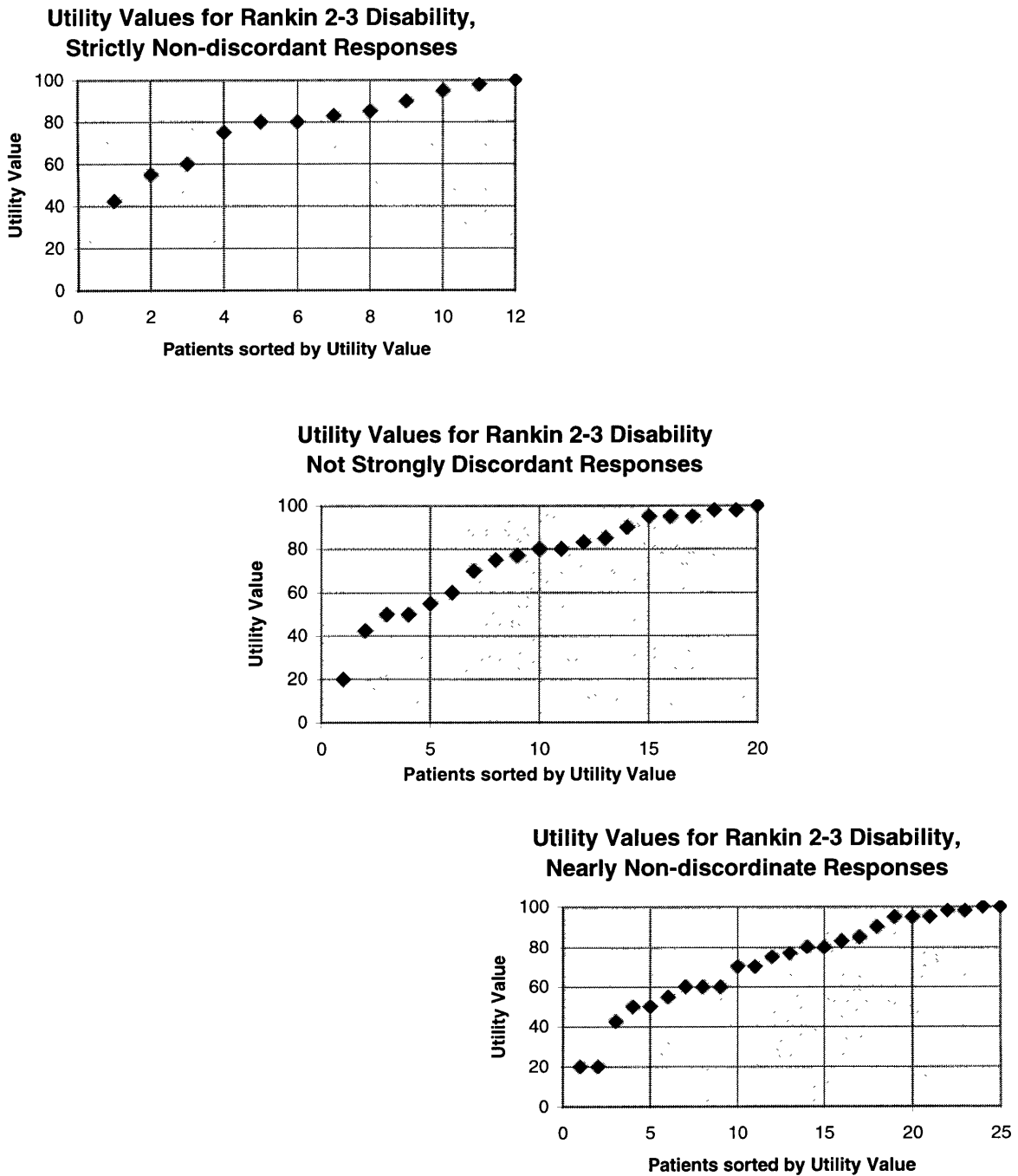
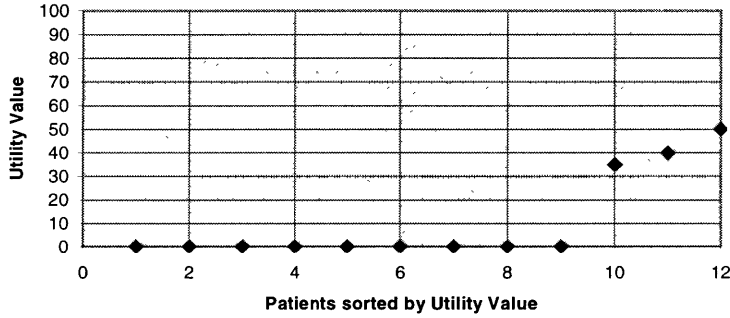
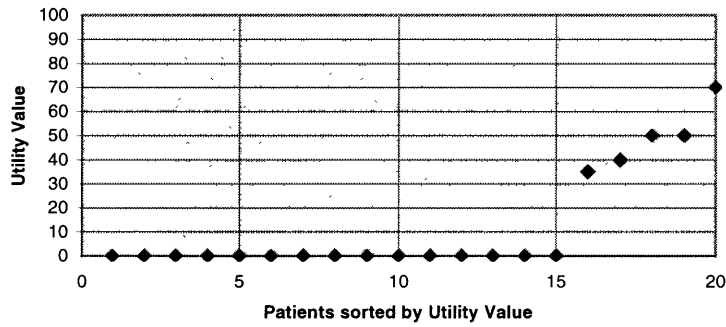


Figure 4.5 Effects of loosening the boundary for discordance in Rankin 2-3 Disability Assessments of Quality of Life.

Utility Values for Rankin 4-5 Disability, Strictly Non-discordant Responses



Utility Values for Rankin 4-5 Disability, Not Strongly Discordant Responses by Rank or Numerical Comparison



Utility Values for Rankin 4-5 Disability, Nearly Non-discordant Responses

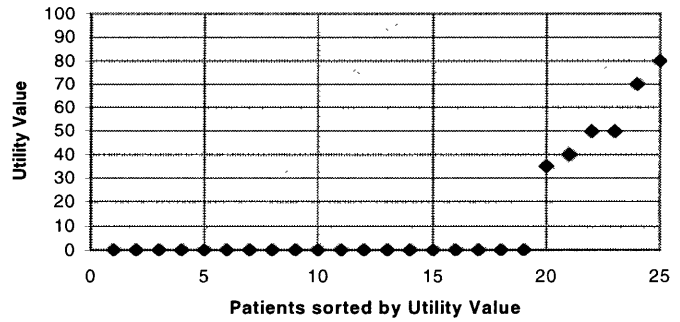
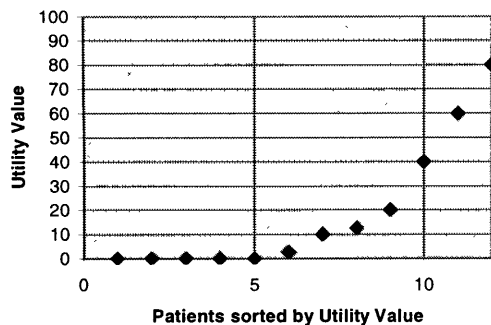
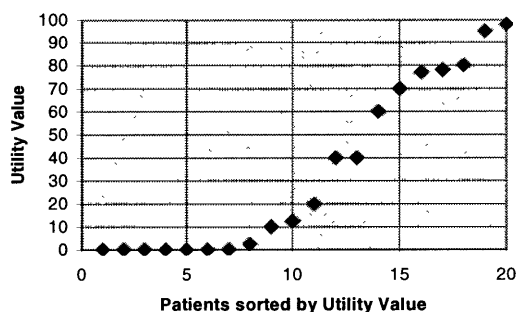


Figure 4.6 Effects of loosening the boundary for discordance in Rankin 4-5 Disability Assessments of Quality of Life.

Utility Values for Acute Death, Strictly Non-discordant Responses



Utility Values for Acute Death, Not Strongly Discordant Responses by Rank or Numerical Comparison



Utility Values for Acute Death, Nearly Non-discordant Responses

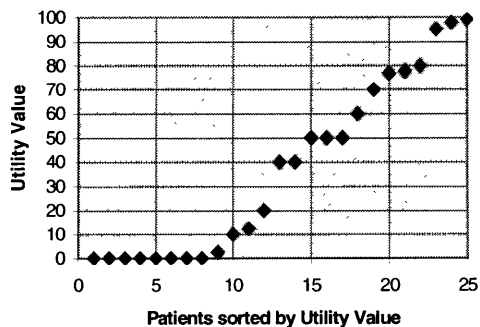
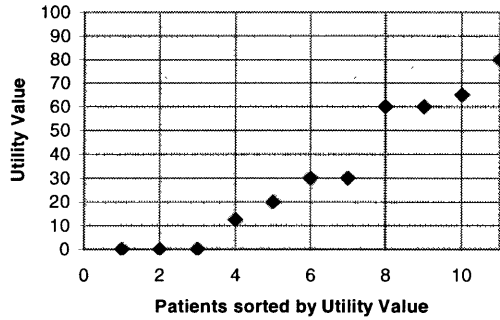
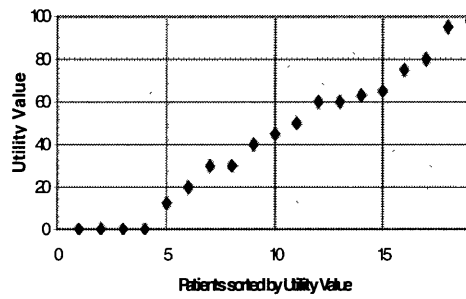


Figure 4.7 Effects of loosening the boundary for discordance in Death within 2 days Assessments of Quality of Life.

**Utility Values for Six Month Death,
Strictly Non-discordant Responses**



**Utility Values for Six Month Death, Not Strongly
Discordant Responses by Rank or Numerical
Comparison**



**Utility Values for Six Month Death,
Nearly Non-discordant Responses**

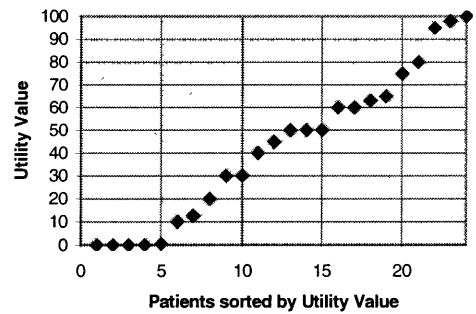


Figure 4.8 Effects of loosening the boundary for discordance in Death within Six Months Assessments of Quality of Life.

If disparate states can be given equal utility values, then it is not too big a stretch to consider disordered values with small margins. These small margins are similarly explained by unappreciated dimensions in the patient's point of view, which are only poorly revealed in univariate utility for states of health. Perhaps the inversion of values is a matter of "noise" in the data, elicitation technique, or value system of the patient. It is compelling to regard responses that are nearly concordant with more regard than those with vast margins of discord. One is left to ask how a person could come so close to being rational or consistent and yet miss by a narrow margin. If we introduce the notion of a tolerance threshold for the margin of discord, we have a parameter for gauging the relaxation of the boundary for discord. In this study varying the tolerance for discord from 0 to 10 changes the collective description of death within 2 days more than it does for severe disability.

Relaxing the criteria for concordance allows a broader representation of the sample population that includes more of the unusual perspectives. It more accurately represents the diversity of the population. In application of utility assessment to individual decision making, systematically relaxed concordance criteria enables more persons to benefit from the expression of their decision in rigorous terms. Health care providers can be informed by further discussing with the patient the meaning of weak or nearly discordant results. As with the traditional method for handling discord, decision support should not be anticipated from this form of utility assessment in the case of responses that remain outside the nearly concordant population.

4.3 We understand discord poorly

The mere presence of results which can be explained yet lie outside the boundaries of strict concordance warn us to be careful in the interpretation of study results where these results are discarded. What is being discarded are not those which make the descriptive statistics more normal. Patients who respond with discord may or may not be doing so for reasons that justify exclusion. Hence, false negative assessments of rationality are possible using definitions for

discord that are not strict. This study illustrates the beneficial effects of qualifying standard gamble assessment with rank order assessment when the states of health can be ordered. It also illustrates the fact that not all responses which first appear concordant turn out to be so. Care must be taken in the interpretation of results from both traditional strict concordance based studies and loosened concordance based studies.

Two experiences in this study illustrate a potential for rational perspectives that explain the responses captured as marginal to those strictly concordant. One patient explained her ranking of death within 2 days as the best possible outcome. She pointed out her devotion to her work in a lab and identified all the outcomes, even mild disability, as prohibiting her return to work in that lab. She would simply rather die than not be able to go to her lab. A second elderly lady, having her discordant response on the scoreboard pointed out, responded, “When you get as old as I am, you will understand these things.” Humorous or sad, these patients clearly are convinced of their own rationality and will be asked to employ the same if their “informed consent” is solicited for any decision made.

4.4 We need a systematic approach to fill in our lack of understanding

When utility assessment is employed for practical purposes, it must be decided how to handle discordant responses. Irrational responses in utility assessment will lead to erroneous conclusions. If the individual is not operating under a consistent code of rationality, then he or she is acting unpredictably and value assessment is for naught. If there is some consistent rationality to a person’s system of values and preferences then we would ideally like to capture that in utility assessment. If there are discordant responses that do not represent irrational value systems and preferences, as characterized in previous sections of this chapter, then we need to know how to interpret those responses.

In previous studies, the choice has been to disregard any discordant responses. This reflects a disregard for the value of these responses – because they are not rational, they are useless. It assumes all discordant responses are equal. How can a person come close to being

rational and just miss by a small margin? Is it a stochastic accident or something with more meaning? Furthermore, what is it that causes the confidence of such responders? Do they know something we don't yet? If their reasoning was strictly wrong, wouldn't it be replaced or at least be subject to admitted uncertainty?

Lenert's CAMPA method for recognizing consistent responses heightens the motivation for understanding discordant responses because it increases the population. CAMPA exposes responses to utility assessment that disagree with each other for one individual. What otherwise would be recognized as rational responses, when considered independently, are recognized as discordant. If there is to be any hope of valid utility assessment in such individuals, we must be able to systematically make sense of what the various types of discord are telling us.

4.5 How does this proposed values representation and elicitation fill in our understanding?

4.5.1 Decomposable model.

Keeney and Raiffa ⁵⁷ maintain that to minimize ambiguity, a utility model should be decomposable. Currently employed utility assessment methods (analog rating scale, standard gamble and time-tradeoff, etc.) do not provide decomposable output. When the numeric result of one of these methods is faulted, there is no additional information to suggest why it is faulted. There is no inherent way of knowing where the reasoning went wrong. With no gold standard for utility assessment there is no way to gauge which result is erroneous when two methods give conflicting results. For any multi-dimensional model, the hope of diagnosing which contributing dimension is at fault promises an opportunity to isolate a fault if one truly exists. If discord is asserted merely on the basis of insufficient evidence of rationality, then the multi-dimensional view offers more detail for the context, perhaps illuminating the interpreter's

omission that could remedy the insufficiency. With more detail, the chain of inference that successfully leads to the assertion is harder to miss.

4.5.2 Words of the individual.

If the vocabulary of the individual is used to express patient values rather than the biomedical or utility assessment community, an insight to the individual's chain of inference is more likely available. The individual is more likely to be able to explain him or herself in his or her own words. If we force the individual to explain his or her values in unfamiliar terms, it should not surprise us to find gaps in explanations and inconsistencies, both apparent and real.

What do the numbers found in the previous figures of this chapter tell us about the patient? They give us a bottom line with no explanation. A "43" tells us nothing about how the individual arrived at a 43 or how he or she was reasoning when forced to make a choice between medical outcomes. We observe the population of responses change as the definition of discord is altered, but dots and descriptive statistics give us no clue as to why those individuals are not like others. If patients were willing to try, could we reasonably expect them to explain their reasons accurately in unfamiliar biomedical terms? Only when patients explain themselves in their own words can we expect to minimize the losses that otherwise occur in translation. We know why one female laboratory worker thinks a mild disability is worse than death because she told us in her own words.

4.5.3 A Complete, Comprehensive and Operational Utility Model

If the utility model is complete and comprehensive, it may become more evident how certain factors play an unexpected role in the preferences of the individual. There may be either canceling factors or relative weights that play vivid roles in the overall evaluation of health that are obscured by looking at a single utility value for each of a few specific health state descriptions. If each individual dimension of the multi-dimensional model is expressed in

operational terms, i.e., terms that enable others to use the same scale in the same way, it should be easier to see how the individual arrives at the values ascribed.

4.5.4 Example from early investigation

One patient interviewed in this study gave responses that resulted in surprisingly high values for health states with the IMQOL measurement. The results indicated a score of 0.961 on a scale from 0 to 1 for his own quality of life and everything from no symptoms to death was rated higher. At face value this might have been interpreted, as a discordant response were it not for the multidimensional aspect of the IMQOL model. Previous standard gamble utility assessment of the same states of health with this individual gave more middle range values distributed in the range from 0.01 to 0.7 on a scale of 0 to 1. In both cases rank order corroborated with the order of the scores. Admittedly, the results of traditional utility assessment would not have raised an eyebrow in this case, but the IMQOL model illustrates both skewed values that can have a rational explanation and the advantage of decomposable utility assessment in seeing the explanation. I will first explain the table of results (Figure 4.9) and then tell the story of this individual that makes sense of the hypothesis suggested by those results.

Attribute	Relative Weight	Self Assess	Mild Disability	Moderate Disability	Severe Disability	Death w/in 2 days	Death w/in 6 months
Mental Health	0.96	1	na	na	0	1	na
Active Life	0.0024	0	0.6	0.6	0.6	0	0
No chronic illness	0.036	0	1	1	0	0	0
Obsessiveness	0.00077	1	na	na	na	na	na
IMQOL Results		0.961	0.998	0.998	0.002	0.961	0.961
Standard Gamble Results			.7	.5	.18-.2	1	1

Figure 4.9 Results of an interview showing how multiple dimensions explain results. “na” indicates the individual felt there was no relationship between the state of health in that column and the attribute of that row. In that case, the score he assessed his current health was used to compute the overall score.

The individual revealed four attributes that added up to what he though defined health. More detail was elicited regarding these attributes than is shown in the table, but topical labels suffice for the purposes of this discussion. The relative weights for each attribute, derived in the interview, are given in the second column. The rest of the columns above the gray bar contain the value for the individual attribute (on a normalized scale from 0 to 1) in each row. Values are given, from left to right, for the individual’s self-assessment and the disability levels from the study discussed earlier in this chapter (Section 4.2). Cells containing “na” (for *not applicable*) indicate this individual felt that the state of health in that column had no bearing on a person’s rating in that attribute (e.g., a weakness or numbness in the face considered a “mild disability” would have no impact on a person’s mental health). To compute the results found in the gray row of the table, the value in each row is multiplied by the relative weight for that row and added to the same result in other rows. The row’s value for self-assessment is substituted for “na” wherever it occurs (by coincidence that happens to be 1 in all cases of this example). What is most notable about the values represented is the extreme weight given the attribute of mental health. This causes overall scores to be relatively high despite extremely low ratings in two of the four dimensions in this IMQOL model. Results of standard gamble utility assessment for the same states of health are included on the bottom row of the table for comparison. These values were obtained before the IMQOL method interview. Results were not obtained for the individual’s own state of health with the standard gamble method.

The zero for mental health in the severe disability state deserves a brief explanation. This individual interpreted the description of that state as peculiarly affecting the mental health of a stroke victim. There was ample evidence that had the analyst corrected the patient’s misinterpretation of the description, the individual would have ascribed a value of 1 or “na”. Since this was not the practice of the analyst at this point of the study, the results are presented without alteration. The extreme impact of a poor mental heath rating is nonetheless clear.

Serendipitously, this individual commented after the standard gamble elicitation and before the IMQOL interview, that he suspected the analyst would find his results “quite

skewed.” This middle-aged white American male looked unremarkable for appearance sake. The individual then volunteered the fact that he had just experienced the death of his dearest friend after three years of caring for him in terminal illness. He further revealed that he too was terminally ill. In the course of the IMQOL interview, he revealed that he was forced to remain at home on IV fluids most of each day accounting for his poor rating in the attribute of “Active Life”. His cheery countenance was as puzzling as his IMQOL scores until you consider what the relative weights imply. The IMQOL model suggests that this individual has adapted his view of life to accommodate his misfortune and to appreciate what he has going for him. His extreme positive regard for mental health causes a skew to the quality of life assessment that allows him to regard his own circumstance with high overall ratings. Although this was not explicitly confirmed with any direct query (owing to the formative stage of the research plan at the time), the hypothesis is consistent with the rest of the discussion entertained by the interview.

Traditional standard gamble results did nothing in this case to expose this adaptive view nor to explain the numbers given. In fact, those results did not even support the individual’s suspicion that we would find his answers “skewed”. It would not be out of line to suspect this individual was a victim of HIV/AIDS. The hypothesis suggested by the IMQOL model for quality of life in this case is consistent with the findings reported by Joel Tsevat at the 1997 Meeting of the Society for Medical Decision Making⁵⁸. Tsevat finds that utility assessment in HIV patients indicates a significant number assert their quality of life to be greater after contracting the disease than before.

After I have explained the IMQOL model representation and elicitation details, I will return to this example using details to illustrate the application of IMQOL measurement results.

4.6 Summary

In summary then, this chapter has revealed how discordant responses are not only endemic but symptomatic of the need for a better understanding some individuals. If

widespread use of utility assessment is to be expected at the individual level or accurately employed to represent populations at the social level, we must categorize discordant responses and capture the particular values of those individuals rather than dismiss them. I have alluded to a representation of values that provides an opportunity to do just that. In the next chapter, I will give the details of this representation.

Chapter 5

Patient Perspective Representation

The key to knowledge-based systems in computer science and artificial intelligence is often the representation of knowledge. In this chapter, I will discuss in detail the representation of the value system used in this protocol for individualized utility assessment. The representation forged by traditional utility assessment seems unnatural in its use of a single numeric score for health. Evidence supporting this assertion includes the many qualifying questions people ask, seeking more details and conditioning their responses, when traditional elicitation methods are employed to discover their values. When I ask people how some of their friends or family are alike, as healthier than others, the response is frequently in the form of a list of characteristics. Formally asking such questions of twenty-eight strangers on a Boston subway suggested that most people could think of 5-15 dimensions of healthiness before their ideas were exhausted in a 20-minute interview. I suggest a more natural representation than traditional utility assessment inspired by these heterogeneous lists of responses. Furthermore, I

suggest the need is for a complete list of specific attribute scales that are comprehensive and operational.

5.1 Multiple Factors Contribute to Values of Health

The patient's perspective has many factors that contribute to preferences. These factors may influence the patient's decisions in different ways when considered separately. This is easily demonstrated by asking the patient what matters regarding health. The patient is not likely to give a single answer. Left to his or her own in responding to such open questions, respondents employ a heterogeneous list of dimensions. Most dimensions are categorical but vary in the number of categories making up any dimension. Some dimensions may be binary, e.g., smoking versus non-smoking, while some dimensions may be made up of a few to several levels, e.g., mild/moderate/extreme exercise. Other dimensions are potentially scalar in nature, e.g., days missed from work. Another thing quickly evident from such explorations is diversity in answers. Such natural features of human expression motivate the use of focus groups to discover the salient issues of target populations in many domains inside and outside medicine. This underscores the need for a representation of these values that accommodates the heterogeneous nature of multiple factors that contribute to an individual's view of what healthiness is.

Categories of Constructs Used

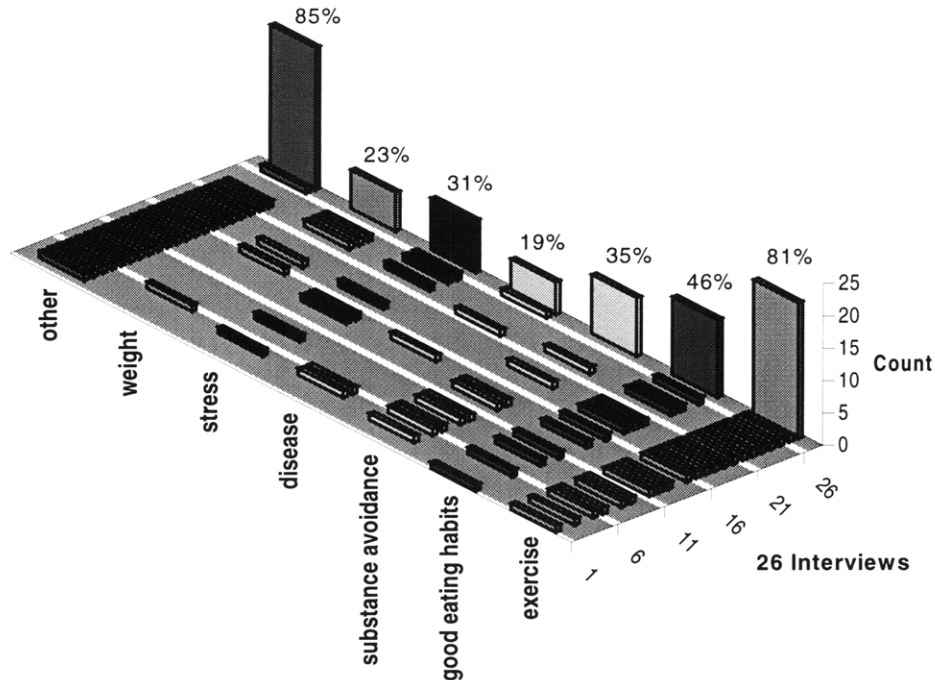


Figure 5.1 Categories of constructs given in interviews.

To discover more about the multiple dimensions people might naturally employ, I asked 28 adults on the Boston subway to think of three people and to tell me how they were alike as healthier than the third. After the interviews, I tried to derive a classification of the topics in all the responses. Not all the responses were direct physical manifestations of disease (symptoms). Not all were physical. I decided the responses could be sorted into exercise, eating habits, substance avoidance (e.g., cigarette smoke, alcohol, or drugs), disease, stress, body weight and “other” categories according to what I regarded as common sense. Distributions of the responses among these groups are visually displayed in Figure 5.1.

The “other” category of responses could be divided into physical and non-physical. The responses that were not physical in nature could be sorted into those that were matters of emotion (e.g., “laughing”/“boredom”, “relaxed”/“worrying about emotions”), matters of

intellectual exercise (“mental well-being”/“lack of mental well-being”, “learning new things”/“apathy”), and matters of will or desire (e.g., “pushing the envelope”/“home watching tv”, “tenacious”/“does not try”). There were responses which were not physical that could not be classified as any of those yet mentioned. These frequently contained direct reference to the “spiritual” realm of a person’s existence (e.g., “faith”/“no-faith”, “spiritually grounded”/“no spiritual dimension”, “peace of mind”/“can’t find peace of mind”, “fulfilled”/“empty inside”). Distributions of the responses among these groups are visually displayed in Figure 5.2.

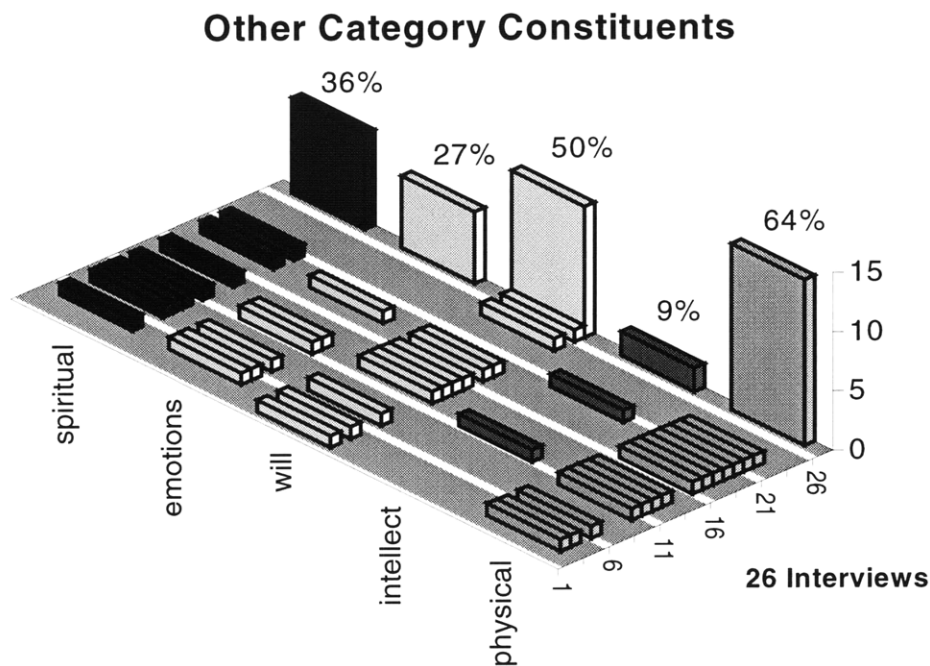


Figure 5.2 Categories of "other" constructs given in interviews

It is easier to understand, looking at the graphed distribution of responses, how difficult it would be to represent everyone with a single model of combined factors. This study did not quantify the relative importance of any dimension versus another, but it was subjectively evident that the relative importance could easily be nearly as diverse as the subject matter of these dimensions. These results also emphasize the importance of non-physical features of healthiness. The observation of no more than 20% of the respondents using disease to indicate healthiness, along with the non-physical topics used, support a conclusion that healthiness

described in strictly biomedical terms would fall far short of adequately representing these patients' views.

A utility value system that would enable us to represent the individual among such diversity without ambiguity must be complete, comprehensive, decomposable, and operational. This list of features is inspired by the similar list provided by Keeney and Raiffa⁵⁹.

It is important in any decision problem that the set of attributes be *complete*, so that it covers all the important aspects of the problem; *operational*, so that it can be meaningfully used in the analysis; *decomposable*, so that aspects of the evaluation process can be simplified by breaking it down into parts; *nonredundant*, so that double counting impacts can be avoided; and *minimal*, so that the problem dimension is kept as small as possible.

I propose a representation of the individual's values which can be described as complete and comprehensive as the result of two inherent dimensions: the length of the list of dimensions and the scope of each dimension. The representation is decomposable by the fact that it is a multi-attribute model. In the next sections, that representation is presented by discussing its properties of completeness, precision, comprehensiveness and operational quality. Following that, I will explain why the latter two properties advocated by Keeney and Raiffa are considered inappropriate in representing the individual.

5.2 A Complete List of Relevant Issues

To fully represent the individual, a full list of relevant features must be included in the representation. I refer to this as the "completeness" of the list of dimensions. To be adequate, the model of healthiness should not leave out any salient issues relevant to health for the represented individual. Patients are sensitive to issues they think are left out. Numerous interviews in the course of this work have included remarks by patients about dimensions they feel are left out of decision making in health care. Right or wrong, the patient is less likely to heed advice that does not understandably address dimensions that he or she deems important.

Some might argue that a complete list is not required if a sufficient list of the most important issues is represented. This could justify a more economic elicitation process by avoiding long interviews to extract the exhaustive list of dimensions of health for each

individual. This may be an efficient approach but the problem with it is where to cut off the list. Patients may not necessarily reveal the most important issues first. There is a risk that some dimension of health that is very important might be missed altogether because it did not become apparent until after a quota of dimensions have been reached.

How does one decide how large to make the quota? If the cutoff is arbitrarily placed at n attributes, the $n+1$ attribute may not be much lower in relative weight than the n th. The cutoff imposed would eliminate factors that contribute nearly as much as those included, which is difficult to justify. A stronger argument still is that any number of dimensions may be of the same relative weight. If there should be only room left for some of the dimensions within the quota, which ones should be included and which ones excluded? No matter which one you exclude, some state of health may involve an attribute which solitarily impacts the dimension left out, leading to a false appraisal of overall value or lack thereof.

5.3 Getting the Right Issue

One might imagine that it is sufficient to ask individuals just to name the relevant issues. The problem with simply naming a dimension is the risk of misunderstanding the intended dimension. George Kelly built an entire theory of psychology, Personal Construct Theory⁶⁰, based upon a perspective that bipolar constructs are more precise descriptions than a list of adjectives or modifiers. Simply stated, the underlying principle is that it is less ambiguous to describe concepts with paired – contrasting denotations than with solitary labels. For instance, when a person asserts a route home is the “long” way, it is unclear whether they are talking in terms of distance or time. If the assertion is in terms of “long as opposed to quick” there is little doubt that the concept is related to time. Kelly postulates that humans construe their worldviews implicitly in such bipolar terms—not in unpaired descriptors. This perspective can be employed in the representation of a person’s values such that the issues embodied are less ambiguous.

When a person asserts that healthy people are recognized by “their enjoyment of life—well enough to do what they want to—living very well,” ambiguity remains as to exactly what the person is referring to. Just how much ambiguity remains is more apparent after the

revelation of what the person considers the contrasting opposite. For instance, asserting the opposite to be “those who cannot do anything for themselves” is describing a different concept than a simple assumption of “not enjoying life” as the opposite. This example is taken from actual interviews conducted in the development of this protocol. If there was any uncertainty as to whether this was truly a bipolar construct and not some coincident elicitation of poorly related terms, the exercises which ensued in the interview verified this as two ends of a single dimension in this person’s perception of health. In my experience asking for the patient’s views, there are often times when the direct negation of a description is exactly what he or she wants to say. There are often times when that is not the case and what the patient thinks is quite distant from such an assumption. In representing the individual accurately and precisely, it is useful to capture not only an attribute label but the contrasting pole as well.

Given the bipolar nature of these relevant dimensions of healthiness, the question remains as to what lies between the poles. As mentioned in section 5.1, the dimension might be a simple binary classification of a feature, an ordered nominal scale of classifications, or a continuous spectrum of some feature. The first case is merely a simple case of the second and the third can always be discretized into the second. In the interviews conducted thus far in my study, people rarely volunteer a concept that is a continuous scale as they describe it. Even those features which might be considered a scale of continuous values by academics are expressed in categorical terms—for instance: “body weight as normal for height”, “+/-10 pounds from normal for height” and “obese” instead of a continuous scale of weight/height ratio. Consequently, I have chosen to represent the dimensions of a person’s healthiness perspective as ordered nominal scales which I refer to as “attribute scales” implicitly assuming they are ordered.

5.4 Comprehensive Attributes

Orthogonal to the completeness of the list of relevant features is the comprehensiveness of each attribute scale. To cover all possible states of health, a scoring system’s scope must stretch between all extremes.

Before discussing the coverage of extremes, it is useful to decide the dimensionality of the space for which the extremes must be known. Bipolar attribute scales presume that there are no attributes of health which have two or more dimensions or at least need not be modeled as such. If some attribute is capable of being described in more than one dimension, it is capable of being decomposed and described as one attribute per dimension. The argument was made in the previous section for the inherent nature of all assertions having an implied counter-quality. George Kelly's method for eliciting personal constructs^{61, 62, 60} succeeds in getting people to articulate their notions strictly in single (bipolar) dimensions. This inspires the feasibility of representing each dimension individually.

To say that we have a comprehensive representation of each dimension, we need only determine that, for each dimension, we know each of two extremes with no holes between. In much of the literature regarding Kelly's Personal Construct Theory, the assumption is that these bipolar constructs are naturally elicited in terms of their polar extremes. If we care not to make such an assumption, it is not problematic to determine extremes, which lie beyond the first descriptors used to identify the bipolar construct. We need only ask directly if there are extremes that the person can imagine beyond the terms used so far. Without discussing elicitation prematurely, suffice it to say that the attribute scales, as single dimensions, are easily comprehensive if a small degree of attention is paid to make sure the poles represent the extremes conceivable by the individual represented.

It is natural to wonder about the ability of individuals to conceive, in one sitting, of all possible health states relevant for them. The question is whether the universe of all health states is to be represented or the set of health states imaginable by a particular person. The representation scheme is the same for either case, but the choice should be made explicitly and adhered to consistently. In the interest of representing the individual, we should be content to represent the scope of health states as those that the individual can imagine at that point in time. The potential for some state of health to fall outside these bounds remains. This should form part of the basis for the distinction between a descriptive and a prescriptive or normative representation of the individual's value system. Those who use IMQOL representation should regard the initial elicitation output as descriptive. Noting the difference between its scope and

what he or she considers normative or prescriptive, the health care provider can respond accordingly by educating the individual. The success of that education can be indicated by the broadened scope of the attribute scales in subsequent descriptive representations of the individual's values.

5.5 Measurable Attributes are Operational

An attribute scale that is only capable of being used by its author is of limited use. Others need to be able to use the scale with or without the help of its author. Great advantage comes if a nurse or physician can apply the attribute scale in the same way as the patient. The health care provider can describe or rate a state of health in terms of the attribute scale and compare it to others states in the same terms. One patient I interviewed believes that healthy people are characterized by “hobbies besides their job” as opposed to “no hobbies”, with only one intermediate level possible, “occasionally doing an activity they enjoy.” Naturally, that person has no problem choosing which level best describes his own state of health—in fact, it was the intermediate. The entire scale came from a consideration of a few friends and family so it is easy to assert that there was little difficulty in describing the state of health represented by each person with these terms. Now, if we ask that person to consider a state of health described with some specific degree of disability after a stroke, there is little difficulty in asserting that state as allowing, limiting or prohibiting hobbies. We could ask the same person to use the same scale to rate other states of health such as hemodialysis or peritoneal dialysis. We may need to accommodate the notion that the attribute scale regards something irrelevant to the obvious features of the state of health; or that the biomedical description of the state bears no impact on the issue of the attribute scale. For example, someone might assert that peritoneal dialysis has no bearing on most hobbies. Therefore, any attribute scale should allow the assertion of “not applicable” or “no change”. The title, “no change” foreshadows the implication of such an assertion. It means that whatever the patient was before the change of state remains the best description of the patient's state after the change. For any given state of health, any understanding user should be able to assert the level of the attribute that best

describes the state. The ability to describe many health states in terms of a particular attribute scale is one aspect that makes it operational.

So far, the example illustrates a scale that is operational for its author. If the scale is equally operational for someone else, that is even better. When the doctor or nurse asserts that a stroke victim with severe disability should be rated as “no hobbies”, he or she is operating with the attribute scale of this patient. The health care provider and patient can assess states of health in the same terms by using the same attribute scales. This allows the two to collaborate in judging the value of such states as outcomes of treatment. A collaborator should know what the terms mean that are used to describe each level of the scale.

The patient and collaborator should be capable of agreeing on which level of the attribute is correct for a given state of health. Any two individuals might have different preconceptions that they bring to bear on their first appraisal of a health state description. Whatever the patient asserts, the collaborator may reasonably be inclined to assert an alternative rating on the same scale. If the patient and collaborator disagree at first, the two parties should be capable of coming to a mutual understanding of an agreeable assertion with discussion or education. Specifically, the patient may have an opinion regarding the state of health that the physician knows to be otherwise—e.g., only people who smoke a lot are unhealthy. The physician is compelled to make it clear that smoking to any degree is damaging to your health. To be operational, the attribute scale need not be expected to elicit the same rating from all users. It is the difference in rating that provides one benefit of this protocol. If the collaborator is using the terminology as the patient intends, but is basing an alternative rating on something not thought of by the patient, the discrepancy in perspectives is made more explicit. What would otherwise be an abstract assertion that one state of health is less healthy becomes an assertion in terms that are more specific.

I have been careful to suggest that operational quality implies only a capability of agreement not necessarily an agreement. The capacity to reach an agreement is limited by ambiguity. If either party has an ambiguous understanding of an attribute scale, they are incapable of using it in judging a specific state of health. So, if the patient uses a term that the physician does not understand, the physician cannot contribute an opinion. This is bad only if

the health care provider is not allowed to ask questions stimulated by the attribute scale terminology. Knowing the words used to describe a poorly understood attribute is at least a step towards understanding. If the physician does not understand what the patient means by “hobbies besides job,” he or she may ask what “hobbies” means rather than some more general query about why one state of health is more preferred than another. If the patient and collaborator agree to disagree, the basis of the disagreement is more explicit in these terms than when expressed in broader comparisons of health states. Ideally, the assertion of which level best fits without ambiguity should be possible without further explanation. However, if there is any need of explanation or clarification, the details necessary for resolution are closer at hand with this exposure. The operational requirement is satisfied if the collaborator’s use of the terminology is acceptable to the patient even if the conclusion differs.

If the attribute scale is capable of use by others in the absence of its author, the values may be used to influence decisions on that person’s behalf in his or her absence. The hallmark of success as an operational attribute scale is the capacity to use the scale independent of the person who authored it. Until the validity and reliability of this representation is well established, it may be safer to employ the attribute scale in the presence of its author. This would result in a collaboration in which the person whose values are represented defines the terms of discourse, which still represents an advance for incorporating individualized patient values. Nonetheless, the attribute scale needs an operational quality for collaborative use as well as independent use.

Finally, it is the operational quality of the attribute scale that enables the methods used to assess the quantitative values with this proposed protocol. This is both a useful feature and a strong indicator. The structure of the elicitation process helps to shape the attribute with an operational quality. As the patient is granted the opportunity to see the early application of the attribute scale to sample states of health and invited to verify its use as intended, more confidence is invited in the reliable representation of the individual’s values. The endorsement, by the author of the principle, for a pupil’s application is a powerful statement that a functional understanding has been acquired. So, the final mark of operational quality is the ability to teach

someone else how to use the attribute scale resulting in skills that meet with the teacher's approval.

5.6 Redundancy?

Keeny and Raiffa⁵⁷ suggest non-redundancy as a quality necessary to avoid ambiguity a utility value system. I suggest that the accurate representation of a person's individual value systems would require that we NOT represent them with a non-redundant model in any circumstance where they are redundant. It may be true that a utility value system that contains redundancy risks more ambiguity, but an individual's values are not guaranteed to be free of redundancy. For any person who is redundant in his or her value system, representing that person's values with a non-redundant value system is a misrepresentation. By selecting not to require a non-redundant representation, I have made the implicit statement that the accuracy of the descriptive nature of the representation is more important than the inherent ambiguity. To minimize the risk of ambiguity, non-redundant attributes should be the guide, but not a necessity. Furthermore, as a descriptive representation, the job of the representation is to illuminate the ambiguity within the person's value system, not conceal it.

Suppose the tree surgeon values both an active recreational life and a robust career of tree pruning. Should we discount one of these values because they both share the requirement for a functioning vestibular organ? Both of these attributes are highly dependent upon good balance. The tree surgeon does not likely discount one value because of the redundancy. So we are likely to introduce error in a descriptive model if we do not capture both recreational life and work.

5.7 Minimal Dimensions?

Keeny and Raiffa⁵⁷ also suggest minimized dimensions as a quality necessary to avoid ambiguity in a utility value system. Similar to the issue of redundancy, the accurate representation of an individual who does not minimize the dimensionality of their perspective is misrepresented by any representation that does minimize dimensions. As with redundancy, the relaxation of a requirement for minimal dimensions is an implicit statement that the

accuracy in representing the individual is more important. Again, the principle should be a guideline but not a necessity. Not all persons may be capable of understanding the principle making it difficult to train them to be otherwise. However, recognizing when they have many dimensions is important in order not to neglect anything of relevance. This reflects back to the completeness of the representation that was established in Section 5.2.

Building a model with too many dimensions is like building a linear regression with too many factors (terms) in the equation. After so many additional terms, the model is said to "over fit" the data, which is to say the predictive performance on novel samples degenerates as additional terms are added beyond some empirically optimal number. This degeneration of accuracy is what I understand to be the ambiguity that would result when the number of dimensions is not minimized. So why not minimize the dimensions of an IMQOL model even if the patient does not do so in reasoning? First, it is somewhat hard to imagine the occurrence of over fitting with the small number of dimensions people generally come up with for what makes up healthiness. Second, as already mentioned, a descriptive model needs to capture the errant ways the patient thinks, not conceal them. Finally, as stated in the explanation of completeness, the difficulty is in choosing which attribute to discard. Nonetheless, it will be interesting, in future research, to see how often over fitting occurs and under what conditions.

5.8 Recomposing the Whole from the Parts

With individual dimensions of a multidimensional model identified, the question remains as to how those dimensions should be mixed for composite scoring of any state of health. If the application of the system of values is to be limited to only consideration of separate dimensions, there is no need to worry about the relationship between dimensions. This would be the case if, for instance, one were to think of a state of health only in terms of some profile of completely independent issues. Such a perspective does not, however, fulfill a need to judge the relative merits of two outcomes that excel on different dimensions. To address tradeoffs, a relationship between dimensions must be represented that makes sense and more or less captures the dynamics of the value system.

The choice of model for this multidimensional system is not trivial. Keeny and Raiffa⁵⁷ have explored the theory of multi-attribute models for personal preferences. Extensions of this work are numerous. Assuming a multi-attribute nature of preference models, Michael Wellman developed ways to construct, analyze and reason with them³³. What has remained unclear is how to determine which dimensions to select, include and exclude for any specified purpose or task. It is the ambition of this thesis to fill this gap by assuming the multidimensionality of the values which impact patient preferences and developing an elicitation suited to that assumption without making an immutable commitment to the nature of the relationship between the dimensions. It will suffice, at this point, to be able to identify the dimensions and a simple linear model. More complex models of the system can be explored in light of well developed utility theory, but evidently not until some systematic way to arrive at a determination of the salient dimensions is developed.

One might contend that experts and/or analysts need only decide what the issues are—by empirical evidence, analysis, or omniscient decree—and assess the targets with established psychometric tools. This is the perspective of the Medical Outcomes Study Short Form-36 Health Status Measure (SF-36) as originally authored⁶³. Based on the expertise gained by the Rand Corporation of Santa Monica during the 1970's⁶⁴⁻⁶⁶ and results of a 20 question precursor published in 1988⁶⁷, the Short Form 36 is a response to criticisms that a few issues were left out in the earlier version⁶⁸. The result of this applied expertise is a model composed of eight dimensions: physical functioning; role limitations due to physical health problems; bodily pain; social functioning; general mental health covering psychological distress and well-being; role limitations due to emotional problems; vitality energy or fatigue; and general health perceptions. A twenty-year history of considerable investigation has produced a good measure of health status. It provides a measure by which the state of health of many people can be compared in holistic terms. Those terms are, however, decidedly biomedical and, by necessity, derived from widely held views, not necessarily used by all individuals. The results of this status measure do not tell us what patients prefer. They do not suggest any basis for how the patient's values resulted in the status indicated. For example, neither a solitary SF-36 result nor a declining result in one or more dimensions over multiple assessments tells the analyst

whether the patient *wants* to exercise more or not. The SF-36 does not help resolve whether a patient prefers to avoid long term disability or any increased risk of death in the choice of treatments for stroke. It does not help resolve whether a patient prefers hemodialysis or peritoneal dialysis for treatment of kidney failure. It does not reveal the fact that a tree surgeon must regard threats to his sense of balance differently than an office executive. The shortcoming for representing patient preferences lies mostly in its inflexibility. The choice of dimensions is fixed no matter who it is used for. Fixing the dimensions affords the ability to compare what is measured across populations and to derive statistically or analytically the relationships between dimensions. However, if we are willing to forsake the benefit of a fixed set of dimensions, we gain the ability to represent diverse people with more accuracy. We may also derive the relationship between dimensions for each person recognizing the unique combination of weights any individual might place on the dimensions.

The model by which the patient composes the individual attributes into an overall judgment of healthiness is a critical step in value system determination. If we are willing to employ a linear model for composition, we need only determine the scaling coefficients for each attribute in the equation. A linear model serves as a baseline for ways we might combine the dimensions of multidimensional preference models to generate composite scores for health states. It is an attractive model because the linear model is easy to manage, it has precedent, it has credibility⁶⁹ and it is supported by a perspective of factor analysis popular in psychometrics. An added benefit of the factor analytic perspective is the abundance of knowledge about factor analysis that may be called upon for thinking about multidimensional patient preferences and the familiarity many people have with that way of thinking. Those inclined to complain of the oversimplification in such models should be reminded that all models are an approximation and that significant improvement in outcome estimation must be demonstrated to warrant the increased complexity of non-linear models. There is no argument to be made against the claims that interdependence is highly likely between the dimensions of a multidimensional health preference model. I would argue for the choice of a linear model with an opposite approach. I would argue that the data points we are talking about are a cloud, if you will, of neighboring data points in some dimensional space, all inter-related—a cloud that we

elect to approximate with a line for simplicity sake. That is fundamentally what the perspective of factor analysis is.

The opportunity exists to explore the consideration of utility independence and preference independence at this juncture³³, but the value of that complexity is not yet clear. A more complex model might be multiplicative, accounting for second, third or higher order interactions. It might be a hybridization of linear (additive) components and multiplicative components arranged in some hierarchical structure. Any increase in complexity, however, comes with a cost. More complex models will require more steps to determine the nature of the model. This translates into more questions, making the elicitation process longer. It also results in more complex questions. This may challenge the patient's attention span and willingness to complete the interview process of elicitation. It may also challenge the attention span of the health care providers who must retain enough interest in the underlying theory to accept and actually use it. With the linear model, we do have a feasible starting point. The scaling coefficients can be determined by using the descriptors authored by the patient to describe hypothetical yet "familiar" extremes in combination for standard utility assessment methods. How this is done will become clearer when I discuss the methodology of elicitation in a latter section. No specific health state is required. The coefficients can be determined using hypothetical states of health using only the terminology generated as elements of the patient's attribute scales.

The formalism I propose is systematic enough to suggest successful implementation with computer aided self administration. This raises the possibility of more familiar and leisurely environments for administration than disconcerting physician offices and hospitals. It also allows for repetition, which fosters familiarity with the instrument. All of this increases reliability judging by more traditional applications of utility assessment. If the individualized patient values elicitation involves significant investment of time, it pays back that investment in reusability. The absence of dependency on specific health states frees the elicited perspective to application across decisions without limitation as long as the patient remains stable in his or her perspective. The freedom of the administration also allows more frequent administration should a change of perspective be suspected (e.g., patient education, change in health state). This

provides greater opportunity for longitudinal studies, i.e., assessment can be done with more ease and greater frequency.

In my final support of the linear model for a composite score, I would like to raise the issue of whether most persons think in terms of linear models or ones that are more complex. In so far as we are emulating human intelligence and not behavior, we might wonder how human intelligence works. Disputing the non-linearity of observed behavior would be ill advised. However, if the human mind works in non-linear models, it should follow that such models would be easy to understand and explain. Getting most people to understand non-linear models, much less hybrid linear/non-linear models with more than second order interactions is difficult. Asking them to explain their non-linear values is worse. Perhaps the best we should expect is to be able to coerce an understanding of a linear approximation of more complex behavior if it gets us close to where we need to go.

These are all questions that may be addressed by investigations that follow the establishment of feasibility for the simplest of models. A linear model gives us a baseline from which to start so I choose to begin there in representing multidimensional value systems.

5.9 Descriptive Versus Normative or Prescriptive Representation

It is worth underscoring that this representation is descriptive. No claim is made that it is or should be normative. After we first manage to describe the patient's view then we may explore ways in which that differs from what experts think would be more normative. Only after we have explicitly made the differences between such descriptions and "normal" can we focus prescriptive efforts with optimism.

If this representation is descriptive, it is subject to more volatility in the sentiments of the person represented. My ambition is to develop a protocol that is easy enough to use that it enables re-assessment at a rate that can keep up with the volatility of people's values. It would be reasonable to assume that the values a person has in regard to life long or life threatening decisions are to be based on life long history and experience. This would be expected to stabilize those actual values we are trying to capture. Nonetheless, we are unreasonable if we presume a person's values will not change. I think of the value system used for medical

decision making to be like that used for writing one's legal will. Such documents are subject to change at a modest rate, a rate that would be easy to accommodate with a stand-alone computer program.

5.10 Labile Values

The value system must be allowed to change with the evolving perspective of the patient. What can be said about the values of a naïve person unaware of some looming threat? Will the values represented be a faulted representation for use in evaluating a procedure that may result in an outcome with features of importance that are not represented as such? It is very much like laboratory results for serum chemistry assays (e.g., glucose, Blood Urea Nitrogen, liver enzyme levels). A health care provider would be remiss to neglect immediate sampling because he/she has results from some months ago. Similarly, the provider would be remiss to proceed on the basis of an old value system information without inquiring whether the patient has reviewed or been given reason to review his or her value system. The IMQOL model represents values held at a single point in time. Just how labile these values are is an interesting question. I will return to this issue of labile responses in the chapter on Application of Elicited Results integrating a plausible explanation for such labile responses over short intervals as an expression of poorly differentiated values (preferences) put forth by Shiell, et al⁷⁰.

5.11 Summary

I have described then an individualized multidimensional quality of life model, linear for now, which is complete as a list of relevant issues, each one being a comprehensive ordered nominal scale of mutually exclusive categories for each given dimension. These attribute scales are fully specified by the terms of the patient in words that others can use with some degree of approval. These scales in combination should be capable of scoring any described health state if those judging can come to agreement on the most fitting level of each attribute scale that describes the scale's aspect of the described health state.

Chapter 6

Elicitation of Patient Perspective

Having presented, in the previous chapter, the chosen representation of values to be used in comparing states of health, the task of eliciting information that conforms to the IMQOL representation scheme can be described. I will refer to this elicitation process as the IMQOL method for the remainder of the thesis. The task can be subdivided into three subtasks that correspond to the structure of the values representation. The first is the acquisition of a complete list of relevant dimensions, their extremes and intermediate levels, leaving no gaps in order to fulfill the comprehensive requirement of each attribute scale. Adopting, by convention, a range from zero to one to numerically represent the range of most detrimental to most beneficial level for each scale, the second subtask is the determination of numeric value equivalents for each of any intermediate levels of all the attribute scales. Finally, the relative weights for each attribute scale must be determined for composite scoring of health states.

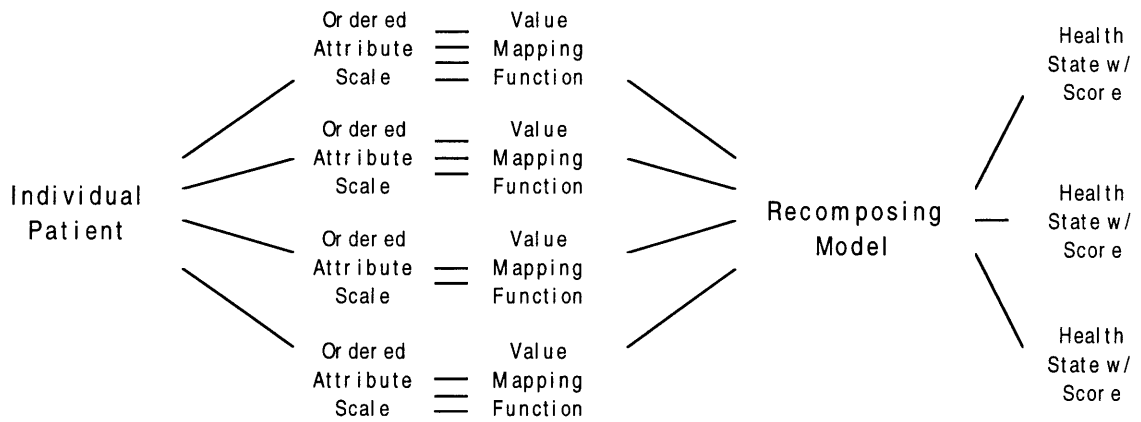


Figure 6.1. Overview of decomposition and re-composition of patient values. The individual's perspective of healthiness is composed of one or more attribute scales, each scale having a number of possible levels. Each level has a corresponding value on a numeric scale from 0 to 1. Knowing the level for a given health state for each attribute scale and the model for recombining the attribute scale scores in a composite score allows the use of the individual's perspective to score that health state.

6.1 Determination of Attributes

To elicit a list of dimensions representing a complete registration of relevant issues from the patient's point of view, a structured interview is used to stimulate the patient in revealing words and concepts used to discriminate healthy from unhealthy friends and family. Actual people are used to provide concrete ideas. Familiar people are used to capitalize on the experience and natural thoughts of the patient. Identifying features seen in people other than the patient, help to elicit idioms that are more observable and less abstract. This predisposes the results to more likely have the necessary operational quality. The result of this step in the process is a collection of bipolar concepts

Rules regarding the formulation of subsequent questions provide structure to the interview. These rules are designed to preserve the absence of bias while focusing the conversation on the desired elicitation. An informal presentation of those rules follows in a description of the interview details. Graphic visual aids facilitate the understanding of the questions throughout the interview. The first pair of questions is the same for all participants. Each subsequent question is formulated from the previous responses.

The interview begins with asking the subject to think of three friends or family members specified by a role or relationship, e.g., father, best-friend, best-friend's father. Three roles are presented for the subject to fill with someone they know. Names are not required but the age of the role-filler is recorded for clarity and to facilitate any necessary reference during subsequent discussion. Once this introduction is completed the first question follows. That question is, "With those three persons in mind, tell me an important way in which two of them are alike as more healthy than the third. If it is easier, you may choose to tell me how two are alike as less healthy than the third." Each separate response is regarded as one end of an attribute scale. It represents one dimension by which the patient discriminates healthiness from unhealthiness. For each such response, the person is asked, "How, then, is the third less healthy [or healthier] than the other two?" This provides the 'first draft' of a bipolar attribute. The same triad of roles may be employed for multiple bipolar attributes or changed if needed to solicit fresh ideas. The interview continues until no more ideas come to mind. The patient is encouraged with this question to think of any attributes not represented in the specific people brought up in the interview. One of four sets of roles are used to focus the subject on three familiar people with whom they may have enough acquaintance to hold some opinion of their healthiness. Convenient substitution is allowed. The specific role filled is not so important as that three distinct individuals of satisfactory familiarity are fixed in mind. Example sets of roles used are as follows:

Your father or nearest male relative with whom you have most frequent contact
Your best male friend
Your best male friend's father

Your mother or nearest female relative with whom you have most frequent contact
Your best female friend
Your best female friend's mother

Your boss or employment supervisor
Your most frequent customer or client
Your friendliest business associate

Your nearest neighbor in any direction from your home (someone you are familiar with)
Your nearest neighbor in the opposite direction (again, someone you are familiar with)
Your most frequent neighborhood visitor who does not live nearby

Additional properties of the attribute scale are solicited to verify understanding and for purposes of validating comparability with other responses from the same person or other interviews. Providers interested in changing the patient's behavioral risk factors will have an interest in certain properties that these bipolar attributes possess. These properties might include activity, origin, control, causality, implication and/or measurability. These properties may be solicited with the following sorts of questions. Is the concept of the attribute scale one that requires energy or happens without energy required (activity)? Is the nature of the attribute scale something that comes from within a person or something that comes from somewhere outside the individual (origin)? Is the concept of the attribute scale something that is within the individual's control or something that is beyond the control of the individual (control)? Does the concept of the attribute scale cause healthiness/unhealthiness or is it a result of being healthy/unhealthy; both or neither (causality)? Measurability is determined by the interview administrator or self reporting such that if the response is deemed immeasurable, the patient is asked for explanatory subordinate concepts with a question framed from the response, e.g.,- "What kind of a person is one who is indicated healthy by [response substituted]?" To minimize the length of the interview, results of previous interviews could be employed. If the response is like one heard previously within the same interview or outside interviews, the interview may simply ask for verification of the properties as registered in previous experience as a less taxing elicitation.

It is worth noting that the chosen wording of the questions asks for the means by which the person discriminates healthy from unhealthy friends and family and is far from arbitrary. It capitalizes upon the experience of those who have developed the repertory grid elicitation process ^{71, 72}. Questions are carefully fabricated to avoid the ambiguity of feelings about healthiness. They instead focus on the characteristics by which persons discriminate healthy from unhealthy. Before the interview is completed, they will be asked to evaluate their willingness to trade one state of health for another; each described solely in these personal terms. If the dimensions prove inadequate for such an exercise, the opportunity to revise remains.

Each session is ended with a final query: "Now that you know the sort of thing we are looking for, are there any other ways in which you discriminate between healthy and unhealthy people regardless of any particular person we have brought to mind?" The patient is encouraged with this question to think of any attributes not represented in the specific people brought up in the interview.

The informal rules of the interview are then:

1. Refer to the "healthy" or "unhealthy" root attribute as an anchor for each question.
2. Each new elicitation begins with "How are two alike as more[less] healthy than the third?"
3. Each item elicited must be paired with a contrasting partner term by asking, "...how then is the third less[more] healthy than the other two?"
4. End the query when the subject cannot think of any more means by which he or she might discriminate the individuals of the role triad based upon their healthiness.
5. The subject will be asked to express his or her personal standing in regard to each elicited attribute as a demonstration that the interviewer properly understands the bipolar attribute as he or she intended it (discussed below).

6.2 Determination of intermediate level values for attribute scales

The next step of the values elicitation process is to solicit intermediate levels the attribute scale may potentially have. The interviewed patient is asked if he or she can think of any person who does not belong to either extreme of the scale but somewhere in between. If not, the assumption is that no intermediate levels exist in the perspective of the patient. If a person is brought to mind, specifically or categorically, the patient is asked for a categorical label for such a person. This query is repeated for each bisection created until possibilities are exhausted. The result is a series of nominal categories that are ordered by nature and constitute all plausible levels of the attribute in the patient's perspective. The patient can be asked directly if each extreme, in fact, covers the most extreme cases that can be imagined. Attribute scales from previous interviews may be less taxing to verify or modify than built from scratch

and therefore future developments of the methodology should include some utilization of past responses without undue bias.

To further validate the understanding, the interview involves formulating a question from each attribute scale elicited. The purpose of this step is to test the interviewer's understanding. If we can agreeably phrase a question with their bipolar attribute, higher confidence is associated with the understanding. The question to be formulated is one that asks the subject to rate his or her own personal standing on the attribute scale. "You seem to indicate that a person who is <first elicited pole> is healthy and one who is <contrast pole> is not so healthy. Do you see yourself as healthy because you < first elicited pole> or as unhealthy because you are <contrast pole>?" For scales with more than two levels, the question should be framed to include all levels as potential responses. Explanation is to be offered to the subject of the fact that the answer does not matter as much as the question. It is simply an opportunity for the subject to recognize any misunderstanding and provide opportunity for restatement or clarification before approval.

Care is required when articulating the question for this verification step. If the question is not carefully worded, it is easy to raise an objection. Some individuals balk at the notion that just because a person is poor in the specific attribute of the question that does not mean the person is unhealthy. Some of these respondents can be put to ease by explaining the analogy of a grade school report card. Pointing out how a student who has low grades in only one or two subjects can still have a good overall grade can be very helpful. Most individuals who take to the multi-dimensional model well are more forgiving on this detail proceeding without hesitation. The problem seems to occur most easily if the leading statement is worded, "a person who is <first elicited pole> is healthy and one who is <contrast pole> is not healthy." Taken literally, the objecting people have a valid point. If, instead, the statement is worded "a person who is <first elicited pole> is *healthier* [or *more healthy*] and one who is <contrast pole> is *less healthy*" the problem can usually be avoided.

The completed attribute scale is then assessed to determine the utility values for each level. Utility assessment techniques of Keeney and Raiffa⁵⁷ are employed to determine utility values for attribute scales with inherent cardinality. The ordered attribute scales with no

inherent cardinality have two or more discrete levels. For scales with only two levels, the extremes of 0 and 1 are assumed for the worst and best ends of the attribute scale respectively and the elicitation of the attribute scale is complete. For scales with more than two levels, the interview proceeds with a modification of the standard gamble utility assessment. The patient is told to hypothetically assume he or she exhibit the level of attribute being assessed and offered a hypothetical potion that promises to change them to the best possible level of the attribute but at some risk of ending up in the worst level of the attribute. The specific risk is systematically varied and, as in the earlier description, the probability at which the patient is indifferent or cannot decide between options is the utility value that is assigned to that attribute level. Alternatively, a visual analog rating scale or time trade-off method could be used to determine the utility value of intermediate levels relative to the extremes of the dimension of health represented in each attribute scale. This may be required in cases where severe risk aversion is evident. In any case, the result is a mapping function, which, if given the name of the level for any of the individual's attribute scales, provides a numeric value between 0 and 1 for that attribute of health.

6.3 Determination of a combining model for health state scoring.

To determine the relative weights of each attribute scale for a composite scoring of health states under the linear model assumptions, the patient is asked the following series of questions. Until they catch on to the paradigm, the scenario is introduced in small increments of complexity confirming the patient's understanding at each level. First, the patient is asked to assume that they are currently in a hypothetical health state described completely by one level from each of his or her elicited attribute scales. The attribute of most importance is fixed at its lowest level with one of the other attributes at its highest level. Before proceeding, the patient is asked if he or she would be interested in a treatment that improved his or her level of the most important attribute from the worst level (pre-treatment) to its highest level (post-treatment) with no side-effects or strings attached. Based on everything asserted so far, the patient should not refuse. A negative or indifferent response indicates something is amiss.

After indicating he or she is following along, the patient is asked if he or she would still be interested in the treatment if it had one side effect. The side effect is the loss of the one high level of the attribute in which he or she excelled in the pre-treatment state of health. This is the introduction of a tradeoff between the issues represented in the two attribute scales. The patient is asked to respond at this point with a “yes” or “no” reply. A “no” is then taken to mean that the two attributes are of equal importance and this can be confirmed by a direct query. In that case, the relation between the attributes is known. In event of a “yes,” a time tradeoff framework is employed to determine the relationship between the change of one attribute and the change in the second attribute. The incremental approach allows the patient to anchor the tradeoff between two attributes by making two decisions before the introduction of any change in life span.

When the patient indicates a willingness to trade a high rating in one attribute for a high rating in a more important attribute, he or she is then asked to assume a life expectancy of a given number of years in that state, e.g., 10 years. They are asked if they would trade that outcome for some number of years in a different health state described in the same manner, e.g., 5 years in a state with a higher level of one attribute. The number of years in the second state is varied until the point is discovered where the preference changes indicating equal value for the two hypothetical states of health endured for different amounts of time. Similar equivalencies are sought until a differential in time is scored for a change in each attribute scale granting the circumstance where the coefficients may be computed by solving the simultaneous equations.

To illustrate the method, imagine the oversimplification where a patient measures health by only three two level attribute scales: exercise (some or none), dietary restrictions (none or any), and presence of any disease (true or false). The patient is asked to imagine that he or she will have a life expectancy of 10 years as an individual whose state is best described as $S_1 = \{\text{renal disease, protein restricted diet, and no exercise}\}$. Further the patient is asked if he or she would trade that state of health for one with 5 years life expectancy described as $S_2 = \{\text{no disease, protein restricted diet, and exercise}\}$. For illustration, assume it is determined that the patient is ambivalent about the choice between 10 years in the first (S_1) and 7 years in the

second (S_2) state of health. The patient is also asked similarly for the number of years he or she would trade for in a third state S_3 ={no disease, no dietary restrictions, and no exercise}; let's say it is 9 years. With the values known for the levels of each attribute (1.0 or 0.0 in this case) only the scaling coefficients of the linear model are unknown. Using the method of simultaneous equations, we can solve for each scaling coefficient knowing that the sum of the coefficients themselves must sum to one.

$$utility(S_1) = \lambda_1 Attribute_1 + \lambda_2 Attribute_2 + \lambda_3 Attribute_3 = \lambda_1(0.0) + \lambda_2(0.0) + \lambda_3(1.0)$$

$$utility(S_2) = \lambda_1 Attribute_1 + \lambda_2 Attribute_2 + \lambda_3 Attribute_3 = \lambda_1(1.0) + \lambda_2(0.0) + \lambda_3(1.0)$$

$$utility(S_3) = \lambda_1 Attribute_1 + \lambda_2 Attribute_2 + \lambda_3 Attribute_3 = \lambda_1(0.0) + \lambda_2(1.0) + \lambda_3(0.0)$$

and

$$utility(S_1) 10yr = utility(S_2) 7yr = utility(S_3) 9yr$$

therefore

$$[\lambda_1(0.0) + \lambda_2(0.0) + \lambda_3(1.0)]10yr = [\lambda_1(1.0) + \lambda_2(0.0) + \lambda_3(1.0)]7yr = [\lambda_1(0.0) + \lambda_2(1.0) + \lambda_3(0.0)]9yr$$

means

$$\lambda_1 = 0.587, \lambda_2 = 0.217, \text{ and } \lambda_3 = 0.196$$

This means that the expected utility for this patient for any given health state is given by the formula:

$$utility(\text{given state of health}) = 0.587 \text{ Disease Attribute Level} + 0.217 \text{ Dietary Attribute Level} + 0.196 \text{ Exercise Attribute Level}$$

An example of the patient perspective that results from this elicitation is shown in Figure 6.2. The results are from an actual interview conducted by telephone with a patient who had recently visited the Emergency Department of a hospital. Four dimensions or "attributes" make up this individual's perspective or "abstraction" of healthiness. The words used by the patient have been used to create an ordered nominal scale of categories for each attribute. These are found under the heading, "Ordered Attribute Scales". Each scale is mapped to a numeric scale ranging from 0 for the least healthy to 1 for the most healthy. The mapping is the result of intermediate level value determinations of Section 6.2 and is displayed in the rightmost column. The results of elicitation described in Section 6.3 give the relative weights for each attribute scale to overall healthiness and are found in the second column.

Scale ID	Relative Weight	Ordered Attribute Scales	Value Relative To Attribute Extremes
i	.036	no chronic illness or disability	1
		temporarily ill	0.999999999
		minor health concerns	0.999999
		probability of becoming chronically ill but not yet chronic illness **	0.3
			0
ii	.0024	active life	1
		hypochondriac	.53-.54
		somewhat active or limited	0.4
		forced sedentary life **	0
iii	.96	mental well being **	1
		mentally ill	0.25
		mental decay-marked decline in level of mental functioning	0
iv	.00077	non-obsessive, Non-abusive **	1
		occasional lapses: into heavy drinking, fluctuating weight, yoyo dieter, binge drinking, recreational drug use(no addiction)	0.95
		obsessive personality, drugs or alcohol, self destructive	0

Figure 6.2. Ordered attribute scales of one individual interviewed. All levels of each scale are listed with their corresponding value relative to the scale extremes to the right. The four attribute scales for this patient are arbitrarily numbered on the left. The level which this individual claims best describes his current state of health is marked with two asterisks, “”.**

6.4 Rationale for the Protocol

The individual is directed to use concrete familiar people as the stimulus for revealing his or her terms for health. The request for similarities and likeness make the results more operational. Observed features in comparative terms are more likely to work well in the hands of others than abstract descriptors. The comparative nature is reminiscent of the work by Patrick Winston that formalized learning from examples by analyzing differences and explaining experience ⁷³⁻⁷⁵. The questions asked solicit a positive example and a negative example for the attribute being articulated that result in the bipolar construct. The individual is asked to reveal these ideas from familiar acquaintances. The protocol treats the individual as an

expert on his or her opinion. In keeping with Winston's approach to learning, the individual is treated like a teacher by this protocol and the health care provider like a student learning the values of the patient. Capitalizing on the things with which the individual is familiar is capitalizing on the actual experience of this "teacher". This is naturally more reliable for understanding values and decision behavior because so much of what people choose is based upon their experience.

Every effort should be made to avoid putting words in the individual's mouth if the goal is to truly capture the individual's way of understanding things. There is a tension, however, between this ambition and that of helping the individual who is at a loss for words. Commonly to encounter an individual who obviously knows he or she have something to say but has difficulty articulating the thought; he or she might even explicitly say so in conversation. The trade-off then becomes whether to bias the response with suggestions to avoid omitting any salient attributes or to remain pure and strict in unbiased elicitation at the risk of omission for lack of the right words.

6.4.1 Alternatives Considered for Eliciting Scaling Coefficients

Rating scales might be suggested as a simpler alternative. They certainly are less confusing and easier to administrate. However, rating scales with ambiguous endpoints are inadequate. One might think that asking a respondent to simply mark a spot on a ruled line between the worst and best of multiple attributes would be sufficient to determine the relative importance of each attribute. Let's consider this idea with a clear example and then illustrate why it is inadequate with a very familiar circumstance. Say we determine that a person has four dimensions to his or her perspective of what adds up to healthiness as in Figure 6.1. We want to know what relative importance this person gives to each of the four dimensions, i.e., we want to determine the right values for the second column. Why can't we simply ask the person to indicate how important they think each attribute is on an arbitrary scale from low to high as in Figure 6.3?

On the following scale, make a mark between Low and High that indicates

how important the attribute is in your view of what healthiness is for you:

Chronic Illness	Low	-----+	-----+	-----+	-----+	-----	High
Active/Sedentary	Low	-----+	-----+	-----+	-----+	-----	High
Mental Well Being	Low	-----+	-----+	-----+	-----+	-----	High
Non-obsessive/Obsessive	Low	-----+	-----+	-----+	-----+	-----	High

Figure 6.3 A Short Sited Rating Scale Method for Relative Weights Between Attributes of Health.

To understand why this is inadequate, consider renting a car at an airport and choosing to pay the rental company to fill the gas tank upon return at a discounted gas price. The goal is to return the car with the tank as empty as possible to minimize the expense for fuel. The car is delivered to you with a full tank. Assume the tank has a capacity of exactly 10 gallons and the gauge reads full as in Figure 6.4A.

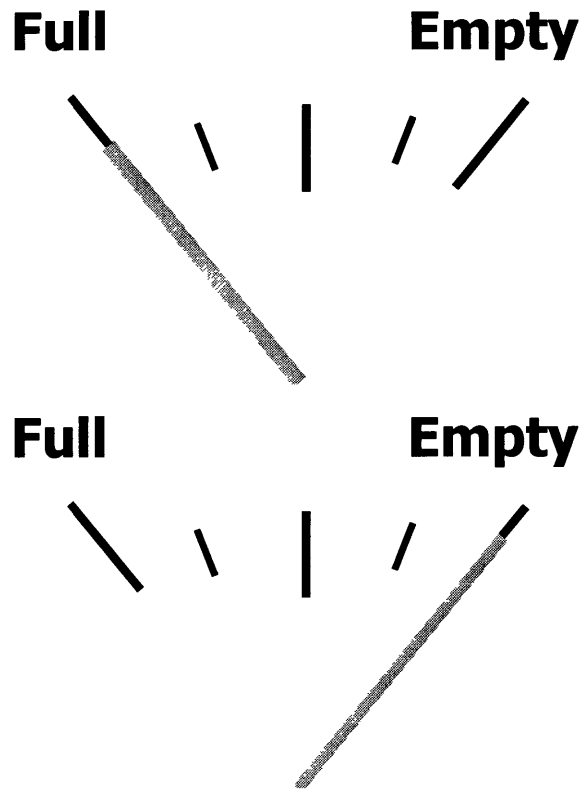


Figure 6.4 A -Fuel gauge when car is rented. B -Fuel gauge after 100 miles.

Assume you drive the car until the gauge reads empty as in Figure 6.4B after traveling 100 miles. If you are on your way back to the airport 50 miles away, will you meet the goal by filling up the tank to the halfway mark? The answer is “no” if “Empty” on this gauge does not accurately represent 0 gallons in the gas tank. You may assume that the halfway mark on the gauge accurately indicates 50% of the gas tank. The answer is still “no”. Ask a few drivers how far they can drive their car when the fuel gauge says empty. The response will vary and someone will say they do not know because they have never gone far enough to run out of gas to find out. If “Empty” actually indicates 0 gallons and “Full” actually indicates 10 gallons, the answer would be clearly “yes”; the car would be returned with an empty tank. If “Empty” actually means there is still a gallon in the gas tank, then the halfway mark actually indicates 4.5 gallons and you will likely return the car with a gallon to spare after 50 more miles. The ratios indicated on the gauge depend upon the absolute reference of the terms displayed on the gauge.

Similarly, if we are to determine the scaling coefficients of the multi-attributed linear preference model with an ordered nominal scale, we will get faulty results if we do not know the absolute values for the terms used to indicate the upper and lower extremes of the scale. The requisite then is a scale that is universal enough for the semantics to remain consistent when negotiating tradeoffs with diverse topics and granular enough to enable the discovery of an indifference point.

“Willingness to pay” is another documented method of utility assessment used⁷⁶, but it has some psychometric difficulties—the most grievous is limited construct validity because of inconsistent value for the “dollar.” A dollar in the hand of a rich person is not the same as a dollar in the hand of a pauper. It is likely that it would be consistent for use within the same individual but difficult in comparing results of different individuals. A novel measure, such as hours at the beach, minutes of exercise, or days of vacation might be employed, but the tradeoff with novelty is the unknown track record. The value of the beach, exercise or days of vacation may vary as much or more as the “dollar”, to say nothing of the semantics. Life span used as a currency by the time tradeoff method has promise of more universal semantics and value. There is a lot more agreement about the meaning of birth and death as well as a more universal distribution of life span among the members of the human race. More pragmatically, there exists a great deal of study of the time tradeoff technique for utility assessment granting a head start on psychometric validity as well as familiarity in the medical domain.

Keeney and Raiffa⁵⁷ describe the theoretical ways to elicit scaling coefficients for both single and multiattribute models that are additive, multiplicative or hierarchical hybrids of additive and multiplicative preference structures. These methods are largely dependent upon what they call “lotteries” that are basically a representation of outcomes with risk, the fundamental option-with-risk structure employed by the standard gamble utility assessment method. The lottery is posed with varying probability of a favorable outcome to determine the point of indifferent preference between the certainty of an intermediate outcome and the advantage qualified by risk in the lottery. To discover a point of indifference for a multiattributed preference model requires a fine granularity be shared by any two attributes

compared. This is problematic for application to the multiple attributes of most models thus far elicited by my protocol.

Traditional time-tradeoff method is, therefore, used in the IMQOL elicitation to determine the scaling coefficients for each attribute of health that is considered relevant. Time trade-off provides a currency that allows fine grain comparison between attributes that is not provided naturally.

6.4.2 Will Difficulties of Utility Assessment Accrue?

One might wonder if the difficulties associated with traditional assessment techniques might not accrue when employed as frequently as is the case in the IMQOL method. The assessment of intermediate values relies on standard gambles to value intermediate positions on each attribute scale. The question is whether the difficulties of this technique will accrue in the IMQOL approach. I argue that it is not known that difficulties "accrue". If we are talking about a systematic error caused by risk aversion, an additive assumption has some logic, but it cannot be said that the compounding of the error due to the risk function of multiple utility assessments will always be greater than the summary measure. Individual components of the multiattributed model may each have risk-adverse, risk-prone or risk-neutral risk functions. I am currently unaware of any work on the modeling of compound risk functions. Independent of the issues of additive versus multiplicative versus hybrid utility models, therefore, we have no idea whether the risk functions of multiattributed utilities are additive, multiplicative, hybrid or otherwise. Second, by evaluating the whole on the basis of its parts no error is accrued for the second order and higher interactions between attributes. Such error, if it exists, is buried within the traditional summary measure making its presence and magnitude unknown.

6.4.3 Assuming that the patient-given attributes cover all possible health concerns

Some might worry that as new clinical states arise in the life of the patient, new issues may come to the patient's mind. For the example, in the case of the tree surgeon with an otic neuroma, we might elicit attribute scales concerning balance, physical comfort, spiritual peace, cooking for friends, mobility, etc. These may even be successfully applied in a decision

analysis. The neurosurgeons then raise the issue that a proposed surgery for the neuroma will put select nerves at risk that account for taste. The potential exists that sweet food will taste like metal after the surgery in some specified number of patients. The patient may have never considered the possibility, and therefore the triads of people used to stimulate attribute scale construction contained nothing differentiating altered taste. It is improper to conclude that the patient places no value on the taste of his food, yet such a value is not represented in the IMQOL model without repeating the assessment in an informed context.

It is fundamentally important that the value system be allowed to change with the evolving perspective of the patient. It is very much like laboratory results for serum chemistry assays (e.g., glucose, Blood Urea Nitrogen, liver enzyme levels). A health care provider would be remiss to neglect immediate sampling because he or she has results from some months ago. Similarly, they would be remiss to proceed on the basis of old value system information without inquiring whether the patient has reviewed or been given reason to review their value system. In the context of a dynamic value system, a responsibility will have to be born by the patient to keep information up to date (as in a legal will) and by the provider to recognize critical events outside the expertise of the patient of which the patient needs educated. In this example, knowing that food taste is at risk, the provider should be expected to do two things with the multiattributed value system. First, the provider should be evaluating the outcome of taste alteration in the terms of the attribute scales represented (e.g., cooking for friends, will be affected by the taste alteration even in a scheme that has no direct representation of taste). Second, the provider should be evaluating the patient's preparedness for the decision. If the value system has what is considered an inadequacy, a prescription is indicated. If the patient has no attribute scale to which food taste bears relevance and the provider deems this inappropriate, an investigation and perhaps an education is in order. The indication is made explicit by the IMQOL model value system. In absence of this explicit value system, the indication is more guesswork, if evident at all.

The absence of any triad of people, who bring food taste into the discourse of the interview, should not present a problem. The IMQOL method does not limit the elicitation to what is represented by friends and family. It uses these familiar persons as stimuli for eliciting

idioms. Once the patient gets the idea of what is being asked for and after the triads cease to stimulate more ideas, that phase of elicitation terminates with a query for any remaining relevant constructs free from any specific person in mind. Only when this last question fails to elicit any novel responses does the protocol move on. In one interview performed in the study described later in the thesis, the patient criticized the IMQOL model as not having elements that he used to judge himself, only those attributes used to judge others. This can easily be remedied by adding one more question, which would be to ask, “Are there any other ways in which you evaluate your own life as healthy or unhealthy?”

Chapter 7

Application of Elicited Results

With the representation of values specified and the means to elicit values of that structure specified, it is now possible to explain the application of value systems elicited this way. The application can be illustrated with a case study from a pilot investigation using the protocol to evaluate outcomes of stroke. Before finalizing the details of the protocol, I explored the feasibility of using IMQOL models of this nature with a few Fellows at the Clinical Decision Making Division of the New England Medical Center. This chapter is based on that experience and the stroke outcome pilot study of IMQOL method feasibility with an Emergency Department patient.

Fundamentally, the task is to use the values specified to measure the quality of life for a specified state of health. The IMQOL preference model provides a complete list of comprehensive attribute scales with which to score the state of health. The multidimensional feature of the IMQOL model requires a score to be ascribed for each dimension separately like a grade school report card. The assessor must score each dimension or “report card subject”

separately, allowing the overall score or “grade” to be what it may according to underlying summary assumptions. The operational nature of the IMQOL allows persons other than the original author to employ the resulting scorecard. This is equivalent to saying that others may now measure quality of life in the terms of the individual from whom the model was elicited.

This chapter includes a section discussing the utility of this application for clinicians. In that section it will be brought out how the IMQOL can be seen as a bridge between values and preferences. As such, the IMQOL method assessment serves to facilitate discovery for the patient by providing a systematic structure for the patient to use in differentiating his or her values. The chapter concludes with a warning about the interpretation of the scaling coefficients elicited by the IMQOL method.

7.1 Rating the Quality of Life for Outcomes

The attribute scales and IMQOL model can be used to score any state of health. It is derived from elicitation procedures that do not use any specific context of disease or state of health. This implies the model may be employed to evaluate any state of health that is conceivable by the patient. The scoring procedure consists of three major steps:

1. Determine outcomes to be scored for quality of life.
2. For each outcome, determine the rating for each attribute using the ordered attribute scale.
3. Tally the overall score using the chosen multi-dimensional model and elicited scaling coefficients.

The health care provider performs the first step. The patient and/or provider perform the second step. The last step is a straightforward matter of computation. A small computer program loaded with the patient’s preference model can instantly provide the results as soon as the second step is accomplished.

To properly elicit the requisites for decision analytic modeling, three guiding principles should be adhered to. The outcomes to be assessed should be distinct, described sufficiently, and should not incorporate probabilities of other outcomes. In decision models the probability of an outcome is captured within the structure of the decision tree. This probability should not be factored into the measurement of quality of life beyond the effect of the impact on the level

of quality experienced by someone in that state of uncertainty. Theoretically, the assessor is asked to rate the quality of life for a point in time. It can be difficult to assert an instantaneous quality of life outside the shadow of uncertain outcomes. However, the assessor should be encouraged to think in terms of a steady state quality of life for each outcome described. This is not peculiar to the IMQOL method of eliciting values; it is true of any utility assessment for outcomes in decision analytic modeling.

The second step is the key to this process. It is the locus where the abstractions of one domain (biomedical) meets the abstractions of the other domain (patient values). The key to this step is to clearly instruct the patient or provider to decide which level of the attribute scale best fits the outcome description. With a comprehensive scale, if the state of health bears any relevance to the attribute measured by the scale, then one of the scale levels will fit better than others. How well the attribute scale relates to the biomedical description of the outcome can vary. If the outcome description is thorough, the attribute is relevant and the attribute scale is operational in everyone's hands, then translation is easy. To the degree that either the attribute scale or the description of the outcome remains ambiguous, the translation may be more difficult. There remains a possibility that the issue of the attribute scale may not bear any relationship to the description of the outcome (e.g., loss of hearing in one ear does not affect one's ability to exercise, dialysis therapy does not force one to change smoking habits, mild degrees of disability following stroke do not necessarily change a person's intellectual pursuits or propensity to be a loner or socialite). Two ways of articulating this absence of relationship have proven useful. One is the notion of "not applicable" to reflect no relationship and the other is "no change" to reflect the lack of effect. Both have the same meaning concerning the implications for rating the quality of life. Whatever level of this attribute a person was before the arrival of this outcome, he or she will still be afterwards. Consequently, for each attribute scale, these two additional responses must be provided as options. Whatever the value of health on this scale was before the advent of the outcome is retained unchanged in the evaluation for this state of health.

For many people, especially health care professionals, multi-dimensional assessment is not a difficult concept. For others, it is difficult to grasp. They are afraid that discrediting a

person's health in one dimension will be a disreputable assertion about that person's overall health. For example, it is not uncommon during human-mediated interviews to hear, "A person who exercises a lot is healthier, but just because a person does not exercise, it does not mean he or she is not healthy!" For patients unaccustomed to thinking in terms of multi-dimensional measurements—especially the abstraction of health—the idea can be presented as analogous to a grade school report card as mentioned previously in Section 6.2.

Following the metaphor, a computer program can be written with an interface that looks something like a report card. The attributes of relevance are specified and the allowable levels displayed in a selectable fashion within a table. One by one the outcomes specified as relevant by the provider are presented and the user asked to select the rating from among the allowable attribute levels or "n/a" or "n/c" on a table. The values for the attribute levels selected are displayed and overall score is computed on the basis of elicited utilities and weights.

7.2 Illustrated Application of Elicited Values

The application of elicited values can be illustrated with the results of an actual interview performed. A middle-aged male Anglo-American consented to the IMQOL interview as a preliminary part of the study comparing the results with traditional standard gamble utility assessment described in the chapter on (Section 4.2). The patient was asked to rate the quality of life for five levels of disability specified by a published gradation of post-stroke disability known as the Rankin scale⁵⁴. Patients in the study were not experiencing signs of stroke and were not facing medical decisions regarding the treatment of stroke but they were asked to tell us, to the best of their ability, what their response would be if they were. After eliciting utility values with traditional standard gamble methods while waiting to be attended by emergency department health care providers, they were invited to participate in an IMQOL method elicitation of values by telephone some convenient time after discharge. The decision that motivated the selection of health states is that of a choice between the use of clot dissolving medication in acute stroke, reducing disability measured three months after the stroke, or not to use such medication because it is associated with a slight increase in mortality within the first two days⁷⁷. The

outcomes germane to the decision model are simply various levels of disability or death, hence the use of the Rankin Scale of disabilities. This is a decision of extreme interest to clinicians of emergency departments. The Rankin scale specifies the outcomes for which they would like to know the patient's opinion on quality of life.

After eliciting this patient's IMQOL model, he was asked to score the five health state descriptions of a modified Rankin Scale with the model's attribute scales. This patient revealed four attribute scales that comprised his IMQOL model. They are displayed as the example in the chapter on the elicitation of patient perspective (Figure 6.2). The five levels of disability derived from the Rankin scale were 1) (*No Significant Disability despite symptoms*) "Some symptoms, for example: slurred speech, numbness in face, or reduced strength in an arm or leg, but still able to carry out all your usual duties and activities", 2) (*Slight Disability*) "Unable to carry out activities you could participate in prior to the stroke, but able to look after your own affairs without assistance", 3) (*Moderate Disability*) "Require some help looking after own affairs, but able to walk without assistance", 4) (*Moderately Severe Disability*) "Unable to walk without assistance and unable to attend to own bodily needs, but not bedridden, not incontinent" and 5) (*Severe Disability*) "Bedridden, incontinent, and requiring constant nursing care". This fulfills the first of the three major steps of the procedure for application of an IMQOL model.

"Consider each of the following outcomes from a decision analysis model. Which of the following best describe the state of health represented?"	
STATE OF HEALTH: After a stroke, you require some help looking after your own affairs, but you are able to walk without assistance.	POSSIBLE ANSWERS, Attribute 1: <ul style="list-style-type: none"> • Mental well being • Mentally ill • Mental decay, marked decline in level of mental functioning • <i>Immaterial or has no bearing</i> • <i>No change</i>

Figure 7.1 An example for accomplishing the second step of application: the rating of Rankin Scale level 3 with an elicited attribute scale involving mental well being versus mental decay. A specified state of health is rated using only the terms of the elicited attribute scale unless the state of health is deemed to bear no relevance or not to change anything concerning the attribute.

Figure 7.1 illustrates the second step of the procedure. Specifically it illustrates the rating of Rankin scale level 3 with the first attribute, mental well being versus mental decay. To avoid confusion, this format can be used to rate each health state with each attribute scale, one at a time, until the assessor becomes familiar with the IMQOL method of assessment. In Figure 7.2, a grid displays all attribute scales with columns for five distinct states of health to be rated. This scorecard can be used by the analyst to chart results or directly by decision-makers who are familiar with the protocol and prepared to proceed rapidly without confusion. Figure 7.3 illustrates the scoring grid after the patient had rated all health states.

The third step in the application procedure is merely to compute the values according to the elicited IMQOL model. At any point after the assessor has rated the health state with the attribute scale, the numeric values associated with those attribute levels are substituted as displayed in Figure 7.4. Applying the linear equation with the elicited scaling coefficients for each attribute provides the summary scores. This computation can be performed in real time during the elicitation process with any spreadsheet software using pre-constructed templates for this purpose. This provides an immediate opportunity for approval, feedback or troubleshooting. More sophisticated and friendly user interfaces are possible for those who find this one difficult.

Scale ID	Ordered Attribute Scales	Relative Weight	Value Relative To Attribute Extremes	Health State 1	Health State 2	Health State 3	Health State 4	Health State 5
				Adj Score	Adj Score	Adj Score	Adj Score	Adj Score
i	mental well being **		1					
	mentally ill		0.25					
	mental decay-marked decline in level of mental functioning	.96	0					
ii	no chronic illness or disability		1					
	temporarily ill		0.999999999					
	minor health concerns	.036	0.999999					
	probability of becoming chronically ill but not yet chronic illness **		0.3					
iii	active life		1					
	hypochondriac		.53-.54					
	somewhat active or limited forced sedentary life **	.0024	0.4					
iv	non-obsessive, Non-abusive **		1					
	occasional lapses: into heavy drinking, fluctuating weight, yoyo dieter, binge drinking, recreational drug use(no addiction)	.00077	0.95					
	obsessive personality, drugs or alcohol, self destructive		0					
	Sum:							

Figure 7.2. Blank Scoring Grid for Five Health States Using an Elicited IMQOL model.

Scale ID	Ordered Attribute Scales	Relative Weight	Value Relative To Attribute Extremes	Health State 1 Adj Score	Health State 2 Adj Score	Health State 3 Adj Score	Health State 4 Adj Score	Health State 5 Adj Score
i	mental well being **		1	na	na	XX	XX	na
	mentally ill		0.25					
	mental decay-marked decline in level of mental functioning	.96	0					
ii	no chronic illness or disability		1	XX	XX			
	temporarily ill		0.999999999					
	minor health concerns probability of becoming chronically ill but not yet chronic illness **	.036	0.999999 0.3 0			XX	XX	XX
iii	active life		1					
	hypochondriac		.53-.54	XX	XX	XX		
	somewhat active or limited forced sedentary life **	.0024	0.4 0				XX	XX
iv	non-obsessive, Non-abusive **		1	na	na	na	na	na
	occasional lapses: into heavy drinking, fluctuating weight, yoyo dieter, binge drinking, recreational drug use(no addiction)	.00077	0.95					
	obsessive personality, drugs or alcohol, self destructive		0					
			Sum:					

Figure 7.3 Scorecard with all health states rated before substituting numeric values associated with attribute levels.

Scale ID	Ordered Attribute Scales	Relative Weight	Value Relative To Attribute Extremes	Health State 1 Adj Score	Health State 2 Adj Score	Health State 3 Adj Score	Health State 4 Adj Score	Health State 5 Adj Score
i	mental well being **		1	<i>0.96</i>	<i>0.96</i>	<i>0.96</i>	<i>0.96</i>	<i>0.96</i>
	mentally ill		0.25					
	mental decay-marked decline in level of mental functioning	.96	0					
ii	no chronic illness or disability		1	0.036	0.036			
	temporarily ill		0.999999999					
	minor health concerns probability of becoming chronically ill but not yet chronic illness **	.036	0.999999 0.3 0			0	0	0
iii	active life		1					
	hypochondriac		.53-.54	0.0012	0.0012	0.0012		
	somewhat active or limited forced sedentary life **	.0024	0.4 0				0	0
iv	non-obsessive, Non-abusive **		1	<i>0.00077</i>	<i>0.00077</i>	<i>0.00077</i>	<i>0.00077</i>	<i>0.00077</i>
	occasional lapses: into heavy drinking, fluctuating weight, yoyo dieter, binge drinking, recreational drug use(no addiction)	.00077	0.95					
	obsessive personality, drugs or alcohol, self destructive		0					
			Sum:	0.998	0.998	0.962	0.961	0.961

Figure 7.4 Scorecard from previous figure filled in with assertions of the interview process. Italicized numbers represent a rating of “not applicable” and hence a substitution of the subject’s self assessment on the attribute scale.

7.3 The Health Care Provider Role In IMQOL Application

The role of the health care provider in medical decision making is to 1) identify the current necessity of decision (diagnose a disease in need of a treatment decision), 2) lay out the available strategies of treatment, 3) describe the potential outcomes for each treatment choice and 4) facilitate the patient in making the best choice of treatment (and thereby the choice of potential outcomes) for the patient's welfare. The application of a person's values plays a key role in the evaluation of the outcomes. Patients have no control over the pathogenesis of the disease or the modes of treatment available. They have control over the choice of treatment (in an ideal health care system) and under rational assumptions, should exercise that control based upon their personal values mapped into the possible outcomes. One reason they need the health care professional, besides diagnosis and prescription, is that they don't know enough about the possible outcomes or the possible treatments to decide what treatment is best. To make that decision the patient must be informed of the possibilities and consequences. These must be linked to the patient's condition and the decision options immediately available. The health care provider can do this. If the possibilities and consequences are expressed in the terms of the patient, we can expect the patient to be making a more informed decision than if he or she is considering the possibilities and consequences expressed only in someone else's terminology. Since the patient is unaccustomed to applying his or her values to a new health outcome, the patient may need assistance in applying these values to the possibilities and consequences. The challenge is to enable the health care provider to aid the decision process by using the patient's values system to rate the outcomes⁷⁸. In this way, the health care provider not only supplies the knowledge of what outcomes are possible, therefore in need of judgement, but also serves as consultant on the evaluation of the value in those very outcomes in personalized patient terms. This fulfills the provider role of facilitating the patient in making the best choice of treatment for the patient's welfare. The insight and experience of the professional is merged with the patient's values rather than, metaphorically speaking, left on the stoop a few steps away from the patient's values. The common language creates a genuine facilitation in contrast to a strictly

biomedical-physiological description of treatment rationale and outcomes that might obscure what actually is in the patient's best interest.

If we are admitting that the patient-provider interaction is a communication between dissimilar languages or perspectives, then the proposed application of the elicited value system represents a shift of the health state evaluation from the doctor's linguistic domain over to the patient's linguistic domain. So why does the health care provider have any less difficulty than the patient in dealing with the other party's terms? Perhaps there is no less difficulty for the provider. However, the health care profession is one of the most educated and therefore it seems more likely that the professional side of the interaction would be capable of evaluation in new terms than the non-professional. For sake of experience, it is far more likely that the physician understands what the patient expresses in his or her own terms than the patient would understand what the physician expresses in professional terms. After all, the physician was once an ordinary person and often a patient too.

In an ideal outcome of IMQOL modeling, the physician will be able to work with the descriptors and attribute scales on his or her own without need of patient assistance. A realistic view will likely include concern for the accuracy of application in the absence of the patient, at least until the validity of the instrument is established. The potential to restrict use of an IMQOL model to circumstances in which the patient is available to assist ameliorates this concern. Even in a view tempered with reality, it seems reasonable to think physicians can learn to use these attribute scales with minimal training. That training would come in the form of collaborative interaction with the patient. After becoming familiar with the usage of the attribute scales, the physician should not need the patient so often. This is simply a formal version of a doctor getting to know his or her patient, currently done in ad hoc fashion. An essential step of the IMQOL model elicitation is the interviewer's demonstration of a proficient use of the attribute scale and its terminology. If the interviewer can master the use of these attribute scales in the course of the interview to the satisfaction of the patient, why shouldn't the physician be capable of the same in some short order of time?

To apply the values of the patient to any given outcome description, the description must be clear. Lenert, et al, describe the decreased variability between persons when

descriptions of health used in utility assessment were in greater detail⁷⁹. Although this suggests that more detail in health state descriptions will result in greater utility assessment reliability, test-retest reliability was not strong enough to support such a conclusion in this study. The evidence in this study suggests a learned effect rather than a random effect in the test-retest reliability. This is consistent with the argument put forth by Shiell, et al, who suggest that preferences are poorly differentiated values and as such are a more volatile level of worth than well-differentiated values. Shiell, et al, further suggest that experience adds up in the making of a well-differentiated value⁷⁰. Faced with an unfamiliar outcome or decision, we are forced to deduce our preference based upon some calculus and inference built upon our well-differentiated library of values. Recognizing the frequency with which health decisions involve novel states of health, Shiell encourages us to view utility assessment as an aid in the construction of values. The IMQOL model should be evaluated by the patient and health care provider to verify its adequacy when applied to actual states of health. This can be as much a values clarification or self-discovery process as an elicitation process. The ease of systematic application enabling repetition facilitates the discovery process accommodating frequent re-assessment until the values become well differentiated and more stable. If the IMQOL model does not give agreeable results at the summary score level, the model can be examined regarding which scaling coefficient is out of proportion or which attribute scale is incomplete or bound to ill-distributed values.

As stated previously, it is the role of the health care provider to call upon experience and research to specify what outcomes are of relevance to any medical decision. Given a list of relevant outcomes that are dictated by the diagnosis or differential list and the work-up or treatment options, the patient can be motivated by the presence of the decision outcomes and granted the opportunity to evaluate the outcomes. The provider can assist in that evaluation both directly and indirectly. The patient and provider may apply the patient value system to the list of outcome descriptions independently, in collaboration or by an iterative asynchronous process. The provider may assist directly by collaborating with the patient and coming to a mutually agreeable understanding. The provider may assist indirectly by answering questions, filling in gaps in the patient's understanding of the outcome descriptions and disambiguating

the outcomes. After the patient asserts the value of each outcome, the provider can do the same or qualify the patient's assessment. Fundamentally, there is no reason why the provider could not assert the value of each outcome before the patient does, but my experience has revealed an advantage to the order proposed. Some of the physicians shown examples of values elicited and asked to rate health states with them were reluctant to make a firm commitment of their opinion until they knew something about what the patient thought. They were much more ready to specify a needed correction in the patient's rating with the system, suggesting the patient's assessment best precede the provider's. In some cases, the patient is reluctant to make a firm commitment without knowing what the physician says, but this seems to occur less frequently and is easily mitigated by the promise of the provider's willingness to review the patient's assessment.

The experience of the health care provider can be employed to assist the evaluation or application of a patient's IMQOL model. If the provider is convinced something is missing, this presents the opportunity to repair the deficit through patient education. If something is inappropriately balanced, efforts can be initiated to effect a different value structure or balance of values (e.g., the patient is asserting an unseemly importance to furniture damage should she move in with her offspring rather than a unaffordable nursing home). The provider can contribute to the quality of the preference model at two occasions. The omission may be obvious to the provider at first glance enabling them to contribute to the face value of the model. Alternatively, an omission may become apparent when efforts are made to apply the model to specific health states. The provider can also contribute to the quality of the preference model at more than one level. The provider can capture errors made in the perception of disease or treatment that are leading to erroneous evaluations (e.g., a perception that there is no evidence that smoking causes cancer). At a deeper level of reasoning, the provider may understand the underlying values that are well-differentiated as manifest by the IMQOL model but observe "mistakes" of inexperience in the construction of preferences poorly inferred from those values. The provider may offer alternative ways to put the fundamental values together with alternative results in preferences expressed (e.g., some attention to exercise and diet can minimize burdens to an essential supportive spouse caretaker who has dietary constraints of his

or her own). This extended capacity to contribute to the quality of the preference model is in sharp contrast to the traditional utility assessment methods that fundamentally do not expose anything about what the patient uses to arrive at a given choice in a standard gamble or time tradeoff.

Furthermore, I can say that my experience (limited as it is) of putting hypothetical sets of attribute scales before clinicians has not been met with lack of confidence. Providers may want to correct an IMQOL model's inadequacies or the attributes may stimulate an explicit question they want to ask or statement to tell the patient, but they have offered little resistance for using what is there. I am suspicious that they are like starving persons more than ready to eat, deprived as they are of a systematic way to understand patient preferences. As of this writing, a small study is being designed to evaluate the utility of sample IMQOL models from actual patient interviews with clinical focus groups and longitudinal validity studies. The patients are even more confident. Rarely does one ask for the doctor's answer to any question before committing himself or herself to an answer. In the chapter that discusses the outcome of the feasibility study it will be revealed how many patients want these preference models used to support decisions even if they do not fully understand how they were derived.

7.4 The Utility Of This Application For Clinicians

The question may arise in the reader's mind regarding the potential for clinicians to ascribe descriptors from an IMQOL model to clinical outcomes. In many ways, the difference between health care providers and patients can be like a cultural difference. The depth and magnitude of domain knowledge required to practice medicine often distances the provider from the patient on the order of a cultural barrier. As with any cultural language barrier, there is a risk that translation will be difficult from one culture's language into another. Busy patients in a modern paced age do not have time to learn what they might be able to otherwise and expect their provider to take care of everything but the bottom line of actual decisions. The demand created by this situation is for a discourse and language of convenience to enable translation between "cultures". As pointed out in the chapter describing the patient perspective representation, traditional utility assessment has been accused of eliciting poorly differentiated

values as preferences rather than deeply held, well differentiated and stable values. I propose the values elicited with the IMQOL method are more likely to be on the well-differentiated end of the spectrum since they are originated by thinking of familiar friends and family. As such, they are exactly what the providers want to understand for both descriptive and prescriptive purposes. In so far as my proposition is true, this application of IMQOL model attribute scales to outcomes directly embodies the inference of preferences concerning unfamiliar health outcomes from deep-rooted values.

The question remains whether the descriptors of the attribute scales in the individual's unique terms can be related to the biomedical descriptions of outcomes. If it can be done at all, can it be done efficiently enough for clinical application—will it be cost effective? Will the provider be able to spend the time required to learn the constructs of the patient as they change? In the previous section, I suggested that the provider is more likely able to understand the words of the patient than the patient to understand the words of the provider. I suggested also that the provider can learn to use the operational attribute scales of the patient well. The answer regarding how often IMQOL models will be used and their cost effectiveness depends upon how long it takes for application and how often the model changes by what degree. If we measure stable values with accuracy, we can expect a minimal amount of migration, implying a minimal cost for recurrent use following the initial cost of the first elicitation. Until we begin to collect and analyze values with the IMQOL protocol, this remains an open question. The demand for empirical evaluation is clear.

From another perspective, we can view the application of the IMQOL model as a first draft of applying the patient values so clarified to the health outcomes of relevance. It can be considered a systematic approach of discovery. As a consultant, the health care provider can offer insight and advice concerning the “experimental” expression of preferences that result from the inexperienced patient's attempt to infer preferences from deeply held values. The patient be naïve in rating the state of health (e.g., rating moderate disability following stroke as a “hypochondriac” and not as “somewhat active or limited”). The experienced provider can offer alternative definitions, additional description details, and even alternative chains of inference to assist in the discovery process. This is made possible by the nature of the IMQOL

method that calls upon the familiar and here in application asks both patient and provider to apply those values explicitly to decision outcomes. The reusable nature of the IMQOL model insures that the time invested by the provider in learning to see the patient's perspective will likely be used again and again in so far as the IMQOL model embodies more deep rooted values than labile preferences.

7.5 Interpreting Coefficients

Keeney and Raiffa⁸⁰ point out the error in interpreting the scaling coefficients of an additive model as the relative importance of the attributes. The basis for this observation is the dependence of the scaling coefficient upon the range of the attributes used to elicit the utility of multidimensional examples. To find points of indifference, from which we infer quantitative relationships between the attributes, an election is made regarding the range over which each attribute will be varied. This range might be large or small. In an additive model, a larger range will result in a larger scaling coefficient, all other things being equal. For example, if in comparing job offers, the difference in salary is small and the utility is normalized over that range, location of the job might have an equal or larger contribution to the preferred job. We can easily expect the scaling coefficient for location to be greater in the additive model for job preference than the scaling coefficient for salary. If however, the range of utility is normalized over a larger scale, by an order of magnitude for instance, it is reasonable to expect the scaling coefficient of the resulting additive model for job preference to have a larger coefficient for salary than for location.

To further illustrate with an quantified example, let us assume an additive function for a three attribute IMQOL model.

$$utility(Attribute_1, Attribute_2, Attribute_3) = \lambda_1 utility(Attribute_1) + \lambda_2 utility(Attribute_2) + \lambda_3 utility(Attribute_3) \quad (7.1)$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$, $utility(Attribute_i^*) = 1$ and $utility(Attribute_i^0) = 0$

Let's further assume that the individual asserts that 10 years with $Attribute_1$ at its highest level, all others at their worst, is equivalent to 20 years with $Attribute_2$ at its highest level, all others at their worst.

$$utility(10 \text{ years}, Attribute_1^*, Attribute_2^0, Attribute_3^0) = utility(20 \text{ years}, Attribute_1^0, Attribute_2^*, Attribute_3^0) \quad (7.2)$$

=>

$$\begin{aligned} utility(10 \text{ years}, Attribute_1^*, Attribute_2^0, Attribute_3^0) &= 10 * \lambda_1 utility(Attribute_1^*) = 10 * \lambda_1 \\ &= utility(20 \text{ years}, Attribute_1^0, Attribute_2^*, Attribute_3^0) = 20 * \lambda_2 utility(Attribute_2^*) = 20 * \lambda_2 \end{aligned} \quad (7.3)$$

=>

$$\lambda_1 = 0.5 * \lambda_2 \quad (7.4)$$

Let's finally assume that the individual asserts that 5 years with *Attribute*₁ at its highest level, all others at their worst, is equivalent to 20 years with *Attribute*₃ at its highest level, all others at their worst.

$$utility(5 \text{ years}, Attribute_1^*, Attribute_2^0, Attribute_3^0) = utility(20 \text{ years}, Attribute_1^0, Attribute_2^0, Attribute_3^*) \quad (7.5)$$

=>

$$\begin{aligned} utility(5 \text{ years}, Attribute_1^*, Attribute_2^0, Attribute_3^0) &= 5 * \lambda_1 utility(Attribute_1^*) = 5 * \lambda_1 \\ &= utility(20 \text{ years}, Attribute_1^0, Attribute_2^0, Attribute_3^*) = 20 * \lambda_2 utility(Attribute_3^*) = 20 * \lambda_3 \end{aligned} \quad (7.6)$$

=>

$$\lambda_1 = 0.25 * \lambda_3 \quad (7.7)$$

Solving for the scaling coefficients with *n* equations for *n* unknowns using (7.4), (7.7) and the assumption of (7.1) results in the following.

$$\lambda_1 = 0.5714, \quad \lambda_2 = 0.2957, \quad \lambda_3 = 0.1429 \quad (7.8)$$

However, if we were to in some abstract way double the range of *Attribute*₁, making *Attribute*₁^{*} twice as far from *Attribute*₁⁰ and normalize the utility function from zero to one over that range we would alter the scaling coefficients. Assuming all things linear, we could expect the points of indifference to be 5 years compared to 20 in (7.2) and 2.5 years compared to 20 in (7.5). The resulting relations between scaling coefficients would be as follows.

$$\lambda_1 = 0.25 * \lambda_2 \quad (7.9)$$

$$\lambda_1 = 0.125 * \lambda_3 \quad (7.10)$$

The scaling coefficients then become as follows.

$$\lambda_1 = 0.7273, \quad \lambda_2 = 0.1818, \quad \lambda_3 = 0.0909 \quad (7.11)$$

If we directly interpret the scaling coefficients as “relative importance”, we would infer that *Attribute*₁ is twice as important as *Attribute*₂ and four times as important as *Attribute*₃ from λ_1 , λ_2 , and λ_3 . However, note that λ_1 , λ_2 , and λ_3 suggest *Attribute*₁ is four times as important as *Attribute*₂ and eight times as important as *Attribute*₃. If we similarly assess the relationships with the utility of *Attribute*₁ normalized over a range half that of the original range, we get $\lambda_1=0.4$, $\lambda_2=0.4$, and $\lambda_3=0.2$. That would suggest *Attribute*₁ equals *Attribute*₂ in importance and only doubles *Attribute*₃ in importance. The exponential relationship between the increase in the feature range over which the utility function is normalized to the relative importance asserted is notable, but it is predicated on knowing that the range was doubled to begin with. If we are dealing with non-parametric ordered nominal scales, determining the magnitude of any change in range considered will be complicated at best. There are prolific studies of psychometric assessment methods based on estimating changes in magnitude (e.g., asking when a light bulb is twice as bright as another is). These might offer ideas for refinements where the range of the attributes for IMQOL models are of concern, but this would further complicate the elicitation process.

I would like to address this issue more fundamentally. In that the IMQOL method seeks comprehensive attribute scales, there is reason to argue that the process does yield scaling coefficients that reflect relative importance. As Keeney and Raiffa carefully point out, it is safe to say that the IMQOL scaling coefficients represent the importance of changing one attribute of health over the considered range relative to changing another attribute over its respective considered range. In so far as the range represented is the full spectrum of conceivable attribute levels, then the gamut is covered (at least for this individual’s way of looking at things). We would be safe in saying that we know the relative importance over the conceivable range of one attribute compared to the conceivable range of another attribute.

As I observe the responses to IMQOL elicitation regarding health issues, I am left to wonder if many people would actually reason in partial ranges if we intended them to. It appears to me that they are making qualitative comparisons between the concepts as a whole when they are formulating their answers. In the multidimensional tradeoffs of the last phase of IMQOL model elicitation it is possible to describe a hypothetical state of health involving

intermediate levels of attributes as easily as states only involving extremes of the attribute scales. Determining points of indifference to collect a suitable set of equations to solve for the scaling coefficients of the additive model are still successful because the values for all intermediate attributes are known. The individual is asked to compare partial changes in attributes rather than full spectrum changes. It is my suspicion in the interviews so far that the individual is very likely to give similar responses in both cases, falsely elevating the scaling coefficient associated with the reduced attribute range. It appears to me as though the respondents have an easier time working with their recently articulated extremes than they do reasoning with partial ranges in quantitative terms. However, no measurement of this suspicion has been made at this point.

Chapter 8

The Feasibility of IMQOL Elicitation and Application in Practice

To evaluate the feasibility of the IMQOL method of values clarification, I interviewed 15 dialysis patients with the protocol. The main objective of this study was to determine if patients could complete the task successfully. Secondly the purpose was to find out if the results were believable and useful. To preliminarily evaluate completeness, comprehensiveness and operational quality, the IMQOL models were used to assess the value of clinical outcomes. To make the study clinically relevant, common dialysis outcomes are used to evaluate the ability to apply the IMQOL models elicited. These outcomes were inspired by the frequent decision faced by dialysis patients between hemodialysis and peritoneal dialysis. As portrayed in the earlier chapter on “motivating examples”, there are many tradeoffs in this decision and a profound interest for clinical nephrologists in the patient perception of quality of life in regard to these differing treatment modes for end stage renal disease³⁹.

8.1 Study Method

A convenience sample of eleven hemodialysis and eight peritoneal dialysis patients were invited to participate from among the patrons of the Dialysis Center of Dialysis Clinics Incorporated located at the New England Medical Center. Candidates for participation had to be English speaking end stage renal disease patients who had been undergoing dialysis treatments for at least three months, were mentally competent for coherent conversation and were at least 18 years old. As part of a larger list of objectives than specified here, each consenting patient was asked to respond to a battery of evaluations. Participants were fully informed about the intent and design of the study and were free to ask any questions. The patients who volunteered to participate were not compensated with anything more than the experience of the interview and the satisfaction of contributing to the investigation. Interviews were conducted at the dialysis center. Hemodialysis patients were interviewed during idle periods while undergoing dialysis (2-3 hour periods). Peritoneal dialysis patients were interviewed in a private exam room in the center normally used for training such patients. Usually, multiple interview sessions were scheduled at the convenience of the patient, for example when a visit was required to collect lab samples or routine provider consultation was scheduled. Some interview sessions were scheduled without any other excuse to visit the center. At the request of one male participant, his interview was completed in one single day with an hour lunch break in the middle. Most interviews with peritoneal dialysis patients were completed in two 2 ½ hour sessions or less. The battery of evaluations constituting these interviews began with the Mini-Mental Status Exam, a six item Satisfaction With Decision questionnaire³⁸, and the Medical Outcomes Study Short Form-36 health status survey^{16, 63}. These were administered before any quality of life measurement was begun. The interview sequence was consistent except for The Medical Outcomes Study Short Form-36 health status survey, which was filled out at varying points of respite during the interview process. Four standardized states of health commonly experienced by dialysis patients and the patient's own health were assessed for quality of life using the standard gamble and time tradeoff method prior to using the IMQOL protocol for assessment.

In addition to evaluating their own true health, the participants were asked to evaluate their quality of life if they were a person 1) on hemodialysis (HD) treatment without complications, 2) on hemodialysis treatment with a complication (HD/c) of a clotting vascular access site, 3) on “continuous ambulatory peritoneal dialysis” (CAPD) with no complications and 4) on “continuous ambulatory peritoneal dialysis” with a complication (CAPD/c) of peritonitis (infection of the belly wall) every 9-18 months. Obviously, two states, one without complication and one with complications were selected for each modality of dialysis treatment. Preprinting them on a single sheet of paper standardized the descriptions used. They were given to the participant to read or read aloud to those with poor eyesight. The two descriptions involving states without complications (HD & CAPD) were derived from those used by the CHOICE study group in order to facilitate comparison of results with that quality of life research. The full descriptions are found in the Appendix and sufficiently define these medical terms. Using the methods described in this thesis, IMQOL models were then elicited and subsequently applied to the same four standardized descriptions of life on dialysis treatment as states of health. Immediately following the application of the IMQOL model, individuals were asked to give their opinion of the outcome and the values elicitation process. They were asked to compare the traditional and IMQOL measurement methods.

8.2 Results

8.2.1 Cohort Characterized

8.2.1.1 Participants Invited

Consent forms requesting interviews every two months for two years, were passed out to twenty hemodialysis patients, chosen by convenience for the schedule of the collaborating nephrologist. Only the results of the first round of interviews are available at this writing. The first eleven patients (3 female, 8 male) to return signed forms were selected for interviews; three females ranging in age from age 41 to 68 (mean 54.5) and eight males ranging in age from 36 to 76 (mean 57.7). One female was a relatively recent immigrant from Africa and had mild to moderate hearing loss. She could understand most words used in the interview and had

no problem asking for clarification if she was uncertain of the meaning of unfamiliar words. All the remaining participants were fluent with English and had no difficulty understanding the language of the analyst. Two male participants revealed the existence of heart disease in the course of their interview and it was laborious for another male to read though he followed charts and visual aids well. These participants varied in experience with dialysis and vascular access complications. All had working vascular access either in the form of a graft (synthetic material shunt or “shortcut” between an arm vein and artery) or a fistula (a transplanted vessel situated as a shunt between an arm vein and artery). One female’s interview was begun immediately before a surgery to replace a failing access. The latter portion of her interview was postponed for the convalescence of the surgery. The eldest female was the last to begin her interview because of numerous days where her blood pressure was not well controlled during dialysis. One male patient had a single interview session postponed for the sake of poorly controlled blood pressure. One patient had indirect experience with peritoneal dialysis when his wife was on that treatment. All participants had some fundamental understanding of what peritoneal dialysis was.

The population of peritoneal dialysis patients at this center is much smaller. Only eight patients fulfilled the inclusion criteria in the judgement of the peritoneal dialysis nurses and all were invited to participate. All eight, three females (ages 26 to 56, mean 42.7) and five males (ages 21 to 64, mean 44.4) consented to participate with same consent forms as hemodialysis patients. Peritoneal dialysis participants were not bound to machines during the interviews, but were willing to schedule extra time at the center for their participation. Some had experienced at least a brief period of hemodialysis, but none had vascular access sites (either a graft or fistula) for hemodialysis. One patient was new to peritoneal dialysis therapy although he had been on hemodialysis prior. Hemodialysis therapy had not gone well for him and providers had encouraged him to try peritoneal dialysis.

All participants, except one (HD male), scored 28 or higher out of 30 on the Mini-Mental Status Exam, which is considered mentally competent by those whose use that evaluation. Only seven (3-HD, 2-male, 1-female; 4-PD, 2-female, 2-male) scored a perfect 30/30. Six scored 28/30 (4-HD, 2-female, 2-male; 2-PD, 1-female, 1-male). The one exception

was a 63 year old male HD patient who scored 25 out of 30. This score was the result of two failures: 1) his inability to either count forward by sevens or spell “WORLD” backwards and 2) his inability to copy, with pencil and paper, a printed design consisting of overlapping pentagons preserving all sides, angles and the intersection as a quadrangle. Prior to the interview, these patients were satisfied with their decision regarding the choice of dialysis treatment modality. All answers to the questions on the Satisfaction With Decision questionnaire were answered with “Agree” or “Strongly Agree” with the exception of three participants. Only one question each in two of these three individuals was answered with “Disagree”. One male HD patient “disagree[d]” that they were “adequately informed about the issues important to their decision between dialysis treatment modality”. This person “strongly agreed” for all other items. One female HD patient “neither agree[d] nor disagree[d]” that she was “satisfied that her decision was consistent with her personal values.” She simply “agree[d]” with all other questions. Finally, one male HD patient agreed that he was adequately informed and that the decision was consistent with his personal values, but disagreed that the decision he made was the best decision possible for him and disagreed that he would be able to carry out the decision. He was equivocal regarding the overall satisfaction with his decision and whether it was his decision to make.

8.2.1.2 *Interviews Completed*

Interviews were staggered over a period of twelve weeks; individual interviews spanning generally 2-3 weeks, no more than 4. Fifteen of the consenting patients completed the interviews.

Because of the segmented nature of the interviews in the clinical setting where nursing interruptions were common, the duration of the interviews was not rigorously tracked. However, the full battery of evaluations rarely took longer than six hours—the IMQOL protocol rarely longer than four hours. One hemodialysis patient did not complete the interview because he went on vacation for a few weeks. One male peritoneal dialysis patient’s schedule got more crowded with work after volunteering. He failed to show for the first appointment and subsequently did not respond to phone messages. Two hemodialysis patients were unable to complete the interview after starting. One female withdrew from participation and one male

was unable to answer the questions satisfactorily requiring the analyst to abort the interview diplomatically. The latter individual was the one who also scored lowest on the Mini-Mental Status Exam. The first two of these four incomplete interviews are considered no fault of the interview process, but the latter two deserve discussion concerning what prevented completion. The female was, in fact, the individual who underwent surgery and convalescence of vascular access replacement with several days in which she did not feel up to conversation. However, other contributing factors played a role in preventing her completion that will be discussed later in this chapter. The remaining fifteen participants were able to complete the battery of evaluations, including the IMQOL model elicitation and to apply the result to their own health and the four states of health specified by the study design.

8.2.2 Data Collected

8.2.2.1 Qualitative Opinions of Participants

After applying the IMQOL model to assess the quality of life in the four dialysis therapy descriptions and their own state of health, 93% of those completing the interviews answered, “Yes,” when asked if they would want their doctor to use this scoring system to help make personal medical decisions. Eighty percent wanted their doctor to use it even if they were not available to assist in its use (e.g., if they were unconscious or mentally incapacitated). When asked if the IMQOL model score for their personal health made sense on a scale from zero to one hundred, eighty percent said, “Yes.” Asked if the score for their current state of health made sense relative to the scores for other states of health, 93% said, “Yes.” Seventy-seven percent indicated the score for the four dialysis treatment descriptions made sense on a scale from zero to one hundred. Eight percent said they would want to use this scoring system personally to help make medical decisions. Among the three who did not, one (HD male) said, “That’s a hard question,” one (HD male) gave a response that did not answer the question and one (PD male) said, “Maybe in a few years after its tested and debugged.” Participants were asked which method of measuring quality of life made them the most comfortable. Six of the twelve that answered the question said the IMQOL method with the remainder split between the standard gamble and the time tradeoff method. Finally, participants completing the

interview were asked the following question. “Given the results from time tradeoff, standard gamble and this new method, which numbers do you believe to be the most accurate representation of your values?” Seven responded with confidence in the IMQOL measurement, two in the time tradeoff, one in the standard gamble and one thought it a tie between the standard gamble and the IMQOL measurement.

8.2.2.2 *Means and Distributions*

Figure 8.4 shows the mean, range and confidence interval of the results of each method for each of the four states of health and self-assessment. It can be seen that the range for all methods is very wide in all states of health evaluated. The range of the IMQOL results were exceeded by the range of both traditional methods except in the case of complicated peritoneal dialysis. In that case, the range of the results for the IMQOL method was zero to one hundred. Measurements for quality of life with IMQOL models ranged from zero to one hundred with an overall average of 72. Measurements using the standard gamble method ranged from 3 to 9999/10,000 (average 66) on the same scale. Measurements using the time tradeoff method ranged from 5 to 59.5/60 on the same scale.

The complexity of the IMQOL models can be characterized by the dimensionality of the individual models and the number of levels found in each attribute scale. The number of dimensions in the elicited IMQOL models ranged from four to 12 (average 7.7). The distribution of the number of dimensions is graphed in Figure 8.1. The number of levels in each attribute scale ranged from two to five (average 3.1). The overall distribution of the number of attribute scale levels is graphed in Figure 8.2. The average number of attribute scale levels for each participant is graphed in Figure 8.3 sorted by the average.

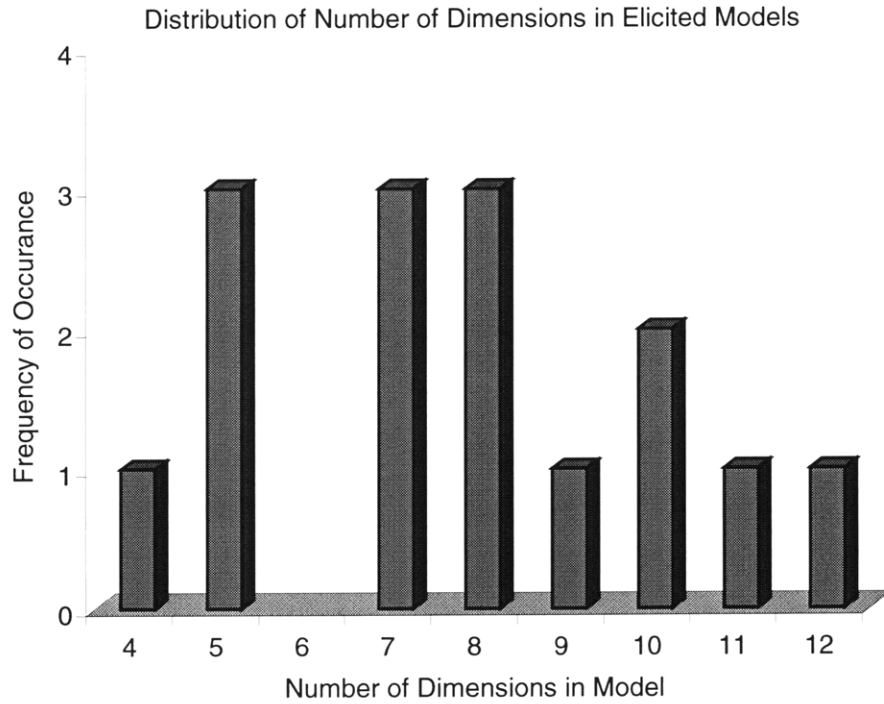


Figure 8.1 Dimensionality of Elicited Models

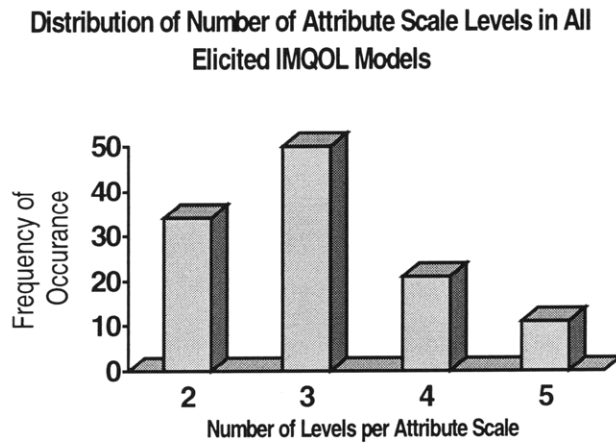


Figure 8.2 Distribution of Attribute Scale Complexity

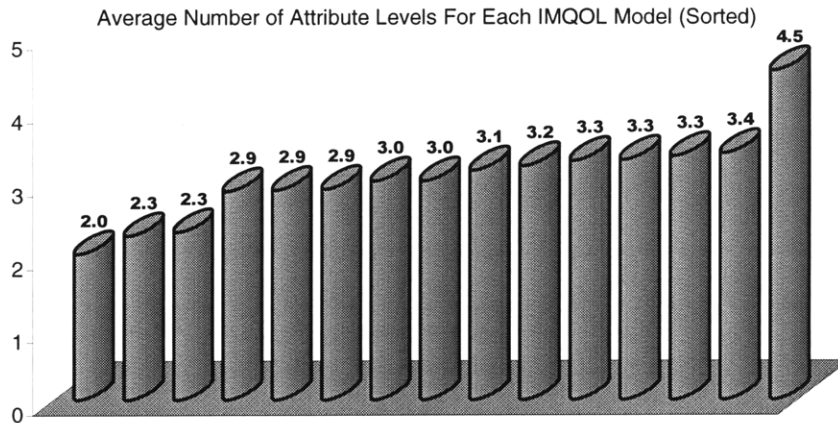


Figure 8.3 Average Complexity of Attribute Scales for Each Participant - Sorted

The mean of the quality of life asserted for all four states of health and the patient’s own health were not significantly different but a weak trend suggested uncomplicated states were better than states of health with complications and peritoneal dialysis subtly higher than hemodialysis. The mean quality of life asserted for the participants’ own health was 75, 75, and 63 for IMQOL, TTO and SG respectively with 95% confidence levels ranging from 12 to 16.6 above and below the mean. The overall mean for the quality of life for all states was 71 out of 100. Uncomplicated hemodialysis results were 76, 66, and 71 for IMQOL, TTO and SG respectively. Uncomplicated peritoneal dialysis was 79, 70, and 68 for IMQOL, TTO and SG respectively. Complicated hemodialysis results were 66, 59, and 61 for IMQOL, TTO and SG respectively and peritoneal dialysis results were 63, 62 and 67 for IMQOL, TTO and SG respectively. The mean rating these patients gave their own health was 75, 70, and 63 for IMQOL, TTO and SG respectively.

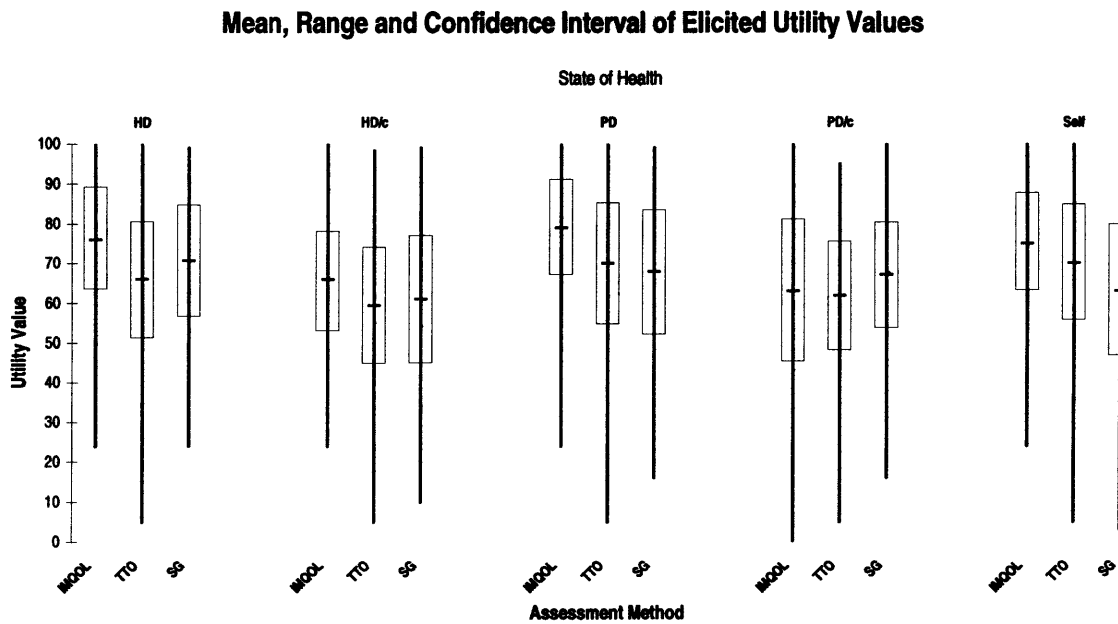


Figure 8.4 Mean (cross-hair), range (whisker), and 95% confidence interval (box) of the elicited utility values for hemodialysis(HD), hemodialysis with the complication of a clotted access (HD/c), continuous ambulatory peritoneal dialysis (PD) and continuous ambulatory peritoneal dialysis with complication of peritonitis every 9-18 months (PD/c). Triplets represent (from left to right) results from the IMQOL, time tradeoff (TTO) and standard gamble (SG) methods of assessment.

Chapter 8: The Feasibility of IMQOL Elicitation and Application in Practice

Table 8.1 Results of Standard Gamble Utility Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c).

Participant	HD	HD/c	PD	PD/c	Self
#01	83	83	50	50	75
#02	29	18	16	16	3
#03	1/100	1/100	40	1/10x4	10-25
#04	60-65	50-55	75-80	65	60-65
#05	75	50	75	75	75
#06	97	97	97	85	75
#07	24	25	19	40	35
#08	40	40	40	40	45
#09	80	60	90	90	80
#10	75-80	60-65	95	85	-
#11	85-90	75	90	85	96
#12	75	60	65	55	60
#13	90	90	90	90	90
#14	40-45	10	75	50	75
#15	99	94	99	80	98

Table 8.2 Results of Time Tradeoff Utility Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c). Results are expressed in the least number of years of perfect health (numerator) considered equivalent to an expected life span (denominator) in the state of health at the head of the column. For example, participant #01 considered 10 years of perfect health to be equivalent to 20 years of uncomplicated hemodialysis.

Participant	HD	HD/c	PD	PD/c	Self
#01	10/20	10/20	10/20	10/20	15/20
#02	2/40	2/40	2/40	2/40	2/40
#03	18/20	18/20	15/20	15/20	18/20
#04	15/20	12/20	15/20	11/20	12-13/20
#05	30/40	20/40	35/40	30/40	30/40
#06	19.5/20	19.5/20	19.5/20	19/20	18/20
#07	20/40	20/40	10/40	10/40	25/40
#08	5/10	5/10	5/10	5/10	4/10
#09	15/30	10/30	20/30	15/30	15/30
#10	18/20	17/20	18/20	17.5/20	-
#11	59/60	48/60	58/60	53/60	55/60
#12	12/20	12/20	18/20	16/20	17/20
#13	10/20	10/20	15/20	15/20	15/20
#14	15/30	10/30	20/30	15/30	25/30
#15	59.5/60	59/60	59.5/60	40/60	59.5/60

Table 8.3 Results of IMQOL Method Assessment for Hemodialysis (HD), Hemodialysis with Thrombosis Complication (HD/c), Peritoneal Dialysis (PD) and Peritoneal Dialysis with Peritonitis (PD/c).

Participant	HD	HD/c	PD	PD/c	Self
#01	100.00	83.74	87.80	34.55	100.00
#02	72.30	55.97	55.97	52.40	55.15
#03	23.94	23.94	23.94	0.00	23.94
#04	75.99	75.99	75.99	71.20	75.99
#05	62.22	59.14	52.96	52.96	49.88
#06	94.87	84.40	94.42	90.31	75.45
#07	73.16	73.16	69.47	4.737	53.16
#08	100.00	100.00	100.00	100.00	75.68
#09	92.08	65.82	95.65	95.65	95.51
#10	98.75	98.75	96.88	96.88	96.87
#11	89.49	50.39	92.43	35.69	92.43
#12	39.84	39.84	75.33	72.07	72.07
#13	72.37	40.67	72.37	71.70	72.03
#14	54.88	48.21	97.39	86.38	97.39
#15	96.57	84.16	96.57	84.18	96.57

Table 8.4 IMQOL Model for Participant #01 (Self-Assessment in Bold).

Coeff	Attribute levels	Value	Self
0.244	loners, self-sufficient, people who put loved one's priorities before their own	1	*
	people who don't know what they want, follow the crowd, mild or moderately dependent on others	0.5	
	selfish	0.35	
	always has to be around somebody	0	
0.203	eat right	1	*
	unknowing, don't know what to eat for their own good	0.65	
	arrogant people who eat what they feel like	0	
0.203	open-minded or always trying to be	1	*
	brought up with a standard ok for themselves	0.8	
	very closed-minded	0.7	
	evil and selfish	0	
0.179	live right, putting children and family first before self	1	*
	not putting priorities in order	0	
0.171	like working around the house	1	*
	not interested if it doesn't benefit himself	0	

8.2.3 *Withdrawn Participants*

The circumstances contributing to two participants who did not complete the interview should be outlined objectively. Some details have already been given regarding the two withdrawn from the interview process. As previously stated, these were both hemodialysis patients, one male and one female, so for convenience they will be referred to by gender for clarity of reference with no implied gender specific causality or implications.

As mentioned above, the female was experiencing difficulties with her vascular access site and several days feeling ill or tired from convalescing. She was the least excited about participating in the interview, having witnessed, at a distance, how long the analyst spent talking with other patients in the dialysis center. She scored perfectly on the MMSE. She was not satisfied that the decision regarding hemo- versus peritoneal dialysis was consistent with her values and disagreed with the statement that it was her decision to make on the Satisfaction With Decision questionnaire. She struggled to answer the questions as described previously and expressed her distaste for the first session of the interview comprised only of traditional assessment methods and MMSE as “boring.” She began the IMQOL method under the analyst’s encouragement that it would be different. However, her enthusiasm did not improve and it became clear that she wished for the analyst’s approval to withdraw, although she would not come out directly and say so. Giving her “permission” to withdraw met with reserved contentment. There was reluctance about discontinuing but a greater disinterest in making judgements called for by the interview. In the process of withdrawing from participation, she complained of being asked to make choices concerning death, judging the health of her acquaintances and being asked questions she did know the answer to. When carefully asked if it was that she did not know the answers or did not understand the questions, she replied, “Both.” She then immediately restated her qualm with questions that she did not know the answers to, omitting anything further about not understanding the questions.

The male who was withdrawn from the interview process was a different story. His willingness to continue was never in question. In his case, an insurmountable difficulty with abstract thinking seemed to interfere with any progress in the IMQOL protocol beyond the point of describing opposites or individualizing elements of health. The first struggles came

when it was difficult to think of friends or family who were still alive in this older gentleman. His recluse lifestyle was no help in identifying anyone suitable for the first step of the IMQOL elicitation. Two triads were eventually found and used to elicit a dozen candidate ideas that could normally be used to construct bipolar constructs. However, eliciting what was the opposite of an idea in his terms was difficult. Asking how the remaining individual of the triad differed from the others in the same respect only brought up new ideas, not opposing poles. More direct query, such as, “What is the healthy opposite of ‘no exercise, walk only to turn cans in to the redemption store’?” only resulted in puzzled looks. Adopting a common sense approach where the analyst took the liberty to assume the opposite was a simple linguistic negation (e.g., walking for exercise) enabled the process to uncover the next insurmountable hurdle. When attempts were made to fashion questions for self-assessment out of the elicited bipolar constructs, the participant could not agree to any proposals. In each case, when any statement was articulated regarding people who are <some stated term> are more healthy than <the stated terms assumed or asserted opposite>, this individual would be compelled to thoughtfully disagree, making long statements about how, “You never know how health that person might be. He might fall over dead any second! One who [exhibits the unhealthy feature] could live forever!” It was not that this individual could not form a strong opinion. He consumed most of an entire dialysis session lauding the ills of being an alcoholic and to the contrary, the “lack of proof” that smoking is bad for your health, demanding to see the data that proves it. He would not agree to any construct of smoking versus non-smoking although he would admit that some people who smoke get lung cancer. Being an alcoholic was the only thing that came close to constituting an acceptable dimension for an IMQOL model but even that could not be agreed to in the end. His justification was that there are people who drink excessively and live a long life while there are those who are “Tea-totallers” who “might fall over dead any day.” Subjectively judging that this interview was neither going to help the individual nor any other people, it was cordially discontinued with full appreciation expressed to the participant. It helped to identify limitations to the protocol that will be characterized in the discussion below.

8.2.4 Analysis

8.2.4.1 No significant difference between methods

It is clear looking at the means and confidence intervals of Figure 8.1 that there is no significant difference between the means of quality of life for the five states of health. T-tests show no difference in the means between methods and F-tests show no difference in variance. Analysis of variance for the results between methods shows no significant difference in variance accounted for by factors of age (above or below median for the sample), gender (male or female), or MMSE score (29/30 and above or 28/30 and below). Treatment modality is a significant predictor of quality of life assessment with most method/state combinations but the 95% confidence interval for the coefficient always includes zero when this is true.

8.2.4.2 Discord

Seven out of fourteen participants (50%) who completed the interviews and rank ordered the five states of health, gave discordant responses when comparing the rank order with standard gamble or time tradeoff assessments. Among those who did not complete the interview, all who performed the traditional assessments (3 HD patients) distinguished treatment modalities but gave *weakly severity-discordant* responses[†] for time tradeoff assessments. The female HD patient who was withdrawn from the study gave one of only two cases overall of *strong severity-discordant* responses to standard gamble assessment. The other case was a woman who was a native of Africa and did have a mild difficulty with the English language. When asked to rank order the four specified states of health before the IMQOL elicitation process, participants often indicated no difference between complicated and uncomplicated health within the same treatment modality. These assertions were verified by direct questions. We can define this as *weak severity-discordance* and find that it was commonplace, occurring seven times in standard gamble method responses given (5 HD, 6 PD). It occurred eight times in the time tradeoff method responses (8 HD, 6 PD). Only five of these participants gave *weak severity-discordance* in both traditional methods. Only two of

[†] Refer to the chapter on discordant responses in Classical Assessment to review the definition of *weakly severity-discordant*. Specifically, this idiom refers to a response which does not distinguish between health states considered to differ in severity from a biomedical health perspective.

those were consistent in giving scores that differentiated health states to the same degree with both traditional methods. The others either 1) differentiated modalities in time tradeoff but not in standard gamble, 2) differentiated modalities in standard gamble but not time tradeoff or 3) gave *weak severity-discordant* response for PD in time tradeoff but not standard gamble. There were seven cases of *weak severity-discordant* responses to the IMQOL method (5 HD, 4 PD); four in common with standard gamble and four in common with time tradeoff *weak severity-discordant* responses. Only two participants were consistent in giving *weak severity-discordant* responses for all three methods of assessment. One additional participant was consistent in giving *weak severity-discordant* responses for all three methods for HD treatment modality only. There were occasions where the two modalities were regarded as equivalent in quality of life as well. This occurred twice in standard gamble responses, only one of those shared with similarly equivalent time tradeoff responses. This individual also indicated equivalence with the IMQOL method. In addition there was one other participant who indicated equivalence in time tradeoff responses alone.

Of those who consented to be interviewed but did not complete the interview, one (PD male) never began, two (HD male & HD female) completed the traditional method assessments for the four dialysis treatment states but did not get to rank order them, and one (HD male) whose vacation interrupted the IMQOL process did complete the traditional method assessments and rank order for the four dialysis treatment states and his own health. The latter participant did get as far as the elicitation of attribute scales (7 scales - 3 to 5 levels per scale, avg. 4.1) but not to the point of eliciting the values for intermediate levels or scaling coefficients and, hence, there was no opportunity for verification of operational quality. His traditional assessment responses were *weakly severity-discordant* for both treatment modalities with the standard gamble method and *weakly severity-discordant* for hemodialysis states with the time tradeoff method. The time tradeoff responses were the most consistent with his rank order of the states. The male HD participant who was unable to complete the interview once he began was not asked to rank order the states but did succeed in giving *strictly concordant* responses for both treatment modalities with the standard gamble method, similarly for the PD states with the time tradeoff method and *weakly severity-discordant* responses with the time

tradeoff method for HD. This individual volunteered that he had no personal experience with peritoneal dialysis but that his wife had been treated with it before her death several years prior. The remaining female hemodialysis participant who discontinued the interview was *weakly severity-discordant* with the time tradeoff method and *severity-discordant* with the standard gamble responses for hemodialysis. Her interview response was *weakly severity-discordant* with the standard gamble method response for peritoneal dialysis. Her initial responses to most questions during the traditional assessments were numerically reasonable but immediately followed by bemoaning statements like, “I can’t answer that!” or “I don’t know.” She clearly struggled with the traditional methods.

8.3 Notable Case Studies

Promoting an individualized quality of life measurement without attention to individual responses would be inappropriate. Indeed the more profound observations come from close examination of results for individuals. Just as ethnographic analysis of computer programs and graphic user interfaces benefit most from the first few individuals to whom programs are exposed, these earliest of experiences of individuals may be the most potent reflections on the IMQOL method’s adequacy and needs for maturation.

8.3.1 *Absence of difference between states*

The eldest female HD participant (#08) gave results that offered no difference between any of the four states of health that were not her own state. The result differed for each method used but within each method, no differentiation was indicated between the dialysis descriptions. Standard gamble assessment gave a score of 40/100. Time tradeoff assessment gave a score of 50/100. The IMQOL method gave a perfect score. The same methods gave results of 45, 40 and 75.7, respectively, for her own health. This individual was never convinced that she understood the assessments, but went along with the study. Similar to the female who withdrew from the interviews[‡], she professed disinterest in the assessment. It may be worth remembering that this

[‡] In fact, this woman usually occupied the neighboring chair to the woman who withdrew during the same dialysis shift. The two of them would often chime in together greeting my arrival with castigation for asking such boring questions.

woman had numerous days in which her blood pressure was not well controlled interfering with scheduling our interviews. The responses to the debriefing questions were interesting in this light. Asked if the IMQOL model score for her current state of health made sense on the 0 to 100 scale and relative to the other states, she said yes to both. Asked if the IMQOL model scores for the other states made sense on a 0 to 100 scale, she did not think they deserved a 100/100—“More like 75 or 80.” However, when asked if she would want to use this scoring system to help make personal medical decisions, her reply was, “Yes.” When asked if she would want her doctor to use this system—with or without her help—she replied, “Oh yeah, he knows more than I do; yeah, I think so.” Asked what she thought of the IMQOL method as a means of understanding her values, she replied, “I don’t quite understand it, but it is a good idea.” Finally, when asked about the fact that each analysis suggests that she believes there is no difference between the four states of health she was asked to evaluate, she agreed. “I’ve been through them all,” she stated. Her attitude seemed to be that she had seen it all and was just waiting out her remaining time. In our brief conversation, she convinced me that she really did see all four dialysis descriptions as an equal quality of life.

8.3.2 Extreme Scaling Coefficients #1

With two participants (#04 & #05, both male, HD) it was difficult to elicit scaling coefficients. Both were ultimately able to provide answers and both had favorable post-application appraisals of their elicited IMQOL models. The difficulties posed, remedies found and lessons learned were distinct. The first involved an extreme aversion to pain and the second was fundamentally an unwillingness to trade any length of life for any attribute. What these two did share in common was that they both involved an extreme importance of a single dimension.

To characterize participant #04 as supportive would be an understatement. He liked the idea of the IMQOL model and openly made statements indicating he wanted to see it succeed. Nonetheless, he found it hard work when faced with grave choices. His occupation is computational and similar in nature to the ambitions of this research. He commented numerous times about how hard we was trying to approach things with a mathematical mind set, maintaining proportions as he answered the standard gambles that revealed the values of

attribute scale intermediates and the time tradeoffs to manifest scaling coefficients. However, he declared that he was unable to. He was usually forced to abandon that mind set to simply answer the questions on the basis of what he preferred directly in terms of the nominal descriptors. This individual is notably the only person in all my experience to date that has ever explicitly identified pain as an attribute scale (in stark contrast to the expectations of most health care providers that I have talked with). When it came time to rank the attribute scales, this individual put the pain attribute scale at the top of his rank ordered list of scales. As the first of the scenarios of the scaling coefficient eliciting phase were assessed, it became clear that he was indicating an extreme weight for avoidance of pain. He was willing to trade just about any amount of a 20 year life expectancy for a life with “chronic unrelievable pain or untreatable pain” for life with “no pain or trivial pain.” It was evident after only a few scenarios that any other attribute would pale in comparison to pain. It was obviously pointless to go on with other attribute scales of less importance—all would generate a response of complete willingness to relinquish as much of his life span as presented. This hemodialysis patient openly admitted that he had experienced a good deal of pain associated with a heart problem in recent years and would do “anything” to avoid that. Upon suspecting that this was actually an indication that life with such pain was worse than death, I asked if this was so. He contemplated before responding as if to decide for the first time and said, “I suppose it is.” It was further postulated that he was, in essence, committing what I have since coined as “technological euthanasia.” Although he never spoke in terms of euthanasia or suicide, in the scenarios presented, he was taking any opportunity availed him to actively shorten his assumed life of pain. There was no other indication that he was the kind of person who would take his own life. When presented with these ideas he found them agreeable and affirmed that he would not likely take his own life knowingly.

The clear indication was that the time tradeoff method for assessing any hypothesized state considered worse than death represents a dominated decision in which the respondent under rational assumptions will shorten life as much as allowed. Where the person is cognizant of the meaning implied, they may or may not agree to the explicit terms of euthanasia, but, as in this case, they may operate that way in the expression of their preferences with these tools.

In terms of recognizing the occasion, indications are clear. A response indicating the respondent will relinquish his or her full life span, or nearly so, for a health state free of the ill extreme of an attribute, has indicated a preference of death over life with that level of that attribute. If that level of that attribute is a part of any health state in the elicitation of scaling coefficients for an IMQOL model, there will be no point of indifference found because of the dominated decision. In effect what is created by these circumstances is a hypothetical scenario where the only thing available to the patient for treating the condition is the inverse of time.

To remedy the above and proceed with a successful elicitation of an IMQOL model, the analyst needs only to alter the scenarios to all be free of an attribute level which causes the hypothetical state of health to be worse than death. This was done with this participant. The first adjustment was to simply proceed with hypothetical states of health where the pain was fixed at its optimal extreme. In fact, this person had indicated a point of indifference that related the attribute scale for pain and his second-most important attribute scale, “ability to do what they want when they want” versus “restricted by treatment requirements.” Under an assumption that this relationship would prove adequate for solving with simultaneous equations (offline), the interview moved forward. By this I mean that the protocol was followed for the remaining pairs of attribute scales, not repeating the elicitation for the pair of first and second-most important attribute scales. This was successful in creating reasonably logical scenarios and points of indifference with the time tradeoff elicitation for scaling coefficients. I will return to the discussion of results later.

Taking advantage of the enthusiasm of the participant for this study, I pursued variations in the elicitation of scaling coefficients to reinforce the eventual interpretation and validation of his scaling coefficients. One variation was to avoid the potential confusion of a tradeoff between two attributes and the third dimension of life span (time) by performing a simple time tradeoff assessment for a change in a single attribute[§], fully specifying that all other

[§] It has been suggested that posing a tradeoff between two attributes compounded by a foreshortening of life is likely to confuse the patient. When presented this way at the outset, this is a valid criticism. Ideally, we would like a measure that directly compares two attributes (e.g., exercise and non-smoking) rather than indirectly as does a method comparing one attribute at a time to a common currency—time in this case (exercise vs. life expectancy and non-smoking vs. expectancy). The single attribute approach can lead to illogical difficulties—for example, a treatment that cures smoking, improving that attribute, but it shortens your life. In practice confusion does not seem to happen as frequently as might be expected. It is my hypothesis that this could

attributes would remain unchanged. In effect, this then becomes a tradeoff between time and the single attribute. Comparing all attributes to time—one at a time—should provide the common measure of value by which to determine scaling coefficients of the linear model. In typical time tradeoff fashion, a hypothetical pre-treatment state of health is assumed for a specified expected life span and compared to a post-treatment state of health with a shorter life span. The only difference between pre- and post-treatment is in terms of one attribute scale. The only logical requirement is that the post-treatment state must be more attractive than the pre-treatment so that the expectation is reasonable that the individual will be willing to negotiate a shorter life to gain improved quality. For reasons discussed in the section on interpretation of coefficients of the chapter on IMQOL application, only extremes of each attribute scale were used. At what level the other attributes are fixed remains another variable in the approach. In what I will call the *pessimistic* approach, the other attributes are fixed at their worst extreme. This is the equivalent of asking, “If all attributes of your health were poor and a treatment could cure just this one aspect of your health with a consequence of a shorter life, at what point would it be so short that you would no longer be interested in the treatment?” The opposing alternative would be to fix all the attributes at their best extreme in what I will call the *optimistic* approach. This is equivalent to asking, “If you were very well off in regards to health except for this one last attribute, and a treatment could cure that last thing with a consequence of a shorter life, at what point would it be so short that you would no longer be interested?”

Both the *pessimistic* and *optimistic* approaches were used with this individual. The result was surprising. Prior to this experience, the expectation was that results would differ considerably. However, the outcome of the comparison was that there was little to no difference for this individual. The logic behind the expectation for difference was that a person would desperately desire to improve a poor circumstance, even likely be risk-seeking, but would be less motivated for change in a circumstance of near perfection. There is empirical

be an effect of anchoring the assessor on the tradeoff between two attributes before introducing the notion of a shortened life. This is accomplished by incrementally building up to the direct tradeoff between attributes calling for decision twice before any side effect on life span is introduced as described in the chapter on the elicitation method.

evidence to support the idea and a theoretical explanation of the difference between utility assessment involving risk and not involving risk based upon it⁸¹. What this would mean for how the scaling coefficients would change is unclear—for example, would they go up or go down as a rule? Will all coefficients change in the same direction? In this case, comparing the resulting coefficient for a given attribute scale, the *optimistic* approach yielded higher values than the *pessimistic* five out of seven times; once they were equal and once the *pessimistic* was higher. However, the magnitude of the difference of the answers in years of life span was small—ranging from 5-15%. The magnitude of difference in utility values was 1%, -22%, 15%, -12%, 32%, -70% and -6% from most to least important attribute scale respectively. These differences could be the result of 1) actual differences in relative weights, 2) inconsistencies generated by poorly differentiated values, 3) inconsistencies generated by fatigue or 4) a low signal-to-noise ratio for human judgement or the measurement of it. Assuming they are significant findings, the consequence to the overall score of states of health measured was diluted by the intermediate ratings on the individual attribute scales. The result was very little change in the overall scores. That is to say, the conclusions regarding the application of the IMQOL model for this individual would be the same whether the context be the first step up from the worst of health or the last step to perfection. His scores for both hemodialysis states and his own health were 73.2, 75.7 and 76.0 for the *pessimistic*, *optimistic* and modified original approaches, respectively. His score for the peritoneal dialysis complicated with peritonitis was 69.4, 71.9 and 71.2 for the *pessimistic*, *optimistic* and modified original approaches, respectively. The magnitude of these differences are notably small and disarm the concern that the framing of the scaling coefficient elicitation has significant impact on this individual's estimated values in the vicinity of these health states. Further study is warranted to determine how representative this individual is.

8.3.3 Extreme Scaling Coefficients #2

It was difficult to elicit scaling coefficients with a second male hemodialysis participant. However, the explicit symptom was the opposite of wanting the shortest life possible—he would not entertain any shorter life span for any purpose whatsoever. He was

clear in his rank order of the attribute scales and seemed well suited to the quantification involved in steps prior to that step in the protocol. Seven out of ten attribute scales were binary. He was sure of his answers and quick to respond throughout the interview. Asked directly if he thought all the attribute scales were equal, he said that they may be. He suggested a modified willingness-to-pay approach—simply letting him think about how much of one hundred dollars he would spend on each attribute (with replacement of the money between attributes considered). This was the participant's idea and I went along to see how well it would turn out. He was able to distinguish between most of his attribute scales with a quantitative value this way and agreed most with the order indicated by the outcome of the exercise. Since the result is quantitative and met his approval, these results were used for his IMQOL model. My interpretation of this outcome is that this participant apparently values length of life significantly more than any of the attributes captured in the IMQOL model. This is reasonable and does not represent an incomplete multidimensional model because the dimension of time is strategically avoided. This was the only occurrence of its type. This case would be systematically recognized by the consistent unwillingness to relinquish any life span in response to scaling coefficient elicitation with the time tradeoff framework. An adaptive loop in the protocol's algorithm might call upon an alternative strategy as was done in this actual case.

8.3.4 Empirical Requirement For Continuous Attribute Scales?

Early in the development of the IMQOL protocol, it was anticipated that cardinal scales would be required in addition to ordered nominal scales for attributes of a more continuous nature and measure. Early investigation into the nature of responses likely to the questions employed suggested that nominal scales alone were adequate. No interviews necessitated a continuous scale of infinite levels until one female PD patient was interviewed. This woman, who was an architect by occupation, responded with an incredulous tone when asked for suggestions for what lie between the extremes of her bipolar constructs early in the IMQOL elicitation protocol. She rattled off several levels rapidly in a sentence construction that carried an implicit message that there were plenty more possible levels where those came from. We proceeded with the levels she had already given, optimistic that something operational would

result in the end. She was apologetic for not “complying” and for not seeing how she was being helpful to the study when the first interview session ended a third of the way through the protocol. Even though I was preparing to accommodate her with continuous scale elicitation techniques from Keeney and Raiffa⁵⁷, I was convinced she was destined to be another non-believer in the IMQOL method. She signed up for the study convinced she had something well thought out to say regarding quality of life, but was dismayed by the process midway. Ironically, she simplified the numbers in the assessment of values for the intermediate levels of her attribute scales. If she was asked to consider 25 people out of 100, she described how, to her that was like 1 out of 4. She claimed that the smaller numbers personalized things for her. She claimed not to care about any number out of one hundred. Transforming the number into a whole number fraction of 2, 3, or 4 people made it something she could care about—for example, 30/100 became 1 out of 3, 45/100 “is close to 50/100” and became 1 out of 2, 75/100 became 3 out of 4, and 60/100 became 2 out of 3.

Prior to that, when it came time to see if there were any attribute levels between the ones originally offered, a few of the many she rattled off were consolidated even though intentions were to accommodate her with continuous scales as needed later on. When it was time to elicit values for all the intermediates, she consolidated a few more labels as a result of seeing that she was willing to trade off the same for one or more of the top attribute levels in a few more cases. It was her reasoning that if levels were worth the same in trade, they must be the same category. This is how disjunctive statements wind up as descriptors for attribute levels; level “A” and level “B” consolidated become a single label, “A or B”. Asked just prior to the scaling coefficient elicitation protocol if the nominal scales were going to be sufficient, she consented that the continuous scales had proven to be unnecessary since these were adequate. Over one week later, when we finished the interview, I asked again and she reminded me that I had asked already and affirmed a second time that they were adequate. So it remains to be seen whether we will need to provide mechanisms for continuous attribute scales. In addition, during her post-interview de-briefing she responded in a surprising fashion regarding her IMQOL model application. When asked which set of results, standard gamble, time tradeoff or the IMQOL method, did she believe best represented her values, she was surprised

to be saying it but that it was the IMQOL model. She started out the debriefing by proclaiming that the IMQOL did not tell her anything she did not already know—it simply confirmed what she new beforehand. She thought it was an excellent way for others to know how she thinks.

8.3.5 *Self Discovery in Elicitation*

Both of the latter two case studies gave similar remarks in the post-interview debriefing. The male participant offered the opinion that he would have thought the scores given by the IMQOL analysis were too high before the interview. “If you ask if I was healthy I would probably not have thought so but thinking through the possibilities how it could be worse [in the IMQOL interview] this becomes believable.” He openly confessed that he thought his current quality of life was not as bad as he had thought before the interview.

The participant discussed in the last case study did not discover anything unforeseen in the outcome of the analysis according to her explicit statements. She did say that she was surprised to find the IMQOL results most consistent with her values even if they did not present a new conclusion. It is safe to say that she was surprised by much of the intermediate information revealed along the way: the need, or lack thereof, for continuous scales, the equality of value for some variations in attribute levels, the difference between rating scale results and constrained choice based results, and the overall confidence that quality of life measurement could be decomposed assessed on individual dimensions and recomposed into quantitative values she could agree with.

These are both cases where the participant is learning about their values as elicitation with the IMQOL protocol ensues. It is a personal experiment, if you will, in which the inferences by which outcome preferences are explored. In so far as the IMQOL modeling is accessible for repeated experiments—like a person might use a spread sheet to experiment with a chart of numbers and interdependent calculations—the potential exists for reiterative improvement of the IMQOL model until the user is satisfied with the outcome. Thus the application of the IMQOL modeling paradigm for elicitation and application becomes a developmental tool for values clarification, linking the deep-seated (stable) values of the patient to the manifestation of those values in a preference expression. The more a person reviews the

IMQOL model refining the choice of coefficients and attribute level values, the less labile we might expect his or her indicated preferences in terms of outcomes to become.

8.4 Discussion

The qualitative responses of the participants support the believable and useful nature of the IMQOL model. The completion rate for the interview in these volunteers suggest that most adults will be able to complete the protocol, at least when administered by a human analyst. The length of interviews required to elicit a complete list of comprehensive attribute scales as defined by the IMQOL protocol can be accomplished is two to four hours but this can be accomplished in split sessions over a period of days. Even with lengthy interviews, patients are generally satisfied with the values captured and dimensions represented in the IMQOL model. With split sessions, the threat of poor responses resulting from fatiguing interviews does not present itself in the form of disagreeable model opinions on subsequent days.

The order of attribute elicitation was not apparently systematically biased. Ranks ordering the attribute scales after they were fully elicited did not indicate that the first ones elicited were likely to be more or less important than those elicited later. There were cases where the first scale elicited was the most important and others where the most important scale was elicited last, but few of either of these.

Obviously, the IMQOL model is not operational for all individuals. The two individuals who were unable to complete the interview indicate the method of elicitation is not universal. The examples from this study suggest there are at least three types of individuals for whom the protocol is not well suited: 1) those who have difficulty with abstract thinking, e.g., the concept of opposites or contrasts, 2) those who have difficulty in isolating factors from multi-factorial issues, and 3) those who resist making concrete judgements. As mentioned in the description of the elicitation method, if some steps are not well worded, the respondent may balk at the questions. For example, some individuals balk at the notion that just because a person is poor in the specific attribute of the question that means the person is unhealthy. This indicates to me that they are not thinking clearly of one dimension out of a multi-dimensional model. It is my untested hypothesis that some of these individuals are those that lack skills in abstract thinking

and others are just astute listeners or literalists. The previous discussion of methodology (Section 6.2) offered a way to accommodate people inclined to take things more literally by carefully articulating the question. One individual in this study (the 63 year old male who scored lowest on the Mini-Mental Status Exam) illustrates how difficult it can be for some to follow the abstractions of the paradigm. This example suggests that if an individual has difficulty with the abstraction of opposites, the IMQOL method—and perhaps all utility theory based values assessments—will be crippled. The second individual for whom this method was particularly difficult was one who does not like to make judgements. This is not a question of intellectual capacity, but rather a question of will. It is difficult to force a person to make up his or her mind who does not want to decide. Utility theory is premised by axioms of rationality. Underlying that is an assumption that if we just provide more details that are concrete the rational choice will be easier to elect. This is not necessarily so for those individuals who do not want to make decisions. Perhaps there are those for whom no amount of convincing details will make deciding any easier. It is unclear whether the IMQOL method presents any more difficulty for these individuals than traditional utility assessment. Clearly, the decisions required to complete the protocol are more numerous in the IMQOL method, but whether this translates to an absolute difference in performance remains an interesting question for study.

Although the IMQOL method of quality of life values assessment is not suited for everyone, this study supports it as at least as suitable as the traditional utility assessment methods. It should be remembered that the participants of this study were volunteers and hence the selection of study participants is biased toward those interested in having a voice or actively participating in their affairs, medical and perhaps non-medical as well. This does not represent the more passive and withdrawn segment of society.

The absence of significant difference in means and variance between methods of assessment supports the similarity of the IMQOL measure to the traditional measures. Decision theorists argue that the standard gamble method measures something different than the time tradeoff and rating scale methods because the standard gamble involves risk and reveals the assessor's risk attitude. This difference does not show in the form of significantly different means in this study and by analogy any similar difference between the IMQOL method and

traditional methods may be masked. In these results, the time tradeoff measurement does not produce typically lower values than standard gamble. This may be a function of sample size or of the region and range of utility measured⁸². However, in terms of feasibility and useable output, the new method emulates the traditional, suggesting nothing is lost by the introduction. If anything is gained in the IMQOL method, it is at no expense in the character of the output for this sample population.

No significant difference in the means of the health states makes them quantifiably indistinguishable with such measures. They are thereby no help in deciding which is a better quality of life, hemodialysis or peritoneal dialysis, with or without complications. The study population as a whole, judging by the group means, seems to disregard the difference between the uncomplicated states and those states of health with complications. Only insignificant differences in means suggest that there might be a difference in a larger study with more statistical power. The same can be said of the difference between dialysis treatment modalities. It is difficult to evaluate these issues when the range of responses is so broad and the differences so narrow. Indeed, individual responses frequently indicated blatantly that all states of health were viewed equivalently sometimes in all methods. Other individuals explicitly distinguished the states consistently across methods. While this makes it difficult to generalize, it underscores the need to approach quality of life assessment on an individualized basis. Furthermore, the states are evidently difficult to distinguish with summary measures and stand to gain from decomposable measures.

The complexity of the IMQOL models is not problematic. The range for the number of dimensions of each model and the number of levels in each attribute scale is very manageable. This is likely to be a natural consequence of human intellectual capacity. In so far as the IMQOL method captures the individual's operational discrimination system, it reflects the complexity of the person's natural thinking system. Hence, natural limitations to the complexity of human thought and the number of factors we are conscious of that contribute to such discriminatory judgements will limit the size of the models produced by the IMQOL method. This characteristic of the IMQOL model is likely to be lost when the user is allowed to

deviate from the concrete world of actual people and wander philosophically into abstract realms for ideas with which to construct attribute scales.

Another contribution to the minimization of complexity is the shallow degree to which the concepts captured in the attribute scales are pursued. Personal Construct Theory, on which the first steps of the IMQOL method are based, elaborates on the hierarchical nature of the bipolar constructs elicited by such techniques. Hierarchies are built by asking the individual to explain abstract bipolar constructs with subordinate bipolar constructs and to cluster constructs under super-ordinate bipolar constructs. Should the IMQOL method as present in this thesis prove insufficient in eliciting operational attribute scales, it may be due to the abstract concepts articulated. That is to say that the individual might supply terms for attributes that are not widely understood as measurable making judgements by others on their behalf difficult or inaccurate. Extending IMQOL along lines suggested by more elaborate details of the Personal Construct theorists could provide needed measurability if one assumes as I do that abstractions are merely a substituted linguistic simplification for a multidimensional collection of more measurable constructs. Extending IMQOL would take the form of more questions investigating each immeasurable attribute pole to elicit more constructs to give the abstract ones meaning until all terms are well defined in measurable terms. If the IMQOL method elicits attribute scales that require more elaborate hierarchies to be operational, then we would certainly expect more complex models than those captured in this first feasibility study.

If one adopts equivalent scores as a manifestation of discord in these results then the discord rate of 50% might be regarded as high. However, the abundant evidence has been revealed supporting the equivalence of value between studied health states for some individuals in this group with both traditional summary measures and the IMQOL measure. In this context of complex tradeoffs and indistinguishable states of health, the taxonomy of discord does not contribute as much insight as in domains where equivalence is more puzzling like that of the previous chapter on discordant responses in stroke outcome value assessments.

This study suggests the attribute scales are operational for the patients themselves in that they successfully produce scores not unlike those of traditional utility measures. It remains to be seen if the same is true for clinicians using the patients' IMQOL models to evaluate states

of health. Proposals for evaluating the operational quality in clinicians' hands will be discussed in a later chapter.

Scores for all outcomes specified and the participants' own state of health were successfully generated. Patients generally indicated by their overall approval and answers to direct query that nothing had been left out in the representation of their values supporting the completeness of the IMQOL models. In so far as this small sample of specified outcomes evaluates the comprehensive nature of the attribute scales, the IMQOL models elicited were adequately comprehensive to address outcomes of dialysis. Rigorous evaluation of attribute scale comprehensiveness is possible by application to broader and more remote outcome domains. Values were acquired for clinically relevant states of health for the domain of dialysis patients. Attribute scales rarely involved terms specifically related to end stage renal disease or dialysis therapy, but proved to be adequate for the generation of scores comparable to standard gamble and time tradeoff assessments for quality of life.

Chapter 9

Serendipity

In the course of any investigation, there are unexpected discoveries. Naturally when investigating the values of diverse individuals such discoveries will abound, more so when trying to capture those values with a structured interview as in the IMQOL method. It is the purpose of this chapter to share some of the more meaningful surprises discovered in the course of the feasibility study presented in the previous chapter.

9.1 States are not trivial to specify

To assess the value of states of health the states must be specified. Determining which states need to be assessed is not as trivial one might suspect. Consulting decision analysts spend significant time deciding what the decision tree should consist of and what the outcomes are. They do this by eliciting the branching structure of the lotteries from the domain expert in terms of what might happen under various preconditions and what the value of those events happening might be. When asking health care providers what the possible outcomes are, one is likely to get a variety of responses.

One common response is vague terms that clump together multiple possibilities. For example, the outcome of kidney toxicity could be acute renal failure requiring dialysis. As you

know by this point in this thesis, there is a wide variety of life styles that fall under the general label, “dialysis.” The response of the clinician may not be definitive. He or she may describe the events possible because of chosen actions rather than outcomes. For instance, the risk of choosing a population-based dose of a nephrotoxic drug in a patient with small body mass can be said to be acute renal failure. This statement does not specify anything about the outcome even if an assumption of “loosing the lottery” is made. Acute renal failure implies more about the immediate time frame than it does about the eventual outcome. “Acute renal failure” does not tell us if dialysis is required or, if required, for how long. Another type of response from domain experts is a perpetual unfolding of possibilities. This might take the form of a chain of events (e.g., the secondary sequelae to renal failure or the sequence of alternate sites used as vascular access sites fail), a branching pathway of undeterminable dimension (e.g., an overlapping list of complications possible with vascular access sites) or a network of loops begging for some Markov modeling (e.g., hemodialysis patients converted to peritoneal dialysis and vice versa). Another possible response is one with embedded temporal features (e.g., temporary catheter placement, maturing vascular fistulas or strategies that, once started, should be sustained for an interval before passing judgement).

These ambiguous, non-definitive, compound, or non-instantaneous descriptions offered as outcomes can be difficult to avoid. Even clinicians familiar with decision analysis, but who mentally spend more time in the framework of clinical practice than decision analytic modeling, will easily fall into these patterns of response. These responses must be disambiguated and teased into distinct outcomes or the result of evaluating with utility assessment will be an assertion, at best, of some conclusion about a grab bag of health states with ambiguous implications.

Time must be kept distinct in these models. Decision analytic theory sorts out the temporal and probabilistic elements of decision from the value of the outcomes. The temporal component of decision analytic models is captured in the time adjustment of quality-adjusted life years (where the measure of quality of life is multiplied by the time spent in that state) or, in a Markov model, it is captured in the number of cycles spent in any particular health state. The task, therefore, is to quantify the value of outcomes modeled independent of time, i.e.,

without consideration for time spent in the state. It is apparent to me that this is not a natural way of thinking as evidenced by these typical responses of the domain experts in clinical practice. These typical responses would indicate that we tend to compound our evaluation of outcomes without teasing apart the elements distinguished by decision analysts. It would follow that it would be even harder for patients in general to assess value completely independent of time.**

It is important then to distinguish individual concrete outcomes to be evaluated. The states to assess must be specified in order to apply the IMQOL model. Appropriately segregating individual outcomes can quickly grow to an overbearing quantity. To say nothing of the time required eliciting them individually from the health care provider, asking the patient to assess the quality of life for an overbearing number of intermediate states of health can deter participation and the quality of assessment. A select number of states can be assessed, but the selection of that set deserves investigation. It can not be assumed that intermediate states of health have no bearing on quality of life. The health care provider could select the states to be assessed on the basis of clinical experience. The decision analyst might be able to offer an empirical and analytical basis for selection. Insignificant intermediate states need to be eliminated from the evaluation list to minimize the time and effort burden.

It is the obvious job of the decision analyst to determine the significance of short-term morbidities. However, the issue of social versus individualized perspective plays a role in evaluating the significance of short-term morbidities as much as it does in long term outcomes. As individualized utility assessment is enabled, there may be justification for re-examination of previously studied issues of short-term morbidity impact in specific disease contexts that were previously viewed only in terms of consensus opinion.

Consequently, care must be taken concerning the quality, quantity and choice of health states to be assessed.

** One way that occurs to me to evaluate this would be to measure the utility assessment of paired health state descriptions with specified duration, one long and one short. These results could be compared to utility assessment results where the same health state descriptions are used minus any statement concerning duration. Comparison could be made to see if the responses are closer to the values assessed for shorter or longer descriptions. Conclusions may vary with the individuals tested.

9.2 Death States Assessed In Two Frames Of Mind

Having discussed the importance of the choice of outcomes to measure for decisions, I can characterize two distinct approaches taken by respondents in evaluating states of health regarded as near or worse than death. Two fundamentally different approaches were observed in response to the IMQOL model application to such states.

In the pilot investigation of quality of life measurement using the IMQOL method for outcomes of stroke, a number of grave states of health were assessed (see the chapter on discordant responses). Grave, here, refers to states that might be considered by some to be worse than death; near death in most other people. Attempting to assess the quality of life in states near death can pose interesting challenges.

The assessment of value for states near or worse than death are of interest to many parts of the medical community where there exists a growing abundance of means to prolong life. The decision motivating the study presented in the chapter on discord explicitly involves the distinct outcomes of long term disability, immediate death and death within six months. As stated above, the value of the outcomes should be asserted independent of duration in that state or of the probability of being in the state. Death, as we know it, does lend an advantage of being characterized as unchanging and also as a permanent (absorbing) state, however, when evaluating a state described as “death within six months,” there is a question as to what the assessor is being asked to evaluate; the time before death, the manner of death, the “life” after death or the loss of life interval before death.

Standard gamble assessment has successfully been employed to evaluate quality of life for states worse than death⁵⁶. Basically, the task is accomplished by insuring that the worst conceivable state is assigned the value of zero offering it only as the negative outcome for the lottery offered. The result is an assessment question that may pose certain death juxtaposed to a lottery between perfect health and some state worse than death. This framework does not require the interpretation concerning what the person is being asked to evaluate. It simply infers from the probabilities of the lottery at a point of indifference what the outcome is worth. It is only when we attempt to apply a multidimensional preference model to such states that the issue arises. Naively, we could expect the IMQOL model that covers all conceivable outcomes

to cover states near or worse than death as well as any other quality of life. The test comes when it is time to apply the elicited IMQOL model to such states.

In one case observed in the stroke outcomes study, the individual had a difficult time when it came to rating the states involving death with her attribute scales. The framework she was working with was an attempt to rate the outcome predominantly as a post-mortem quality of life estimate. She was generally inclined to report that the attribute scale was not applicable to the state of death. The application protocol calls for a rating of “not applicable” in this case. However, the computation of the overall score for “not applicable” is to substitute the self-assessment rating. This makes sense only where the change of state causes no change in the patient’s attribute. Clearly, this woman’s death would change things—a case where “not applicable” is not equivalent to “no change.” The resulting scores for states involving death were naturally erroneous while her IMQOL model worked fine for states above death. In the post-interview debriefing, this person suggested that the model did not work because it was not built with anything but living examples in mind. It remains unclear how to systematically stimulate attribute scale construction from deceased acquaintances, but her point is well taken. If her IMQOL is comprehensive as it exists, the proper observation should be that death renders an extremely low rating in most attributes. Some means of rephrasing the question or emphasizing the criteria of change rather than applicability in such cases, might ameliorate this problem.

A second individual in the same pilot study measuring stroke outcomes approached the problem in a different manner. He interpreted the same questions posed in the application process to be asking him to rate the quality of life, whatever its length, prior to death. For the health state involving death within two days, he tried to evaluate life for those two days. For the health state involving death within six months, he tried to evaluate the quality life of those last six months, knowing what was eminent. In his framework, there simply was nothing to evaluate after death. This individual had no difficulty with the protocol and found the results consistent with his direct qualitative estimations.

None of these observations result from the IMQOL modeling directly. These observations are a combination of self-reporting and the investigator’s subjective assessment.

These assertions regarding the first case are the result of post-interview conversation about why the scores were discordant with her rank ordering and traditional measures. The disclosure in the second case resulted from the investigator's request as to why the IMQOL worked so well, pointing out the nature of difficulty others had with the grave health states. It is notable that there is nothing directly captured by the IMQOL model that manifests which manner of interpretation is being used for grave health states. It would appear that, in the case of states involving death, the results should be qualified by some statement about the interpretation of death and the approach taken in application of attribute scales. The best way to systematically qualify application to such states remains unclear, but the feasibility of applying IMQOL to them remains optimistically open. These cases also underscore the critical role played by the scope of the outcome and the clarity of its definitive description.

9.3 Operational Versus Knee-Jerk Responses

What started out to be an avoidance of bias turns out to suggest a pattern of behavior in respondents. Without objective measure, I have observed a high frequency of migration from the first approximation for quality of life in a rating scale context to a different value when immediately followed with a decision scenario with either the standard gamble or time tradeoff framework. This observation has led me to distrust the initial response until substantiated with a similar response to a well presented hypothetical decision, i.e., a context in which the respondent must exercise a choice. This phenomenon is readily observed and can be pointed out to the assessor in justification of the added queries. The respondent will likely endorse the value indicated by the more operational context than the initial response following the added effort.

To avoid anchoring, the IMQOL protocol asks the patient first to offer some number to seed the discussion. In the second phase of the IMQOL method, this takes the shape of a formal rating scale assessment before the standard gamble is employed to determine the value of the intermediates of an attribute scale. In the elicitation of scaling coefficients, this can take the shape of an open question asking how much shorter a life span in the healthier state would cause the assessor to lose interest in the treatment. In both cases, the number volunteered by

the patient can be used to graphically represent the assertion. Henceforward, it is safe to say the graphic representation is not biasing the responses because it is consistently the patient's claim that is visualized by the graphic aid. It was frequent that the first response to a question framed as a rating scale method type of question was subsequently altered. The change involved a concretized statement posing a decision. Until applied to a decision many would have gone on thinking they had accurately expressed their values. After seeing the effect of a decision context and the exercise of a choice, none have disagreed with the indicated adjustment in their values representation.

9.4 Values Clarification

As reported in the chapter on the feasibility study, some participants explicitly stated ways in which the IMQOL protocol shapes their thinking. One indicated that the exercise revealed to him that he was not as bad off as he previously thought, after considering all the alternate possibilities. One was surprised to admit it but found the results of the application to be the most believable results. She claimed however that they only confirmed what she knew before. It was clearly an unanticipated way to arrive at the same conclusions, but it would help everyone else catch on to her thinking.

This overt evidence of learning supports the suggestion of Shiell, et al⁷⁰, that proposes preferences are not the same as values and that health state valuation should be viewed as formative more than definitive in many cases. Statements made are built on an assumption that well differentiated values are stable underlying principles. Furthermore, observable preferences are based on these underlying values. The suggestion is that individuals are able to settle on stable expression of values (preferences) where frequent opportunities exist to engage in evaluative activities. If health care were all headaches and colds, valuation of health states would be eliciting values of frequent experience with stable preferences observed. Most health states evaluated being something less common, the assertion is that minimal experience with severe disease states renders "few hav[ing] had the opportunity or the need to reflect on the value they would ascribe to the amelioration of such states."⁸³ Under such circumstances, individuals are purported to derive specific preferences from their basic underlying values by

some process of inference. The inference may vary from time to time until the individual has had enough opportunity for the expression of values to become well differentiated and stable.

Consistent with my observation above, Shiell, et al, point out the difference between the constrained choices used in economic surveys and the rating and ranking procedures of other disciplines, but suggests, “this may not be enough if respondents need help to construct and differentiate their values.”

“In educational psychology, the ‘values clarification’ literature suggests that people need to be taken through three steps in order to help them differentiate or clarify their values for fundamental goods⁸⁴. In addition to exercising a choice, these include affirmation and action. Each of these steps challenges the respondent to defend his or her values, in the process of which he or she comes to realize precisely what it is about the issue being discussed that is held to be of value and how valuable it actually is. The process of reflection provides people with the opportunity and the challenge they need to formulate and then reformulate their ideas. The respondent comes to ‘know’ himself or herself during the process and becomes reconciled with values which he or she may not have previously confronted.”

“It is unlikely that people would hold well differentiated values in [infrequently experienced severe disease states]. It is debatable, therefore, whether the questions that economists pose when conducting health utility interviews or surveys about willingness to pay or conjoint valuation elicit the respondent’s stable values or whether instead they elicit an under-considered and therefore potentially unstable opinion or preference.”⁷⁰

These authors suggest that the modest test-retest reliability of traditional utility assessment methods may be a function of the instability of the preferences measured more than a shortcoming of the measuring instruments. They site as supporting evidence a study in which one third of the respondents said that the first interview prompted reflection that directly caused them to change their answers at a second interview⁸⁵. They conclude their discussion by

suggesting this evidence endorses the proposal of economists that the process of values elicitation be used “to help respondents construct their preferences in the first place.^{86, 87, 70”}

The IMQOL method of assessment is a fulfillment of this proposal. It provides a structured means of linking values to preferences. The attribute scales elicited from familiar circumstances in familiar terms can be expected to represent well-differentiated values. The systematic fabrication of scales may represent novel articulation, but at least the fundamental construct underlying the volunteered construct is likely to be an expression of some forethought. The application of the scales articulated would then represent the inference by which preferences may be derived. The exercise of choice in the IMQOL elicitation and application is an opportunity to experiment and develop preferences facilitating the value settling process. Indeed, the greatest contribution of the IMQOL method may prove to be the role of values clarification more than values elicitation. The role of values clarification intensifies the indication for inexpensive and highly accessible IMQOL model elicitation tools rather than expert human IMQOL method administrators.

Chapter 10

Next Steps

Sprinkled throughout this thesis are suggested directions that could be taken in the further development of the IMQOL model elicitation and application. This chapter will focus the statements on specific ideas that I see as important as of this writing. Some are already initiated but have no reportable results. Others are more far reaching and represent only natural extensions to the findings of my investigation to date.

This chapter begins with the state of the prototype for a computer program to administrate the IMQOL model elicitation. The program is included in the chapter on Next Steps because of the infant stage at which program development stands. Numerous systematic biases, shortcomings and inconsistencies in human judgment have been documented⁸⁸⁻⁹¹. Completion of the programming of a prototype will be instrumental in evaluation studies, particularly for the sake of consistency in protocol administration and avoiding biases. Trying to avoid biases can be a two edged sword—for example, the use of index marks can be criticized for anchoring responses while other critics argue the accuracy of the human eye

without index marks. Discussion of the program will be followed by mention of ways in which I propose the protocol be evaluated following the asserted success of feasibility thus far. As with other health status and utility assessment instruments, this evaluation should consist of demonstrations of feasibility, reliability, content validity and construct validity. The chapter will conclude with a collection of ideas for the extension of the IMQOL model and elicitation.

10.1 SOLOMON Program

10.1.1 Implementation of IMQOL protocol

To implement the IMQOL protocol, I wrote a program titled the Systematic Obtainer of Language Overtly Multidimensional and Operational in Nature (SOLOMON)^{††}. It has been continually refined by a series of undergraduate assistants, most notably Richard Chen who has contributed significantly to the most recent upgrade. There is still work to be done before it can be said that the protocol is fully automated, but the present state of SOLOMON's development is sufficient to substantiate the promise of programmability. A procedural program that simply executes the protocol with a compliant understanding user would be an easy task. Fulfilling the requirement for a more intelligent program as set forth in the chapter on background and significance is a much more arduous task.

Since the IMQOL measurement is predominantly a restructuring of previously developed tools and since there are examples of those tools already in the form of computer programs, it is easy to anticipate the programmability of the IMQOL protocol. Examples that inspire the SOLOMON program are included here to further substantiate the claim that each component of the IMQOL protocol.

The SOLOMON program has not undergone any user testing at this time, but is fully operational. It would be inappropriate to suggest that the same results can be obtained with the computer program as with human administrated protocol until such tests are conducted. It is difficult to reveal the dynamic nature of a computer program on paper in a thesis. This is made

^{††} Similarity to the Hebrew Monarch of that name in I Kings 3:17 is not accidental; it was originally given this persona in recognition of the wisdom of listening with understanding.

worse with increasing use of powerful graphic manipulations and mouse events to control the program or data entry. SOLOMON has many elements that cannot be represented in paper media, such as the dynamic display of pop-up windows for extended labels of attribute scales or “draggable” text employed in rank ordering procedures. Furthermore, the warmth of color in the program is lost in the grayscale figures. Nonetheless, a detailed series of screen shots are presented in the Appendix.

10.1.1.1 Utility Assessment Program Precedents

Precedents for the implementation of utility assessment tools exist. Three programs, aimed at the medical community, are worth noting. These all emanate from recognized medical decision analysis investigators. *Gambler Examiner* is an implementation of analog rating scale and standard gamble methodologies for Windows 3.1. *U-Titer* is an implementation of standard gamble and time tradeoff for Apple MacIntosh platform. Representing the entry into World Wide Web Internet administration capabilities, *IMPACT* is designed to be accessed by any Web browser. It follows that if the IMQOL elicitation is built from the same underlying principles supporting these demonstrably programmable tools, it too is programmable and capable of Web administration.

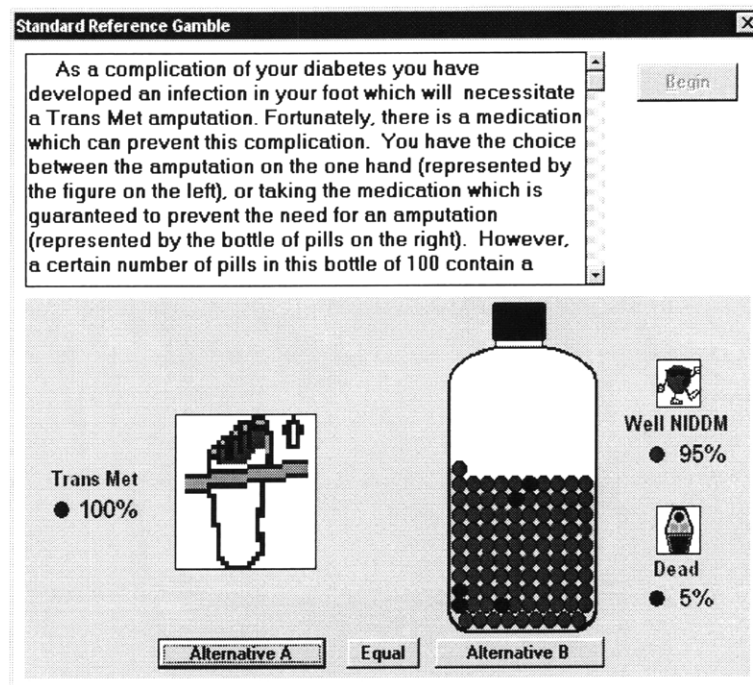
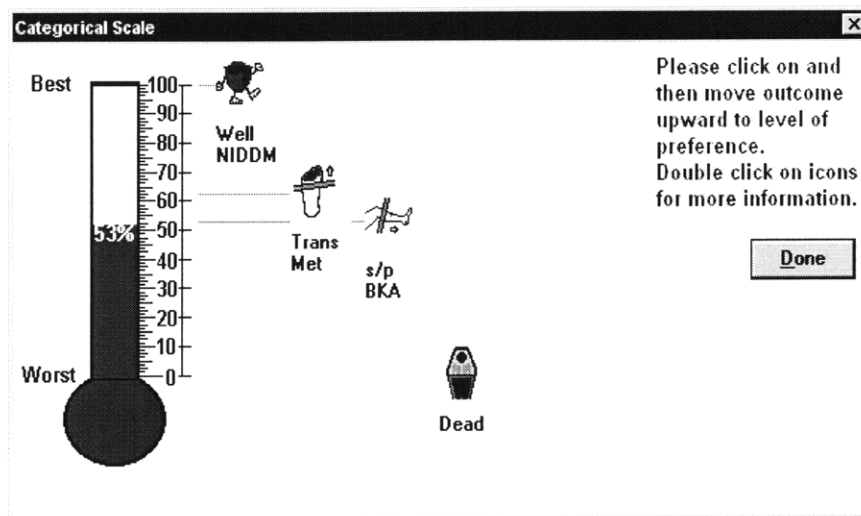


Figure 10.1 Screen shots captured from Gambler Examiner computer program. The top window demonstrates the capacity to program analog rating scale methodology. Bottom window demonstrates the capacity to program the standard gamble methodology of utility assessment. Normally the program is rendered in color. Pills in the bottle are an assortment of red and blue in the same proportion as the outcome probabilities, Well and Dead, respectively, for Alternative B.

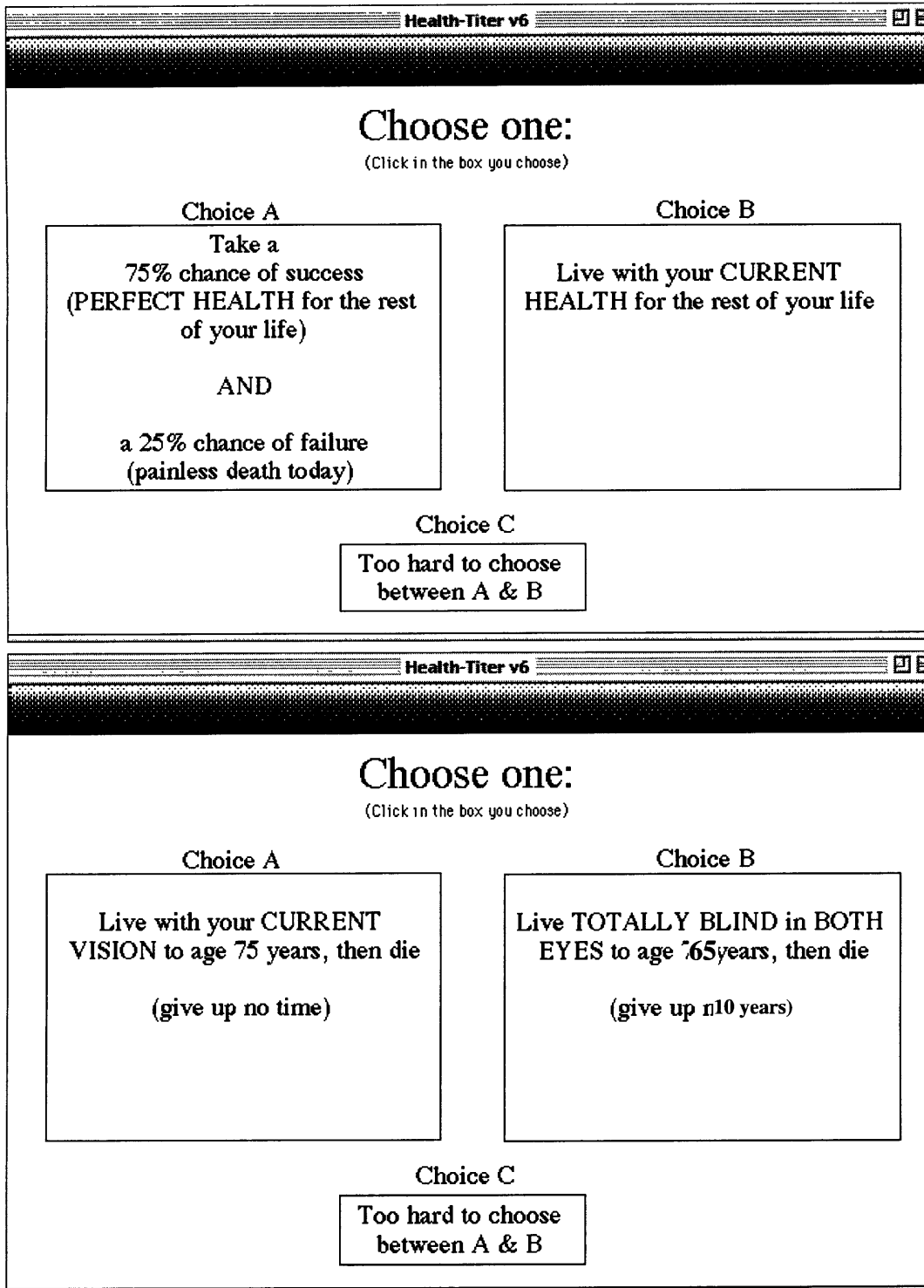


Figure 10.2 Screen shots captured from U-Titer computer program. The top window demonstrates the capacity to program the standard gamble methodology of utility assessment. Bottom window demonstrates the capacity to program the time tradeoff methodology of utility assessment.

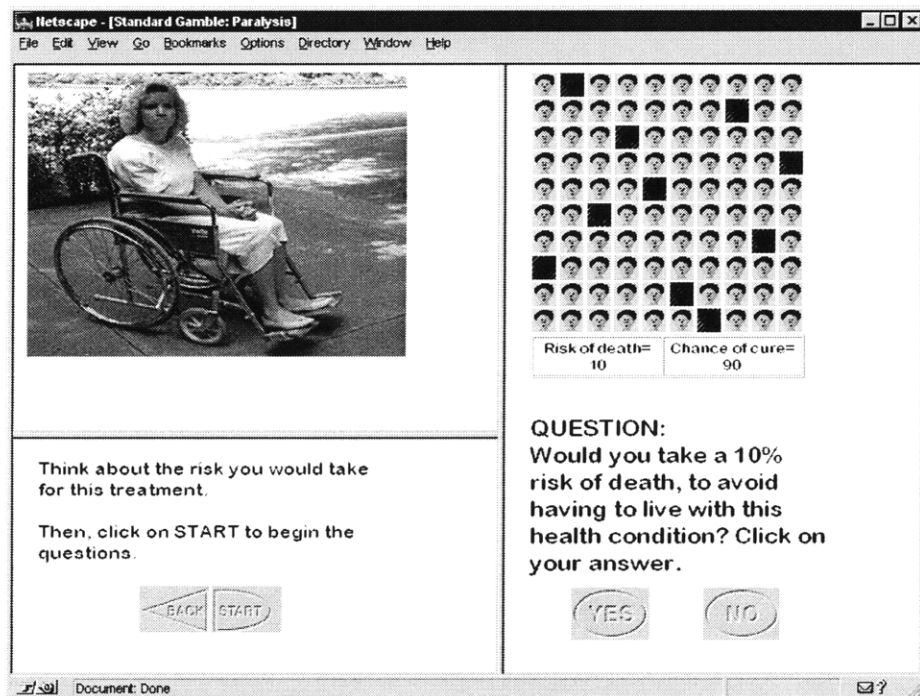
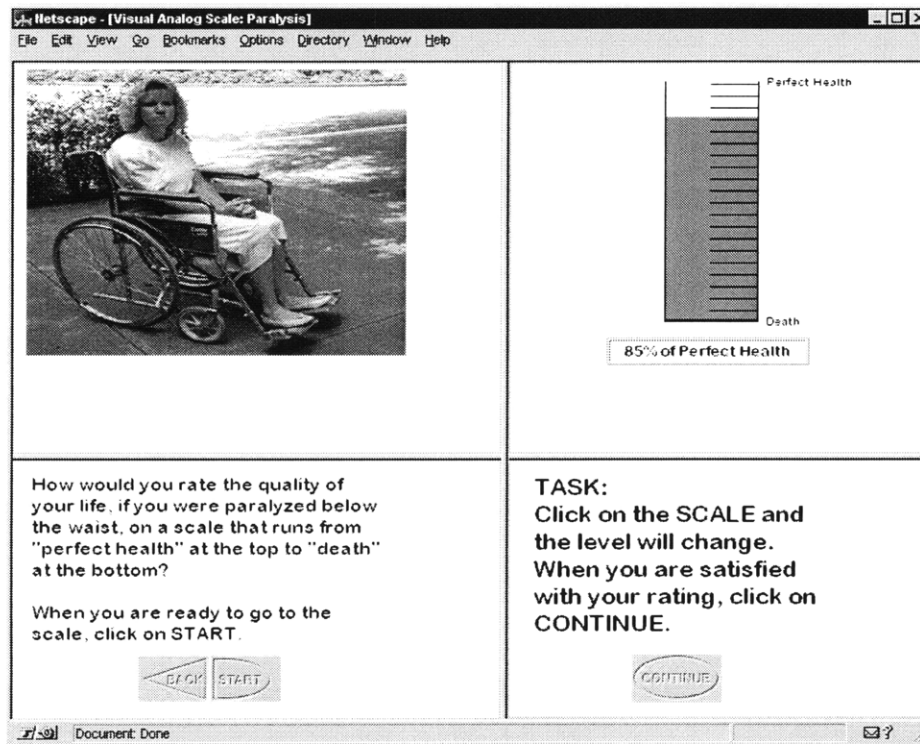


Figure 10.3 Screen shots captured from IMPACT computer program. The top window demonstrates the capacity to program the analog rating scale methodology. Bottom window demonstrates the capacity to program the standard gamble methodology of utility assessment. This program is run on the Internet with any Web browser.

10.1.1.2 Adaptive Program Required for Automation

Practical implementation of the IMQOL protocol is dependent upon computer programming. It is not likely that a protocol that requires up to four hours for trained analysts to administrate, even if only for the initial elicitation, will ever be affordable or prove cost effective. The extremely dynamic nature of the protocol makes it difficult to conduct in a paper survey format. The only hope for such a lengthy protocol is if it can be programmed. To be fully automated, the SOLOMON program will have to be able to detect the difficulties encountered as discussed in previous chapters. Then it will have to exercise adaptive strategies during administration (e.g., detecting an extreme value of length of life and switching strategies for eliciting scaling coefficients from a time tradeoff to a willingness-to-pay framework). Successfully automating the IMQOL method will add the benefit of accessibility when coupled with the Internet and Web TV.

Potential problems include the impossibility of complete automation. If it turns out that the interview and/or scoring process cannot be completely automated, effort can be made to automate as much as possible to minimize the need of human mediation. Computer aided mediation would still provide some advantage over a completely manual elicitation process. I have described a baseline approach to values elicitation and will, below, propose a hypothesis based validation in a systematic experimental framework which shows promise and a preparedness to accommodate problems or entertain alternatives which fit the same framework.

10.2 Evaluate Utility for Clinicians

A study is planned to establish the operational quality of the IMQOL models for clinical use. The first investigation will simply ask members of a nephrology division to score the same four states of health used in the feasibility study described in this thesis. The same four descriptions given the patients will be used to define the outcomes the clinicians are being asked to evaluate. The task will fundamentally be the same as that described in the chapter on application of the IMQOL model. The clinician will be asked, as a consultant, to assist the hypothetical patient in scoring the quality of life with an IMQOL model. A few representative IMQOL models will be selected for this purpose, keeping the identity of the patients unknown

for purposes of confidentiality^{††}. Posing as the patient, I will be able to answer some questions the clinician may want to ask to qualify their judgements or recommendations. Having conducted the original interviews, I am likely to have the necessary insight to enable an undemanding pilot investigation. After a preliminary investigation to determine the length of time required per evaluation and suitability of format, a more extensive study can be planned with a larger number of clinicians. Comparison of clinician scores with actual patient scores will be possible. The clinicians will be asked to qualitatively comment on the exercise after completion. The objective will be to determine qualitatively if they think the information is useful and if it would alter their clinical recommendations. A more objective measurement of clinical impact is proposed below.

10.3 Scope of Utility

Since the IMQOL elicitation does not require a specific health state, it will be interesting to determine how well the clinician can use the attribute scales to assess outcomes in non-targeted domains. This is distinct from employing IMQOL models on behalf of the patient without requiring the patient's help in performing the assessment. A generic IMQOL model should be useful without the patient's help in medical domains beyond those targeted at the time of elicitation. Just how far this can be successfully accomplished should be investigated. For example, the IMQOL models elicited from the dialysis patients may prove clinically useful for dialysis treatment modality decisions, but the same models might be employed to assess the outcomes of stroke. Should this be possible with little to no help from the patient, a truly powerful instrument will have been devised.

10.4 Evaluate Psychometrics

As proposed in the chapter on background and significance, psychometric precedents exist for establishing the feasibility, reliability, and validity of value assessments. As suggested

^{††} Revealing alleged values that have not been proven reliable without patient consent would be unethical.

above, feasibility of computer programming must be completed and the feasibility in terms of clinically operational quality for the IMQOL model must be established.

10.4.1 Reliability – Repeatability Over Time

As mentioned in the chapter on background and significance, evaluating reliability definitively is problematic—any elicitation that is exhaustive at every application is at risk of learned effects, undermining test-retest comparisons most commonly used for reliability testing. However, tracking IMQOL models over time is possible so long as the protocol is feasible. The participants who produced the results used to demonstrate feasibility have already consented to additional interviews. Comparison of future interviews with those reported here will be useful in supporting arguments for reliability. Of particular interest will be the degree of modification observed in the IMQOL attribute scales (hypothesized to represent well-differentiated values) and the degree of modification in application of the scales to health state descriptions (hypothesized to represent preferences inferred from the values). A low rate of change in model constructs coupled with a high rate of change in application to health states would support my characterization of IMQOL models as a link between values in familiar terms and preferences for medical outcomes.

At this point, we can begin to capitalize upon the advantages of a multidimensional (decomposable) model. Returning to the epistemology of values and preferences of the chapter on serendipitous findings, we can hypothesize that the IMQOL representation of values is like that of natural intelligence. The taxonomy suggested by the IMQOL representation of values poses two types of elements: the nominal descriptors and the quantifiers representing weight and priority. In fact, the possibility of distinct maturation processes for nominal elements and the quantifiers of the value structure lead to an interesting question. Which elements require rumination to become stable values? Do they mature independently? We can explore the stability of individual components of the values structure. With this suggestion, we enter the branch of artificial intelligence (machine representation of values) that pursues the understanding of natural intelligence.

10.4.2 Construct Validity

To support construct validity, the following study has been proposed. English speaking adult patients (~100) of participating nephrologists of the New England Medical Center's Dialysis Center will be invited to participate in the study while undergoing dialysis. The interview structure and schedule will be explained to those responding favorably to the invitation. Consenting participants will respond to the SF-36 survey in either paper or computerized form. Patients will additionally be asked to answer the six questions constituting the Satisfaction with Decision instrument published by Holmes-Rovner, et al³⁸, modified to refer to the decision to undergo the specified length dialysis at the planned interval. Participants will be assigned to one of three groups. One group will have the results of the values clarification interview and outcomes scoring given to their clinician who will also score the outcomes. A second group will be informed that the results of their values clarification interview will not be disseminated to the nephrologist, relying on routine health care communication to proceed normally. A control group will not undergo the values clarification interview, but all other outcome measures of health status, compliance and satisfaction with decision will be monitored. Thereafter, the appropriate group members will be interviewed as described in the discussion of the values clarification methodology. Administration of the values clarification interview will either be conducted in person by myself or mediated by laptop computer (or a technological equivalent, e.g. web TV).

Participating nephrologists will construct a list of potential outcomes for the value-assessed patient group participants to use in preliminary scoring. This list will enable patients to use freshly elicited attribute scales to preliminarily score the relevant outcomes for their personal circumstance. The patients whose values are to be used by clinicians to score outcomes will be so informed. These preliminary scores will be provided to the clinician for consideration in their evaluation of the outcomes with patients' attribute scales.

Patients with a history of dialysis at the center prior to the study and patients initiating dialysis for the first time will be visiting the center during the study (the NEMC dialysis unit experiences a net average of 1 new patient per month)—both with some treatment plan already in place. This study will not be able to collect pre- and post-decision data from patient

decisions that are made prior to the study; however, there are changes and decisions that will occur during the study period. All patients are currently re-evaluated monthly regarding their dialysis prescription, calcium supplementation, and compliance with phosphorous and fluid intake management. Assessment results prior and posterior to these changes and associated decisions can be collected and analyzed.

The attending nephrologist will document the treatment plan either as a matter of course in medical record keeping or on a separate form for this study. This will be done by the nephrologist before any knowledge of the values clarification interview results. Additionally the nephrologist will specify the potential outcomes for this patient and communicate the relevance of the list elements to the patient so that the scoring exercise is not cast as the bearer of any new prognostic implication by the patient. The standard utility assessment for these outcomes will be performed for comparison to the values determined with the proposed method. This list of outcomes will be modified by the clinician as required for any decisions that become necessary in the duration of the study. Preliminary patient scoring of outcomes will be acquired on subsequent visits as required by the advent of novel outcome potentials.

Patients' compliance will be monitored from entry in the study throughout its duration in the following dimensions from the medical record: the original recorded treatment plan and modifications over time, the intervals between dialysis and the duration of dialysis at each visit to the center. Prescription refills, OTC purchases (e.g., Calcium Carbonate supplements) and pill count data will be collected at each visit. Dietary recommendations made by the center dietitian will be monitored with a short self-administered questionnaire or as part of the computerized interview, to be written in one form for all patients in the study, which will monitor the number of meals or snacks in which the patient complied with recommendations.

The attending nephrologist of the experimental group whose values are to be considered will be given the results of the values clarification interview and asked to score all the potential outcomes with the patient specific attribute scales. The clinician scoring exercise will end with two questions: (1) Does knowing the values of the patient as portrayed in these scales change your treatment plan? How? (2) Does knowing the values of the patient as portrayed in these scales change what you intend to discuss in your next consultation with the patient? How? (3)

If the knowledge of patient values so represented did not change your treatment plan, do you consider it useful? How?

After the next patient-clinician consultation in which treatment decisions and patient value specific scores for possible outcomes are discussed, the patient will complete the satisfaction with decision instrument again. Health status reflected by the SF-36, compliance measures described above, and response to satisfaction with decision questions will be monitored on a monthly basis. The nephrologist will be interviewed or complete surveys regarding the impact of their knowledge of the patient values at six-month intervals.

Over time, it is anticipated that some of the participating patients will face dietary modification, dialysis duration, dialysis frequency, vascular access, and transplantation decisions. In the advent of any change in the treatment plan, the nephrologist will be asked to score any newly relevant potential outcomes that may result. Concomitantly they will be asked to complete the questionnaire regarding the impact of values knowledge if the patient is in the appropriate experimental group. These sub-populations delineated by specific decision types and contexts can be evaluated in the analysis of the data collected in the course of the proposed research. Analysis of variance on the variables of co-morbidity and renal disease etiology or health status severity can be anticipated where sufficient cohorts result from the study.

10.5 Extensions to the Representation

There are those who would argue that the IMQOL method does not measure utility values if it does not ascertain whether there are interactions between the multiple attributes. I have been careful to present no prohibition of more complex models for recomposing the attribute scales in a preference model. However, before I leave the simplicity of linear models I would point out that if 1) the definition of utility assessment is the quantification of preferences for medical outcomes and 2) there is no gold standard for measuring utility, then we still don't know what it is exactly that we are measuring with utility assessment. While it is true that a linear approximation is measuring something different than utility that is measured by those who concern themselves with justification of additive or multiplicative models, there is no basis to decide which one is measuring the quality of life. The experience comparing *pessimistic* and

optimistic strategies for scaling coefficients reported in the earlier chapter on serendipitous lessons raises an issue. That issue is whether the magnitude of difference caused by interactions between attributes is significant to the overall measure of quality of life. Is the signal to noise ratio large enough not to be dominated by noise or unavoidable anchoring and framing effects?

Linear models are not without precedent. For information concerning research that has been done concerning the use of linear models to determine utilities and values for decision making the reader is referred to Von Winterfeldt and Edwards⁶⁹.

The following two subsections cover non-linear models, an extension of the elicitation used in the IMQOL method that might supply hierarchical relations by another means. An alternative analytical approach to scaling coefficients for linear models and an extension of IMQOL application follows.

10.5.1 Non-linear Multiattribute & Hierarchical Models

Should the linear model prove inadequate or oversimplified, there is nothing prohibiting the extension of IMQOL models with non-linear models. If the additional cost of increased protocol steps is cost effective, more sophisticated models may be used. Previous work with automated utility reasoning systems³³ provide a foundation for exploring more complex models. This reasoning system does not provide suggested attributes for the multiattributed model; it requires that candidates be supplied for evaluation. IMQOL elicitation provides the candidates for such evaluation.

10.5.2 Elaborating the Recomposing Model With Relational Theory

The ideas for eliciting the dimensions of the IMQOL model have been largely influenced by the concepts of *personal construct theory* and *repertory grids*. The potential to employ *personal construct theory* to elaborate a hierarchical IMQOL model of superordinate and subordinate constructs has been alluded to. The use of triads to elicit bipolar constructs is inspired by the first steps of *repertory grid* construction and analysis; none of the customary analytic elements of that perspective have been explored. Data from "the Semantic Differential"⁹² and from *repertory grids* have been analyzed with a *relational theory* methodology proposed

by Bandler and Kohout⁹³. *Repertory grids* have been used to elicit domain knowledge for expert systems^{61, 94}. Constructs of physiotherapists have been utilized to evaluate Parkinsonian patients⁹⁵, as well as other diagnostic data and patient management processes^{96, 97}. Constructs of urban residents have been used to evaluate "quality of life" looking for non-symmetrical dependencies and implications among constructs³⁴. This would lead one to anticipate the ability of using similar methodology for the evaluation of "state of health" or "quality of life" in medical domains.

10.5.3 Conjoint Analysis

An alternate approach for determination of the scaling coefficients that could be substituted in the IMQOL method is known as *Conjoint Analysis*⁹⁸. It provides a way to determine coefficients of an additive utility model accommodating several attributes with different numbers of levels. It was first used to examine consumer weightings of product attributes based on their rank ordering of a set of profiles. It has since been united with other outcome measures and analytic methods including least-squares regression. Typically, the respondent is asked to evaluate a series of multi-attributed cases given the profile of each in terms of the attributes. Systematically varying the attributes and measuring the impact on the respondent's evaluation is the heart of the analysis. It could be argued that the determination of weights in the IMQOL method is a simple form of conjoint analysis, although it was not designed so knowingly. *Conjoint analysis* can minimize the cases requiring assessment to determine the coefficients by employing a fractional factorial design⁹⁹. This has been applied to the study of how faculty members, surgeons, referring physicians and medical students weight observed characteristics captured in decision strategies¹⁰⁰⁻¹⁰⁴. The experience of this research could shed light on efforts to refine the IMQOL method of estimating the relative contribution of the attribute scales to the overall quality of life score while minimizing the scenarios that must be evaluated by the assessor.

10.5.4 Lens Model

Another use for the IMQOL model arises when integrated with the concept of the lens model¹⁰⁵. The lens model provides a framework for comparing and contrasting two linear models for the same set of factors (cues). It is used to compare the weights of an interpretation of multivariate information with the actual weights in regression models of observed data. The lens model has been applied in medicine contrasting the multiple attribute judgements of physicians with linear regression models of observed outcomes in those same dimensions¹⁰⁶⁻¹⁰⁸. The contrast between the weights used in judgement and the weights in actual outcomes can be used as a measure of accuracy. The percentage of the variance explained by a regression model for the weights in repeated judgements by one person can be used to measure the person's consistency.

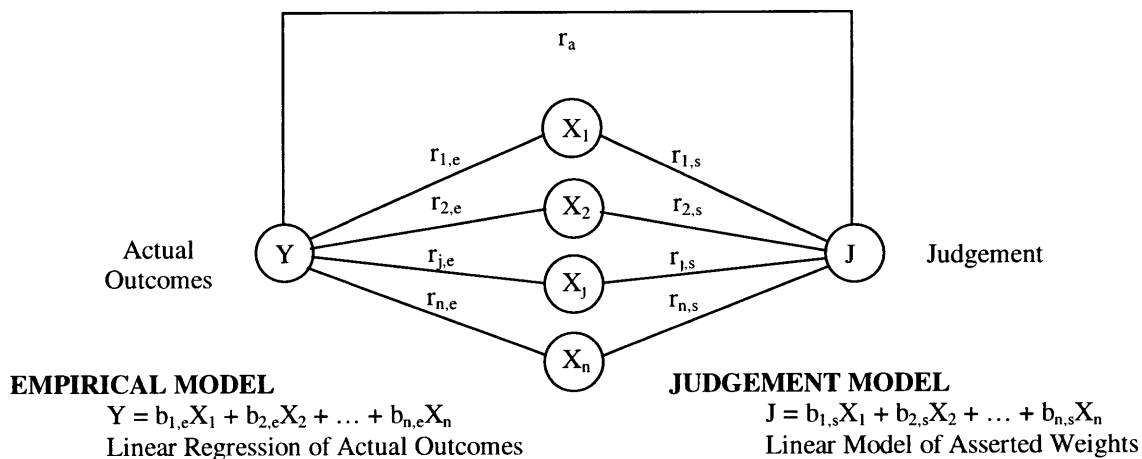


Figure 10.4 The lens model. A linear regression of actual outcomes is calculated on the left side of the lens and a linear model using the judge's assertion of factor weights is calculated on the right side of the lens. X_1, \dots, X_n represent the variables (cues) of a multivariate judgement; r_{1s}, \dots, r_{ns} are the correlations between the individual variables and the judgement made; r_{1e}, \dots, r_{ne} are the correlations between the individual variables and the actual outcomes; and r_a is the correlation between the judgement and the actual outcomes.

The variables of the lens model could be defined as the attribute scales of the IMQOL model. The scaling coefficients of the IMQOL model could be assigned to the arcs connecting to the outcomes representing the linear model. Some decision is required for the representation of the outcomes in some measurable terms, be that the quality of life ascribed, the choice of

outcomes from among a list of possibilities (e.g., hemodialysis versus peritoneal dialysis) or something else that can be measured as an assertion and empirical observation. If a counseling health care provider is compelled to suggest alternate scaling coefficients for the variables present, the comparison between the provider's model and the patient preference model could be made explicit. This would involve using the scaling coefficients of the patient's IMQOL model on one side of the lens model and the health care provider's recommendation on the other side. Alternatively, the provider's estimation of a patient's preferences could be compared with those empirically observed or measured by the IMQOL method. This might be viewed as the contrast between a prescriptive and descriptive model wherever it is possible to consider either side of the lens model as the prescriptive side. Another vantagepoint could be the contrast of the descriptive patient preference model with the linear regression of observed outcomes where data is available to produce such a regression. This would require a context of repetitious decision regarding quality of life. One example of this might be the recurrent nursing decisions of a hospice patient for whom the means of minimizing discomfort is a daily task. In a theoretical pursuit of utility assessment technique evaluation, we might compare the weights of methods of a modified IMQOL method employing only rating scales with the IMQOL model based on decision framed elicitation. Furthermore, a contrast between the patient's ascribed values by IMQOL elicitation and the linear regression of observed decision outcomes could prove very informative for both the provider community and the patient.

A more complex integration of the two paradigms would involve substituting the application of the IMQOL model to a set of health status features (as cues) for the linear model on the right side of the lens model. This would result in a "lens" generating a quality of life measure on one side with the linear regression of those features on the left. The outcome variable on the left would be the quality of life predicted by some direct interpretation or assessment in terms of the health status features. This lens model differs from the former suggestions in that the variables in the center of the lens model are some set of standardized health status features rather than the dimensions prescribed by an individual's IMQOL model. The concept depends upon the applicability of the IMQOL model's attribute scales to the list of features embodied in the lens model. The data necessary to generate a linear regression for the

left side would be much easier to come by in this case, however, the question of which health status features to include remains unspecified. Using the axes of well-studied health status measures such as the MOS SF-36 to specify the variables of the lens model could be fruitful. For example, employing this approach to investigate the adequacy of the SF-36 for measuring end-of-life health status in hospice patients could illuminate ways in which terminally ill patients are underserved by the general measure.

In his paper on the use of the lens model to analyze physicians' decisions¹⁰⁸, Wigton suggests that "technical problems remain to be mastered in constructing linear models of medical judgement." He says these problems "include how to select the correct variables, how to provide a selection of variables broad enough to accommodate individual variations in strategy, how to model intercorrelated variables, and how to characterize and aggregate individual strategies." By analogy, the same could be said of patients regarding the expression of their values as preferences for outcomes. The IMQOL method addresses the problems cited. It provides a systematic approach to the identification of variables accommodating individuals explicitly, enabling the characterization of strategies. The only problem not directly addressed is how to model intercorrelated variables, but having identified the variables, application of more complex multi-attribute utility theory is made possible.

Furthermore, Wigton states that "cognitive feedback which is feedback to the learner of the judgement model derived from previous decisions, is highly effective for teaching complex judgement tasks." The IMQOL method elicitation of preference models, coupled with a lens model contrasting declared and observed values, provides a framework for providing feedback to the patients for learning how they can best instill their values in medical decisions. Seeing how their expressed values contrast with the values demonstrated by decisions and behavior grants the necessary insight to enable well-differentiated values. The result can be a more stable expression of values, an improved resolve in decisions and an increased confidence in the choices made.

Chapter 11

Conclusion

A representation for Individualized Multidimensional Quality Of Life (IMQOL) values has been described along with a fitting means to elicit the parametric and non-parametric elements. The representation is a list of relevant issues cast in the form of ordered nominal scales. The list is arguably complete for the individual's current frame of mind and the scales are comprehensive. These attribute scales are fully specified in terms of the patient's own words, which others can use. Application of the IMQOL model has been demonstrated and empirical results of application have been shared in attempts to convince the reader of the merits of augmenting traditional quality of life assessments with the IMQOL protocol.

As a concluding demonstration that something valuable and revealing is captured by the IMQOL protocol, I suggest an experiment you may perform personally. Often a personal experiment is the most convincing. If you have not already done so, take a few minutes to browse the IMQOL models acquired from actual interviews with individuals found in the Appendix. Then ask yourself if the pages are still simply words organized in grids on a page or has each page (representing a single individual) become personified. Do the words and dimensions of health asserted as relevant draw a personality picture that leaves you with expectations? Do you have a head start on knowing what will motivate this persona and what

may not? Are you prompted to ask any specific questions? Have you any arguments or counterpoints you wish the represented individual would consider?

Moments before you had nothing but pieces of paper in your hand. You may have reservations about how far you will take the impressions you get from these IMQOL models, but you have most likely been moved to form an opinion. Details may or may not be lacking, but details you may want have been identified—the exercise has focused your next step. Assuming the person has claimed this is what he or she wants you to consider as his or her values, you have either been impressed with the completeness of the individual's model or been left with a refined idea of what you would like to see that is missing. You have gained an insight into the scope of the person's view of healthiness by the extremes presented for each dimension. To the degree you feel equipped to use these attribute scales on behalf of the individual, you have an operational tool by which you can express your perceptions of health state descriptions in words that are not your own. You are privy to at least one round of this person's attempt to apply what he or she knows of his or her likes and dislikes to states of health in an expression of preference. The IMQOL model directly compares his or her personal health to those acquaintances brought to mind in the elicitation process. If the IMQOL protocol is fully applied, the model has been qualified by application to one or more health state descriptions and refined as needed. The person represented has thus verified its completeness, comprehensiveness and operational qualities and decided, with that experience, what he or she thinks as well as what he or she wants you to think about his or her values. Those having thought through the IMQOL elicitation process may have questions of their health care providers that were not clear before the exercise.

As a final testimony of the discovery potential for some participants of IMQOL protocol for quality of life measurement, let me tell you of an event that occurred in the final days of writing this thesis. While feverishly attempting to print a final draft, my phone rang. It was one of the people interviewed for the study of post-stroke quality of life assessment. He was one of those who volunteered to participate in the earliest pilot of IMQOL elicitation after completing the standard gamble assessments. The traditional assessment was done in person at the emergency department but the IMQOL protocol was conducted by phone because of the

time involved, hence providing the caller with my phone number. I remembered that this individual had a restricted schedule that made it extra work to complete the interview. However, he was taking the time to call me six months after his interview was over to find out how the study was going. He was excited to know that this thesis was being written; he was also excited by the news of the ensuing study of dialysis patients. He did not take much time. He simply declared that he was happy that the research was going well and wanted to encourage me to continue. He wanted me to know that the interview had caused him think about some things that he would not forget. He was going out his way to thank and encourage me. It is not frequent that subjects of any research go out of their way to do such things. This is more remarkable when one recalls that utility assessment traditionally is of such gravity that it ends with the patient in tears.

Although such testimony is not a rigorous or measurable evaluation, it represents the profound attitude of interest and enthusiasm held by numerous participants. I cannot argue with the contention that this patient satisfaction might have more to do with the personal attention and time spent listening than the IMQOL protocol. It remains possible that patients would give more rave reviews of traditional utility assessment if a decision analyst simply spent an hour just listening to the patient before doing utility assessment. Just getting rave reviews is not the point. The point of this thesis is that additional information has been acquired regarding quality of life in the patient's perspective. The point of this anecdote has been that the contribution of the IMQOL model elicitation has been accomplished with the patient's blessing. Indeed, listening to the patient is good. However, without a systematic approach, we are back to where we started with ad hoc approaches to understanding patients.

Conclusions of a more rigorous basis may be drawn.

IMQOL models elicited from patients thus far reveal a diversity of personalities with a common architecture and a repeatable systematic approach. In contrast to the output of traditional utility assessment as exemplified by the first three figures of the discord chapter, the IMQOL models preserve individuality while adding information. As prescribed by Keeny and Raiffa⁵⁹, the decomposable model that is complete and operational is less ambiguous. What remains is to establish the reliability and validity of the IMQOL protocol output. However, the

feasibility of using the terminology of the individual patient to capture and express the individual's particular, perhaps unique, value system for healthiness has been established, providing a convenient language for discourse. This research has established that it is possible to use individualized patient terminology to represent the values for outcomes in medical decision models.

The IMQOL protocol for elicitation is a rigorous and reproducible way to determine patient values in ways that others can use them. The preference models elicited are not bound by any specific disease context. They are generic in nature and can be reused for application to multiple health care decision contexts as long as the patient's values remain unaltered. Being disjoined from specific disease context enables the building and maintenance of IMQOL models in advance of need for application. The promise of programmability substantiates the expectation that widespread access and administration can be automated through computers and Internet resources.

The findings of this research substantiate the fundamental premise that patients can be treated as experts on their own values and preferences. With the structure of the IMQOL model and its systematic elicitation, the values of the patient can be factored into the evaluation of quality of life on an individual basis. It supports the assumption that traditional frameworks of utility assessment can be applied to constituent elements of values and preferences, to patient-elicited attributes of healthiness as well as to physician-described clinical health states. As such, this thesis is ill applied to any claim that something better than accepted utility assessment methods has been revealed. It would be more accurate to appraise this contribution as an augmentation to traditional methodology. By supplying more information, it delivers a means to hypothesize explanations for more direct measures of traditional utility assessments.

The IMQOL modeling paradigm was not extensively tested in a variety of domains. However, the models elicited were complete enough that participants were, for the most part, willing to agree that there was nothing left out. The models were operational enough to allow quantitative values to be derived for all states of health evaluated, in surprisingly satisfactory ways in some cases. The models were comprehensive enough that no health state explored was incapable of assessment, even when unfamiliar. It must be kept in mind, however, that the

extent of study to date has predominantly been feasibility. Substantially larger cohorts of patients should be studied to assure there are no effects from the sample size in this preliminary work. Investigations designed for reliability and validity testing are required before any universal application is justified.

Exception to completeness could be remedied in subsequent elicitation that allows refinement (not yet provided to participants of this study). Exceptions regarding operational quality and comprehensiveness were limited to attempts to assess states considered worse than death and it was shown how this depends upon the interpretation of the application task. If we ask patients to assess the quality of life for states that include death, we must be clear about how we want them to consider post-death elements. If we wish to evaluate only life up to death, instructions to that effect should make it clear. Evaluating health states with post-mortem aspects may be possible with IMQOL but there are unsettled issues.

The SOLOMON program demonstrates that the IMQOL method for elicitation can at least be aided by computer programming. The ways in which the patient can be confused by the questions of the protocol and what it takes to get successfully back on track remain poorly defined and characterized. To capture errors made and avoid confusion, adaptive loops will need to be embedded in the program, coupled with reliable detection of patient confusion. User training by instructor or self-instructed tutorials may lessen the demand for computer program intelligence. Limited problems can arise from an elicitation of such structured nature. It is easy to remain optimistic about the full automation of the IMQOL protocol. Because of the success in these preliminary studies of feasibility, we have insights that now guide the process of programming both detection and dynamic adaptation to user behaviors. Furthermore, the programmable nature will facilitate the administration of the IMQOL method of quality of life assessment for the purposes of reliability and validity.

Qualitative results of the feasibility study for the IMQOL protocol, administered by expert interview, are rewarding. Things both positive and negative have been learned in the process of this study that can be applied to refining the protocol and as specifications for automation. Specific indicators that may serve as risk factors for failure to benefit from the protocol have been suggested. Quantitatively, the mean summary numeric values of IMQOL

model elicitation and application have been shown, at least in the case of one hospital's dialysis patient population to be insignificantly different from traditional measures (albeit, without the statistical power of a large sample size which is impractical without automated administration).

Serendipitously, attention to individual detail has paid off in the illumination of quality of life assessment issues that should reflect back on traditional utility assessment. Specifically, outcome states for quality of life assessment should be chosen carefully and systematically derived to facilitate application of assessment tools and make valid comparisons. This is especially true when it comes to assessing states of health near to or worse than death. Without disambiguation, risk remains of grave states of health being evaluated in desperate frames of mind with erroneous results. Nonetheless, the composition of multiple attributes of health for one individual appears relatively insensitive to contextual differences in the ranges observed so far. This suggests it will be difficult to justify taxing the elicitation process with further complexities.

The observation of learned effects is particularly interesting. Others have demonstrated that a mere power transformation accounts for the difference between rating scales and the standard gamble or time tradeoff methods of utility assessment¹⁰⁹. The observation of migrating values with the movement from rating scale assessments to other methods suggests that more than a power transformation is involved. More investigation into this phenomenon would be useful in understanding the difference in what is being measured. Of a more practical nature, the role of quality of life assessment as a discovery tool has been engendered in the IMQOL protocol. The call by Shiell, et al⁷⁰, for viewing quality of life assessment as values clarification, aiding the stabilization of values that are expressed as preferences, has been addressed. If we want to know and use the patients' quality-of-life values in medical decision making we will have to help patients find out what those values are and how they play out in preferences for medical outcomes in words the patient fluently understands.

Finally, the potential next steps of investigation have been laid out. Overlapping the establishment of reliability and validity, the experimental variable of individualized patient perspective inclusion in decision making should be evaluated. We can also measure the outcomes of patient care in the terms of self-assessment using their own IMQOL models.

Correlation of trends in the IMQOL model assessments with other status measures can be evaluated. Conclusions may be drawn in terms of multiple traditional status measures, namely: impact on clinical choice of treatment, patient reported satisfaction with decisions, patient compliance, health status, disease and treatment category, and co-morbid complexity. Testing hypotheses through analysis of the correlation between the outcomes measured will help us understand the way the patient is thinking and reasoning. It will substantiate claims regarding the importance, the nature and the impact of patient values in health.

If the interactions between attributes of health are not measured, it may be argued that the IMQOL assessment does not measure utility, if utility is defined as the thing—whatever it is—that is measured by standard gamble or time tradeoff methods. As I have argued, with no gold standard for measuring quality of life values, we have no way to know which method is measuring the most desirable thing. However, we now have an extendible representation and fitting elicitation method that treats the value holder as an expert on what should be measured. Further investigation to establish the reliability and validity of the IMQOL method is warranted. The augmentation of traditional utility assessment and health status measurement may ultimately be more advantageous than supplanting traditional methods. In a Lens model conceptualization, the IMQOL model may be useful in a process of constructive convergence with either traditional utility assessment or health status measurement. What Hammond says of medical diagnostic judgements could easily be said concerning quality of life assessment:

“The author's purpose is to urge the constructive convergence of two current judgment and decision-making research paradigms. ...the differences between the two research paradigms thus becomes apparent; they speak to different problems and appeal to different criteria for evaluating performance. Bringing the two into a constructive relationship to one another, however, will not only double the store of knowledge regarding diagnostic judgment and decision making, but also enhance efforts to achieve a cumulative discipline. Isolating these research paradigms from one another--as is done now--stifles theoretical generality, fragments knowledge, and confuses medical decision makers.”¹¹⁰

With an IMQOL model in a Lens model juxtaposed to utility assessment, health status measures or observed behavior, one could expect constructive convergence to enhance the store of knowledge regarding quality of life judgment and decision making.

When reliability and validation of any refinement of the Individualized Multidimensional Quality of Life protocol is established, it will give physicians a systematic, rigorous, reproducible way to take patient values into account. If, in fact, the clinician who has ranked available treatment strategies on a biomedical or functional status measured basis, can be convinced that the quality of life, as viewed with the patient's values, significantly changes the ranking, the hypothesis of this research will have been substantiated in the most powerful way.

References

- ¹ RB Deber, N Kraetschmer and J Irvine, "What role do patients wish to play in treatment decision making?," Arch Intern Med 156.13 (1996): 1414-20.
- ² RG Justin, "Adult and adolescent attitudes toward death," Adolescence 23.90 (1988): 429-35.
- ³ AM Stiggelbout and GM Kiebert, "A role for the sick role. Patient preferences regarding information and participation in clinical decision-making [see comments]," Cmaj 157.4 (1997): 383-9.
- ⁴ J Von Neumann and O Morgenstern, Theory of Games and Economic Behavior (Princeton, NJ: Princeton University Press, 1947).
- ⁵ IB Wilson, ML Green, L Goldman, J Tsevat, EF Cook and RS Phillips, "Is experience a good teacher? How interns and attending physicians understand patients' choices for end-of-life care. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments," Med Decis Making 17.2 (1997): 217-27.
- ⁶ EB Bass et al., "How Well Do Health Professionals Understand Patient Concerns About Dialysis Treatment Options?," Medical Decision Making submitted (1998).
- ⁷ R Hamburger and e al, "A Dialysis Modality Decision Guide Based on the Experience of Six Dialysis Centers," Dialysis & Transplantation 19.2 (1990): 66-69.
- ⁸ D Breckenridge, "Decisions regarding dialysis treatment modality: a holistic perspective.," Holist Nurs Pract 12.1 (1997): 54-61.
- ⁹ P Groome, T Hutchinson and S Prichard, "ESRD treatment modality selection: which factors are important in the decision?," Adv Perit Dial 7 (1991): 54-56.
- ¹⁰ A O'Conner and L O'Brien-Pallas, "Decisional Conflict," Nursing Diagnosis and Intervention, eds. GK McFarland and EA McFarlane, vol. Decisional Conflict. In: McFarland GK, McFarlane EA, eds. Nursing Diagnosis and Intervention. CV Mosby, Ontario, Canada: 1989, p 573 (Ontario, Canada: CV Mosby, 1989) 573.
- ¹¹ KB Bayley, MR London, GL Grunkemeier and DJ Lansky, "Measuring the success of treatment in patient terms," Med Care 33.4 Suppl (1995): AS226-35.

-
- 12 H Bonarius, "The interaction model of communication: Through experimental research towards existential relevance.," Nebraska symposium on motivation, ed. J. Cole, vol. 24 (Lincoln: University of Nebraska Press, 1977) 291-343.
- 13 MA Gara, "Back to Basics in Personality Study - The Individual Person's Own Organization of Experience: The Individuality Corollary," The Construing Person, eds. James C. Mancuso and Jack R Adams-Webber (New York: Praeger Publishers, 1982) 45-61.
- 14 KB Meyer, DM Espindle, JM DeGiacomo, CS Jenuleson, PS Kurtin and AR Davies, "Monitoring dialysis patients' health status," Am J Kidney Dis 24.2 (1994): 267-79.
- 15 R Kaplan, "Health-Related Quality of Life in Patient Decision Making," Journal of Social Issues 47.4 (1991): 69-90.
- 16 AR Tarlov, JE Ware, Jr., S Greenfield, EC Nelson, E Perrin and M Zubkoff, "The Medical Outcomes Study. An application of methods for monitoring the results of medical care," Jama 262.7 (1989): 925-30.
- 17 KB Meyer, "The outcomes of ESRD and its treatment," Adv Ren Replace Ther 2.2 (1995): 101-11.
- 18 G Nichol, HA Llewellyn-Thomas, EC Thiel and CD Naylor, "The relationship between cardiac functional capacity and patients' symptom-specific utilities for angina: some findings and methodologic lessons [see comments]," Med Decis Making 16.1 (1996): 78-85.
- 19 KA Zug et al., "Assessing the preferences of patients with psoriasis. A quantitative, utility approach," Arch Dermatol 131.5 (1995): 561-8.
- 20 A Fontaine and F Larue, Physicians' Assessment Of The Symptoms Experienced By Their Patients Is It Reliable? Poster Abstract. (Toronto, Canada: 1996).
- 21 PA Groome, TA Hutchinson and P Tousignant, "Content of a decision analysis for treatment choice in end-stage renal disease: who should be consulted?," Med Decis Making 14.1 (1994): 91-7.
- 22 J Kassirer and R Kopelman, Learning Clinical Reasoning (Baltimore: Williams & Wilkins, 1991).
- 23 HC Sox, MA Blatt, MC Higgins and KI Marton, Medical Decision Making (Boston: ButterWorth-Heinemann, 1988) 406.

REFERENCES

- 24 I McDowell and C Newell, Measuring Health: A Guide to Rating Scales and Questionnaires, 2 ed. (New York: Oxford University Press, 1996).
- 25 DL Patrick and P Erickson, Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation (Oxford University Press, Inc, 1993).
- 26 D Yin, HP Forman and CP Langlotz, "Evaluating health services: the importance of patients' preferences and quality of life [see comments]," AJR Am J Roentgenol 165.6 (1995): 1323-8.
- 27 GJ Cuchural, Jr., AS Levey and SG Pauker, "Kidney failure or cancer. Should immunosuppression be continued in a transplant patient with malignant melanoma?," Med Decis Making 4.1 (1984): 82-107.
- 28 PF Pierce, Defining And Measuring Personal Values In Breast Cancer Treatment Decisions. (Toronto, Canada: 1996).
- 29 L Lenert, J Ziegler, K McDonald, M Hlatky and J Hornberger, Utilities For Health States With Arrhythmias. Poster Abstract (Toronto, Canada: 1996).
- 30 K McDonald, E Aton, B Campbell, L Lenert, M Hlatky and J Hornberger, Perceptions Of Patients And Their Families About Near-Fatal Arrhythmias And Treatments To Prevent Recurrence. Poster Abstract (Toronto, Canada: 1996).
- 31 J Ziegler, M Bergen, J Hornberger and L Lenert, The Utility Rating Task: Concerns Of Patients And Normal Subjects In Preference Assessment. Poster Abstract (Toronto, Canada: 1996).
- 32 RF Nease, Jr. et al., "Variation in patient utilities for outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team [see comments] [published erratum appears in JAMA 1995 Aug 23-30;274(8):612]," Jama 273.15 (1995): 1185-90.
- 33 M Wellman, "Reasoning about preference models," MIT Laboratory for Computer Science, 1985.
- 34 M. B. Fishman, ed. E Stiller, S Lovstad, W Bandler and V Mancini, Representing the Social Dimension: Design and Methodology for an Urban Planning System (St. Petersburg, Florida: Florida AI Research Symposium, 1989) 216-220.
- 35 J Kiecolt-Glaser and R Glaser, "Behavioral Influences On Immune Function: Evidence For The Interplay Between Stress And Health," Stress and Coping, eds. T Field, P McCade, and A Schneiderman, vol. 2 (Hillsdale, NJ: Erlbaum, 1990) 189-206.

-
- 36 JK Kiecolt-Glaser and R Glaser, "Psychological influences on immunity: making sense of the relationship between stressful life events and health," Adv Exp Med Biol 245 (1988): 237-47.
- 37 JW Pennebaker, JK Kiecolt-Glaser and R Glaser, "Disclosure of traumas and immune function: health implications for psychotherapy," J Consult Clin Psychol 56.2 (1988): 239-45.
- 38 M Holmes-Rovner et al., "Patient satisfaction with health care decisions: the satisfaction with decision scale," Med Decis Making 16.1 (1996): 58-64.
- 39 E Bass, N Fink, S Wills, A Levev, J Sadler and N Powe, Do Preference Values For End Stage Renal Disease Differ By Dialysis Type? (Houston, Texas: 1997).
- 40 RM Kaplan, JW Bush and CC Berry, "Health status index: category rating versus magnitude estimation for measuring levels of well-being," Med Care 17.5 (1979): 501-25.
- 41 JP Krischer, "Utility Structure of a Medical Decision-Making Problem.," Operations Research 24 (1976): 951-72.
- 42 GW Torrance, MH Boyle and SP Horwood, "Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States," Operations Research 30.6 (1982): 1043-1069.
- 43 LA Lenert, S Morss, MK Goldstein, MR Bergen, WO Faustman and AM Garber, "Measurement of the validity of utility elicitations performed by computerized interview," Med Care 35.9 (1997): 915-20.
- 44 G Torrance, "Toward a utility theory foundation for health status index models.," Health Serv Res 11.4 (1976): 349-69.
- 45 "Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group [see comments]," N Engl J Med 333.24 (1995): 1581-7.
- 46 "Generalized efficacy of t-PA for acute stroke. Subgroup analysis of the NINDS t-PA Stroke Trial," Stroke 28.11 (1997): 2119-25.
- 47 "Tissue-type plasminogen activator for acute ischemic stroke," Int J Technol Assess Health Care 13.3 (1997): 484-6.
- 48 D Chiu et al., "Intravenous tissue plasminogen activator for acute ischemic stroke: feasibility, safety, and efficacy in the first year of clinical practice," Stroke 29.1 (1998): 18-22.

REFERENCES

- 49 R Donnarumma et al., "Overview: hyperacute rt-PA stroke treatment. The NINDS rt-PA Stroke Study Group [In Process Citation]," J Neurosci Nurs 29.6 (1997): 351-5.
- 50 J Grotta, "Should thrombolytic therapy be the first-line treatment for acute ischemic stroke? t-PA--the best current option for most patients," N Engl J Med 337.18 (1997): 1310-3.
- 51 AM Hawn and LM Baldwin, "Thrombolytic therapy for acute ischemic stroke," J Fam Pract 45.6 (1997): 470.
- 52 KD Hoang and P Rosen, "The efficacy and safety of tissue plasminogen activator in acute ischemic strokes," J Emerg Med 10.3 (1992): 345-52.
- 53 JE Riggs, "Tissue-type plasminogen activator should not be used in acute ischemic stroke," Arch Fam Med 6.2 (1997): 102-4.
- 54 J Rankin, "Cerebral Vascular Accidents In Patients Over The Age Of 60," Scottish Medical Journal 2 (1957): 200-15.
- 55 DL Patrick, RA Pearlman, HE Starks, KC Cain, WG Cole and RF Uhlmann, "Validation of preferences for life-sustaining treatment: implications for advance care planning," Ann Intern Med 127.7 (1997): 509-17.
- 56 DL Patrick, HE Starks, KC Cain, RF Uhlmann and RA Pearlman, "Measuring preferences for health states worse than death," Med Decis Making 14.1 (1994): 9-18.
- 57 R Keeney and H Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs (New York: Wiley & Sons, 1976).
- 58 J Tsevat, S Sherman, J Sanchez-McElwee, K Mandell, F Sonnenberg and F Fowler, A Multimethod Approach To Understanding Health Values Of HIV-infected Patients (Houston, Texas: 1997).
- 59 R Keeney and H Raiffa, "Desirable Properties of a Set of Attributes," Decisions with Multiple Objectives: Preferences and Value Tradeoffs (New York: Wiley & Sons, 1976) 50-3.
- 60 G Kelly, The Psychology of Personal Constructs (New York: Norton, 1955).
- 61 J Boose and J Bradshaw, "Expertise transfer and complex problems: using AQUINAS as a knowledge acquisition workbench for knowledge-based systems," Readings in Knowledge Acquisition and Learning; Automating the Construction and Improvement of Expert Systems, eds. BG Buchanan and DC Wilkins (San Mateo, CA: Morgan Kaufmann, 1993) 240-52.

-
- 62 F Fransella and D Bannister, "A Manual for Repertory Grid Technique," (London: Academic Press, 1977) 1-30.
- 63 JE Ware, Jr. and CD Sherbourne, "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection," Med Care 30.6 (1992): 473-83.
- 64 RH Brook, CJ Kamberg, KN Lohr, GA Goldberg, EB Keeler and JP Newhouse, "Quality of ambulatory care. Epidemiology and comparison by insurance status and income," Med Care 28.5 (1990): 392-433.
- 65 KN Lohr et al., "Use of medical care in the Rand Health Insurance Experiment. Diagnosis- and service-specific analyses in a randomized controlled trial," Med Care 24.9 Suppl (1986): S1-87.
- 66 KN Lohr, CJ Kamberg, EB Keeler, GA Goldberg, TA Calabro and RH Brook, "Chronic disease in a general adult population. Findings from the Rand Health Insurance Experiment," West J Med 145.4 (1986): 537-45.
- 67 AL Stewart, RD Hays and JE Ware, Jr., "The MOS short-form general health survey. Reliability and validity in a patient population," Med Care 26.7 (1988): 724-35.
- 68 AB Bindman, D Keane and N Lurie, "Measuring health changes among severely ill patients. The floor phenomenon," Med Care 28.12 (1990): 1142-52.
- 69 D Von Winterfeldt and W Edwards, Decision Analysis and Behavioral Research (Cambridge: Cambridge University Press, 1986).
- 70 A Shiell, P Hawe and J Seymour, "Values and preferences are not necessarily the same," Health Econ 6.5 (1997): 515-8.
- 71 F Fransella and D Bannister, A Manual for Repertory Grid Technique (London: Academic Press, 1977) 1-30.
- 72 A Ryle, Frames and cages: the repertory grid approach to human understanding (London: Published for Sussex University Press by Chatto and Windus, 1975).
- 73 PH Winston, "Learning Structural Descriptions from Examples," Psychology of Computer Vision, ed. Patrick Henry Winston (Cambridge, MA: MIT Press, 1975).
- 74 PH Winston, "Learning by Analyzing Differences," Artificial Intelligence, 3 ed. (Reading, MA: Addison-Wesley Publishing Company, 1992) 349-64.

REFERENCES

- 75 PH Winston, "Learning by Explaining Experience," Artificial Intelligence, 3 ed. (Reading, MA: Addison-Wesley Publishing Company, 1992) 365-84.
- 76 CR Flowers, AM Garber, MR Bergen and LA Lenert, "Willingness-to-pay utility assessment: feasibility of use in normative patient decision support systems," Proc AMIA Annu Fall Symp (1997): 223-7.
- 77 NIOndaSr-PSS Group, "Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group [see comments]," N Engl J Med 333.24 (1995): 1581-7.
- 78 CL Bennett et al., "A comparison of perspectives on prostate cancer: analysis of utility assessments of patients and physicians," Eur Urol 32 Suppl 3 (1997): 86-8.
- 79 LA Lenert, MK Goldstein, MR Bergen and AM Garber, "The Effects of the Content of Health State Descriptions on the Between-Subject Variability in Preferences.," (Working Paper).
- 80 R Keeney and H Raiffa, "Interpreting The Scaling Coefficients," Decisions with Multiple Objectives: Preferences and Value Tradeoffs (New York: Wiley & Sons, 1976) 271-273.
- 81 C Bakker et al., "Patient utilities in fibromyalgia and the association with other outcome measures," J Rheumatol 22.8 (1995): 1536-43.
- 82 P Wakker and A Stiggelbout, "Explaining distortions in utility elicitation through the rank-dependent model for risky choices.," Medical Decision Making 15.2 (1995): 180-186.
- 83 B Fischhoff, "Value elicitation: is there anything there?," The Origin of Values, eds. M Hechter, L Nadel, and RE Michod (New York: Aldine de Gruyter, 1993).
- 84 S Simon, L Howe and H Kirschenbaum, Values Clarification: A Handbook of Practical Strategies for Teachers and Students (New York: Hart Publishing, 1972).
- 85 J Seymour, A Shiell and S Cameron, "Test-Retest Reliability of Standard Gamble and Two Stage Standard Gamble Methods for Eliciting Health State Utilities," Department of Public Health and Community Medicine, University of Sydney and Westmead Hospital.
- 86 R Gregory, S Lichtenstien and P Slovic, "Valuing environmental resources: a constructive approach," Journal of Risk and Uncertainty .7 (1993): 177-97.
- 87 E Nord, "Methods for quality adjustment of life years," Soc Sci Med 34.5 (1992): 559-69.

-
- 88 D Kahneman and A Tversky, "On the study of statistical intuitions," Cognition 11.2 (1982): 123-41.
- 89 D Kahneman and A Tversky, "Variants of uncertainty," Cognition 11.2 (1982): 143-57.
- 90 BJ McNeil, SG Pauker, HC Sox, Jr. and A Tversky, "On the elicitation of preferences for alternative therapies," N Engl J Med 306.21 (1982): 1259-62.
- 91 A Tversky and D Kahneman, "The Framing of Decisions and the Psychology of Choice," Science 211 (1981): 453-8.
- 92 CE Osgood, GJ Suci and PH Tannenbaum, The Measurement of Meaning (Urbanna: University of Illinois Press, 1957).
- 93 W Bandler and LJ Kohout, "Semantics of Implication Operators and Fuzzy Relational Products," International Journal of Man-Machine Studies 12 (1980): 89-116.
- 94 JH Boose, "A knowledge acquisition program for expert systems based on personal construct psychology," International Journal of Man-Machine Studies .23 (1985): 495-525.
- 95 W Bandler and LJ Kohout, "Fuzzy Relational Products as a Tool for Analysis and Synthesis of the Behavior of Complex Natural and Artificial Systems," Fuzzy Sets: Theory and applications to Policy Analysis and Information Systems, ed. P. P. and Chang Wang, S.K. (New York and London: Plenum Press, 1980) 341-367.
- 96 B. R. Gaines, ed. W Bandler and LJ Kohout, The Use of New Relational Products in Clinical Modeling (Louisville, Kentucky: Society for General Systems Research, 1979) 240-246.
- 97 W Bandler and LJ Kohout, "A Survey of Fuzzy relational Products in Their Applicability to Medicine and Clinical Psychology," Knowledge Representation in Medicine and Clinical Behavioral Science, ed. L. J. and Bandler Kohout, W. (London and New York: Abacus Book Gordon and Breach, 1986) 107-118.
- 98 P Green and V Rao, "Conjoint measurement for quantifying judgmental data," Journal of Marketing Research .8 (1971): 355-63.
- 99 P Green, J Carroll and F Carmone, "Some New Types Of Fractional Factorial Designs For Marketing Experiments," Research Marketing .1 (1978): 99-122.

REFERENCES

- 100 JR Clarke and RS Wigton, "Development of an objective rating system for residency applications," Surgery 96.2 (1984): 302-7.
- 101 F Orkin and D Greenhow, "A Study Of Decision Making," Anesthesiology .48 (1978): 267-71.
- 102 DK Richardson, SG Gabbe and Y Wind, "Decision analysis of high-risk patient referral," Obstet Gynecol 63.4 (1984): 496-501.
- 103 RS Wigton, VL Hoellerich and KD Patil, "How physicians use clinical information in diagnosing pulmonary embolism: an application of conjoint analysis," Med Decis Making 6.1 (1986): 2-11.
- 104 RS Wigton, KD Patil and VL Hoellerich, "Enhanced learning of clinical diagnosis through computer graphics feedback of diagnostic weighting," Proc Annu Conf Res Med Educ 23 (1984): 111-6.
- 105 K Hammond, The Psychology of Egon Brunswik (New York: Holt, Rinehart and Winston, 1966).
- 106 T Speroff, AF Connors, Jr. and NV Dawson, "Lens model analysis of hemodynamic status in the critically ill," Med Decis Making 9.4 (1989): 243-52.
- 107 T Speroff, AF Connors, Jr. and NV Dawson, "The lens model--a gem of a method of judgment analysis or just a diamond in the rough?," Med Decis Making 9.4 (1989): 257-61.
- 108 RS Wigton, "Use of linear models to analyze physicians' decisions," Med Decis Making 8.4 (1988): 241-52.
- 109 PF Krabbe, ML Essink-Bot and GJ Bonsel, "The comparability and reliability of five health-state valuation methods," Soc Sci Med 45.11 (1997): 1641-52.
- 110 KR Hammond, "How convergence of research paradigms can improve research on diagnostic judgment," Med Decis Making 16.3 (1996): 281-7.

Appendix

Satisfaction With Decision Questionnaire

Your health status requires ongoing decisions regarding treatment. In regards to those decisions, answer the following questions. [In particular, we asked the participant to bear in mind the decision between hemodialysis and peritoneal dialysis.] Please indicate to what extent each statement is true for you AT THIS TIME.

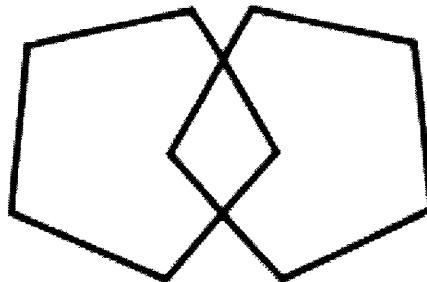
Use the following scale to answer the questions.

- 1 = strongly disagree
- 2 = disagree
- 3 = neither agree nor disagree
- 4 = agree
- 5 = strongly agree

1. I am satisfied that I am adequately informed about the issues important to my decision.
2. The decision I made was the best decision possible for me personally.
3. I am satisfied that my decision was consistent with my personal values.
4. I expect to successfully carry out (or continue to carry out) the decision I made.
5. I am satisfied that this was my decision to make.
6. I am satisfied with my decision.

Mini-Mental State Examination

Items	Points	
<i>Orientation</i>		
1. What is the	Year?	1
	Season?	1
	Date?	1
	Day?	1
	Month?	1
2. Where are we?	State?	1
	County?	1
	Town or city?	1
	Hospital?	1
	Floor?	1
<i>Registration</i>		
3. Name three objects, taking one second to say each. Then ask the patient all three after you have said them. Give one point for each correct answer. Repeat the answers until the patient learns all three.	3	
<i>Attention and calculation</i>		
4. Serial sevens. Give one point for each correct answer. Stop after five answers. Alternate: Spell WORLD backwards.	5	
<i>Recall</i>		
5. Ask for names of three objects learned in Question 3. Give one point for each correct answer.	3	
<i>Language</i>		
6. Point to a pencil and a watch. Have the patient name them as you point.	2	
7. Have the patient repeat "No ifs, ends, or buts."	1	
8. Have the patient follow a three-stage command: "Take the paper in your right hand. Fold the paper in half. Put the paper on the floor."	3	
9. Have the patient read and obey the following: "CLOSE YOUR EYES." (Write it in large letters.)	1	
10. Have the patient write a sentence of his or her own choice. (The sentence should contain a subject and an object and should make sense. Ignore spelling errors when scoring.)	1	
11. Enlarge the design printed below to 1-5 cm per side and have the patient copy it. (Give one point if all sides and angles are preserved and if the intersecting sides form a quadrangle.)	1	



= Total 30

Reproduced from Folstein MF. The Mini-Mental State Examination. In: Crook T, Ferris S, Bartus R, eds. Assessment in geriatric psychopharmacology. New Canaan, Connecticut: Mark Powley, 1983:50-51

Standardized Health State Descriptions Used in the Feasibility Study - 1 of 4

Uncomplicated Hemodialysis

Imagine what it would be like if, over the next few years, you were to receive dialysis treatment that includes the following:

Access

A blood vessel connection, a graft, is inserted in one arm.

You must keep the connection area clean.

If the connection stops working, surgery is required to repair it.

Technique

Dialysis center staff use a hemodialysis machine with an "artificial kidney".

-Staff insert 2 large needles to connect the fistula or graft to the dialysis machine.

-Staff determine the exact treatment, depending on weight and blood pressure.

-Staff take the needles out and change or clean the "artificial kidney" at session's end.

Scheduling

You must go for 3 dialysis sessions a week at the center.

These are scheduled on Mon-Wed-Fri or Tue-Thu-Sat for 3-4 hours per session.

Dialysis staff set fixed times for treatment, taking your preferences into account.

You are free between treatments.

You see the kidney doctor at the center regularly.

Diet

You should:

-Limit your fluid intake to 32 oz plus urine output per day.

-Limit your protein intake (meat, poultry, and fish) to 8 oz per day.

-Limit your dairy intake (milk, cheese, ice cream) to 6 oz per day.

Symptoms

There is momentary pain during needle insertion; local anesthesia can be reduce the pain.

There is a chance of having cramps or feeling faint during or after each dialysis session.

There is a chance of getting an infection in the bloodstream from the access.

Standardized Health State Descriptions Used in the Feasibility Study - 2 of 4

Hemodialysis Complicated with Thrombosis

Imagine what it would be like if, over the next few years, you were to receive dialysis treatment identical to that described for HD but also includes the following:

The access vessel (graft) in your arm has clotted.

Access

The connection has stopped working, surgery is required to replace it with a new graft.

For the 2 weeks it will take for a replacement site to mature to a functional state.

A temporary catheter is placed in your neck until the new access matures.

Technique

The temporary catheter must be placed immediately.

Your scheduled dialysis treatments will be impossible until the catheter is in place and alternate treatments may be required to get you back on track.

Scheduling

Putting the temporary catheter in place disrupts your immediate schedule.

This involves some waiting at the hospital for an unscheduled surgery with nothing to eat.

Alternate treatments may require alternate schedules.

Diet

Not changed. No impact beyond the effect of waiting for surgery without eating.

Symptoms

There may be some discomfort with the graft site during the 2-week healing period.

There is some discomfort with the catheter in your neck.

You may not shower with the catheter; however, sponge baths are possible.

Cramps and generalized ill feelings may occur while your kidney failure is out of control.

If so, they happen less frequently as control returns.

Standardized Health State Descriptions Used in the Feasibility Study - 3 of 4

Uncomplicated Continuous Ambulatory Peritoneal Dialysis

Imagine what it would be like if over the next few years, you were to receive dialysis treatment that included the following:

Access

- A plastic catheter is inserted in the belly for use in dialysis.
- The catheter and exit site must be washed daily, and then kept clean.

Technique

- The dialysis center staff train the patient to use clean techniques to do dialysis exchanges at home: training takes 3 weeks.
- For each exchange, the patient selects the correct type of dialysis fluid (dialysate). The type of dialysis fluid depends on the patient's current weight and blood pressure. The dialysis fluid dwells in the belly, where it is contained within the belly's natural membrane (the peritoneum). To change the fluid, the patient connects the catheter in the belly to an empty bag, and the fluid drains out by gravity. Then a bag of fresh dialysis fluid flows into the belly. This process takes about 30 minutes. When each exchange is completed, the patient disconnects the bag from the catheter and is free while the fluid does its work inside the belly. The supplies, which are ordered for home delivery, are bulky and need adequate clean, dry storage space.

Scheduling

The dialysis fluid is exchanged 4 times a day. The 1st exchange is done in the morning, the 2nd and 3rd are done during the day, and the 4th is done before going to bed. The patient chooses the exact time and place for each fluid exchange. The patient is free between exchanges.

- The dialysis staff are always available by telephone.
- The patient sees the kidney doctor at the center once a month.

Diet

Fluid intake is limited to 48 oz plus urine output per day.

Protein intake (meat, poultry, and/or fish) is limited to 10 oz per day; dairy intake is limited to 6 oz per day.

Symptoms

- The belly is distended because it contains about 2 quarts of dialysis fluid.

There is a possibility of having low back pain.

There is a possibility of getting an infection in the belly (peritonitis) from the access site.

Standardized Health State Descriptions Used in the Feasibility Study - 4 of 4

Continuous Ambulatory Peritoneal Dialysis Complicated with Peritonitis

Imagine what it would be like if, over the next few years, you were to receive dialysis treatment identical to that described for CAPD but also includes the following:

You have repeated infections in your abdomen. These infections occur every 9 to 18 months.

Access

No change.

Technique

Depending upon how sick the infection makes you, you may be hospitalized for several days for each infection, or you may be treated as an outpatient.

The more effort you put into avoiding contamination and monitoring the cloudiness of the fluid drained, the less frequent are these infections. Sometimes the pain is avoided by noticing the fluid is cloudy before the pain.

Scheduling

No change other than hospitalization if needed.

Diet

No change.

Symptoms

Each time you get an infection, you have pain in your abdomen and you feel weak.

There may also be chills and fever.

The pain may progress from mild to severe in a matter of 12-24 hours.

The pain and weakness lasts for several days.

Each infection is treated with antibiotics for two weeks. You will receive pain medicine if you need it.

IMQOL Models of Dialysis Patients

Table 11.1 IMQOL Model for Participant #01.

Coeff	Attribute levels	Value
0.244	loners, self-sufficient, people who put loved one's priorities before their own	1
	people who don't know what they want, follow the crowd, mild or moderately dependent on others	0.5
	selfish	0.35
	always has to be around somebody	0
0.203	eat right	1
	unknowing, don't know what to eat for their own good	0.65
	arrogant people who eat what they feel like	0
0.203	open-minded or always trying to be	1
	brought up with a standard ok for themselves	0.8
	very closed-minded	0.7
	evil and selfish	0
0.179	live right, putting children and family first before self	1
	not putting priorities in order	0
0.171	like working around the house	1
	not interested if it doesn't benefit himself	0

APPENDIX D: Individual IMQOL Models

Table 11.2 IMQOL Model for Participant #02.

Coeff	Attribute levels	Value
0.20	lot of responsibility, doesn't screw up	1
	little understanding of what they are doing, can catch on	0.2
	mental problems	0
0.15	not paranoid, gets both sides of the story	1
	paranoid	0
0.10	no back pain	1
	wk w/o back pain	0.05
	bad back	0
0.05	being stable	1
	exaggerates everything	0
0.10	don't believe everything that is heard	1
	tries to listen but believes a lie	0.5
	no control over nothing, talked into things easily	0
0.08	minding own business	1
	not happy with life so starts trouble	0
0.18	not drug addict or drinking	1
	beer once in a while	0.7
	stopping for a while	0.5
	drug addict, drinks	0
0.08	intelligent	1
	can teach them and they are enthusiastic	0.7
	have to treat them diplomatically like a kid	0.6
	don't understand your words or the point you are making	0
0.05	don't have to take medicine	1
	medication once in a while	0.5
	taking medicine	0

APPENDIX D: Individual IMQOL Models

Table 11.3 IMQOL Model for Participant #03.

Coeff	Attribute levels	Value
0.28	keeping out of the hospital	1
	hospitalized number of times	0
0.24	less need of checkup	1
	regular checkups required	0
0.24	never complain	1
	somewhat tired, little exercise, appetite not too great	0
0.24	pretty good shape, good physical condition	1
	poor shape	0

APPENDIX D: Individual IMQOL Models

Table 11.4 IMQOL Model for Participant #04.

Coeff	Attribute levels	Value
0.479	no pain or trivial pain	1
	pain relieved with non-trivial treatment	0.9
	chronic unrelievable pain or untreatable pain	0
0.12	ability to do what they want when they want	1
	restricted by their body as to what they can do	0.65
	restricted by treatment requirements	0
0.102	eat and drink what you want	1
	recommended restrictions in diet	0.6
	restrictions in diet, rigorous restrictions you have to follow	0
0.09	not overweight	1
	just overweight	0.78
	grossly overweight to the point it affects daily living	0
0.078	never touch drink or moderate social drinking	1
	self medicinal drinking	0.95
	heavy alcoholic drinking	0
0.072	not smoking	1
	occasional but not habitual smoking	0.8
	smoking;	0
0.06	no apparent defect	1
	cosmetic defects	0.99
	defect that can be overcome	0.3
	congenital defect medically threatening	0

APPENDIX D: Individual IMQOL Models

Table 11.5 IMQOL Model for Participant #05.

Coeff	Attribute levels	Value
0.09	people relieved of stress	1
	mental problems causing body change and stress	0
0.12	physical problems that don't have stress	1
	physical problems that cause mental problems and stress	0
0.01	companionship, long term relationships	1
	co-worker but no companion, temporary relationships	0.66
	loner, being alone	0
0.12	no disease	1
	many diseases, sick all the time	0
0.12	staying away from drugs and alcohol	1
	on drugs and alcohol	0
0.06	thin within reason, good physical shape	1
	overweight but better shape or thin and not too bad	0.5
	overweight or thin but not in good shape	0
0.12	mentally in good shape, A student	1
	mentally out of the action	0
0.09	still working, like to work	1
	depressed about job	0
0.12	better condition, get around better	1
	hard to walk around, limited because of medical problem	0.94
	no movement, death	0
0.12	not smoking	1
	smoking	0

APPENDIX D: Individual IMQOL Models

Table 11.6 IMQOL Model for Participant #06.

Coeff	Attribute levels	Value
0.105	very active at work and home	1
	somewhat physically active at work, somewhat active at home	0.95
	sedentary at work, physically active at home, with children, etc.	0.9
	physically very active at work, but TV/six-pack at home	0.8
	sedentary both at work and home, only watch TV	0
0.094	no health problems, no impediments, no life long bad habits like drinking too much or heavy smoking OR one time health problem with period of incapacity but go back to work and are fine now	1
	limited health problem present all the time, like a bad leg or poor eyesight, can live well with this, not incapacitating	0.95
	chronic health problems fine for 4 months, then comes back and goes away	0.85
	imaginary serious health problem that exists only in their mind	0.8
	had health problems and still having them, life long bad habits like drinking too much or heavy smoking	0
0.094	good mental outlook, looking forward to the future with good feelings, enjoying life as it is day to day with family as important, looking to what's happening at work	1
	changes in mental outlook due to change in circumstances	0.95
	moody, sunshine one day, gloomy the next	0.85
	manic depression	0.8
	do not have a happy outlook	0
0.099	working full time	1
	working half-time	0.9
	takes a little job or works in their spare time at home, e.g. 3-6 hours per week	0.85
	working sporadically	0.75
	not working	0
0.105	physically unlimited	1
	limited only in respect to jobs (physically)	0.95
	not having mental capacity to do certain things	0.9
	do not have desire to do things	0.75
	physical problems, limited in what they could do or want to do	0

APPENDIX D: Individual IMQOL Models

Table 12.6 (cont.) IMQOL Model for Participant #06 continued.

Coeff	Attribute levels	Value
0.105	interests are broad, strong interest in current affairs, local or global, interested in their own time, want to travel, never stops learning	1
	confined to several interests	0.98
	parochial, only interested in a certain field, absorbed in one small narrow thing, don't want to travel, just doesn't open themselves for learning, absorbed in his or her work, not interested in the world around - politics or entertainment	0
0.089	reads a lot of books or listens to tapes with spare time	1
	occasional book	0.75
	reads periodicals or newspapers	0.7
	read material only related work	0.5
	not read books	0
0.105	exercise regularly, like programmed walks	1
	forced by something to exercise more	0.98
	Fits and starts, all out couple weeks, then do nothing for period of time	0.95
	involuntarily can't exercise, only able to once in a while	0.85
	lack of any want to exercise	0
0.099	children and interests of others first, responsibilities that go with that OR people who deal with children, who take in children for care, e.g., teachers, nurses, grandparents, etc.	1
	no immediate family but extended family responsibilities	0.95
	no interests, no family, no responsibilities	0
0.089	when they see something not right, they move to correct it, take the bull by the horns and change it, self-motivated	1
	do things spasmodically or spurts of energy then sink back to doldrums	0.75
	only does what they are compelled to do like husband who only does when nagged or they must do it to keep their job or starts and never finishes	0.6
	throw arms up and say, "what can I do about it?"	0

APPENDIX D: Individual IMQOL Models

Table 11.7 IMQOL Model for Participant #07.

Coeff	Attribute levels	Value
0.421	being strong enough to two jobs and school, things you are able to do, more activities, more strenuous work	1
	still want and attempt to do strenuous work	0.5
	can do but don't want to do strenuous work	0.25
	not strong, can not do strenuous work	0
0.158	haven't been in the hospital	1
	go in the hospital like once a year	0.4
	keep going to the hospital for a lot of little things	0
0.105	not taking medications	1
	constant medication or antibiotics	0
0.211	no breathing problems	1
	problems breathing but not a major one, doesn't make them go to the hospital, treat with over-the-counter medication or none at all	0.6
	problems breathing, require prescription medication all the time	0
0.105	doesn't complain of being sick, don't constantly call in sick	1
	have something minor but there don't complain	0.45
	pain and aches, always complaining, not going to work	0

APPENDIX D: Individual IMQOL Models

Table 11.8 IMQOL Model for Participant #08.

Coeff	Attribute levels	Value
0.135	take care of themselves, keep themselves in good shape, have to support families with those responsibilities	1
	don't care, don't take as much care because they only have a few more years, don't have family responsibilities	0
0.135	eat good	1
	eating junk food all day	0
0.135	play sports, exercise	1
	think they are in good shape and don't need to exercise	0.5
	don't play sports or exercise	0
0.11	get up and do something	1
	sit on their rears, watch television all day	0
0.11	take the time	1
	don't take the time or don't have the time	0
0.135	look fine	1
	look ill	0
0.135	don't see the doctor because they don't need to, but still seeing the doctor for checkups	1
	should be seeing the doctor, but think they are their own doctor	0.5
	seeing the doctor very frequently, too much, hypochondriac	0
0.11	work every day	1
	can't work at all	0

APPENDIX D: Individual IMQOL Models

Table 11.9 IMQOL Model for Participant #09.

Coeff	Attribute levels	Value
0.139	chronic uncontrolled life-threatening problems	1
	controlled life-threatening conditions	0.95
	life-threatening in remission	0.75
	nothing but transient or minor episodes	0
0.104	lack of anxiety sound mental health or non-hospitalized controlled mental illness	1
	Unhospitalized bonkers	0.75
	hospitalized bonkers	0
0.035	don't need to see the doctor	1
	need to see, know it and see the doctor	0.999999
	hypochondriac, don't need to see don't know , don't need to , but see the doctor	0.9
	need to see, know it and don't see the doctor	0.75
	need to see, don't know it and don't see the doctor	0
0.132	standard BP	1
	controlled high BP or under treat	0.99
	high BP not controlled	0
0.139	lack of heart problems, sound heart	1
	controlled heart problems	0.99
	heart problems in remission	0.9
	uncontrolled heart problems	0
0.132	standard cholesterol, medically acceptable level for age/wt/ht	1
	controlled high cholesterol	0.99
	uncontrolled high cholesterol	0
0.035	acceptable weight for ht/age/sex	1
	borderline cases or fluctuating wt, in and out of acceptable range	0.99
	overweight	0
0.132	abstain from cocaine	1
	intermittent or rare social users	0.99
	propensity to snort cocaine consistently	0
0.104	monogamous relationships	1
	whore around	0
0.007	lot of exercise	1
	moderate amount exercise	0.99
	cant walkout of his own way	0
0.007	abstinence in drinking	1
	moderate social drinker	0.99
	drink to excess	0
0.007	no ill effects from travel, strange food or strange places	1
	cuisine/location-specific problems	0.999
	intermittent and random problems with travel and strange food	0.99
	consistent travel/food related gastric distresses	0

APPENDIX D: Individual IMQOL Models

Table 11.10 IMQOL Model for Participant #10.

Coeff	Attribute levels	Value
0.125	free to do what they want, not under tension or stress, outlook is that we put up with whatever the day is, thankful, know how fortunate they are, or neutral who could go either way	1
	people happy with work, standardized, have necessities on par with others, enjoy time off, sickness restricts, must pay attention to plans, part of your life is not your own	0.85
	so involved in work that it affects their home life, no time to talk after work, workaholics, no time to enjoy things in life, job involves all of their energies	0.65
	burdened with worry and tension, time taken up with nursing someone, tending to someone else like a husband	0
0.125	companionship, intermingle with others, enjoy others company	1
	doesn't care, gets along with everything, kind of a private person	0.9
	quiet, listener, very observant but doesn't get involved with disagreements, straight shooter, honest in what little he says, very agreeable	0.75
	crabby, being alone, don't mix, not mingling, disagreeable, don't enjoy other's company kept to themselves, don care who understands	0
0.125	talk to younger people, inquisitive how others are doing, likes to see others prosper, thinks more of fellow man than himself	1
	don't care if others make it or not, doesn't offer advice, not glad to help, think just about themselves, like a bump on a log	0
0.125	thankful, naming the best of it, always aware that others are worse off, accept fate	1
	people who can cope with anything, but just go along worrying about own well-being	0.85
	sickness makes them feeling sorry for themselves, go around feeling sorry for themselves all the time	0
0.125	don't get upset, laugh off things that happen, nice outlook on life	1
	say one thing mean another	0.85
	not letting things go, not being forgiving, holding things against others, always holding a grudge	0
0.125	better communication with others, can talk to others, understanding, converse intelligently	1
	shy, kind, soft-hearted, not good at communication, doesn't know how to express self, lost when talking to brighter people	0.85
	derogatory uncooperative, likes to argue	0
0.125	hobbies like music, choir, going to plays, camping (like the outdoors), hikes and day trips	1
	not having something to occupy your mind, watch TV, not broadening horizon, standing still, nothing done with hands to be creative, no outdoor interests	0
0.125	activity sport or other thing, just something to do	1
	time on your hands, not having interests	0

Table 11.11 IMQOL Model for Participant #11.

Coeff	Attribute levels	Value
0.24	walking straight, standing straight	1
	walk around and need a rest to keep on going	0.9
	different walk or stance instead of straight	0.8
	in bed or need machines to help them get around	0
0.21	always active, working or going out	1
	used to be always active but staying home more	0.8
	staying home and doing less	0
0.20	more energy to do things	1
	go out but moderate energy levels	0.95
	go to work go home, only go out 1-2 times a week	0.8
	just got to work and go home, that's it	0.75
	always being tired	0
0.17	don't smoke	1
	smokes	0
0.18	don't complain as much	1
	always complain that they are sick and ill	0

APPENDIX D: Individual IMQOL Models

Table 11.12 IMQOL Model for Participant #12.

Coeff	Attribute levels	Value
0.163	not sickly	1
	kind of sickly	0.75
	very sickly	0
0.138	get lot of strenuous work, get lot of exercise, very active	1
	get exercise but its not extreme	0.75
	don't get much work at all, doesn't exercise, not as active	0
0.138	watching your weight, stable weight	1
	loosing weight	0
0.13	good eating habits, eat better	1
	dieting to loose weight	0.6
	not eating right	0
0.163	walking fast, walking firmly, not stumbling about	1
	weak walk, get tired in the legs, walk slow, not far or walking with a cane	0.8
	on crutches	0.7
	wheel chair	0
0.138	do not need help	1
	only need help at certain times	0.7
	can't do things for themselves	0
0.13	no complaints	1
	complain sometimes or for certain things	0.8
	always complaining	0

APPENDIX D: Individual IMQOL Models

Table 11.13 IMQOL Model for Participant #13.

Coeff	Attribute levels	Value
0.13	lot of energy, always doing stuff, work all day then come home and do projects, talented and able to do stuff that is challenging	1
	spurts of interest but no all the time	0.1
	willing but not able	0.05
	does stuff but bored, not interested in stuff, just watching television all day, in a rut	0
0.07	have not been sick in any major way	1
	sick but not serious, minor problems	0.96
	has had a major sickness	0
0.07	looks well	1
	some days look better than others	0.9
	pale looking	0
0.07	good night's sleep	1
	sleep deprived, more than half the night but not long enough	0.98
	sleep half the night, ~4 hours	0.95
	dose off and wake up	0.5
	sleepless, don't have great sleep	0
0.13	great attitude, look at everything with positive attitude, never complain	1
	complain sometimes	0.9
	down, negative about everything, complaining all the time	0
0.10	very willing to help people	1
	can't do as much to help	0.85
	don't want to help	0
0.13	sociable	1
	social but want times alone	0.9
	not too forthright but pleasant quiet person	0.85
	withdrawn	0
0.07	more forward, form friendships, outgoing, leader or invite people to do stuff but not all the time	1
	sporadically, more moderate outgoing	0.95
	follower, won't initiate, depend on others, always waiting for someone to come to them, feels sorry for self	0
0.03	don't have to watch their diet, able to eat anything you want	1
	don't eat food because of no desire, or can't eat more - filled up	0.8
	had to watch diet	0
0.07	able to do things w/o fear of physical harm	1
	can't do some things	0.95
	can't do anything w/o fear of physical harm	0
0.13	able to drink occasionally, can enjoy a drink	1
	drinking too much	0

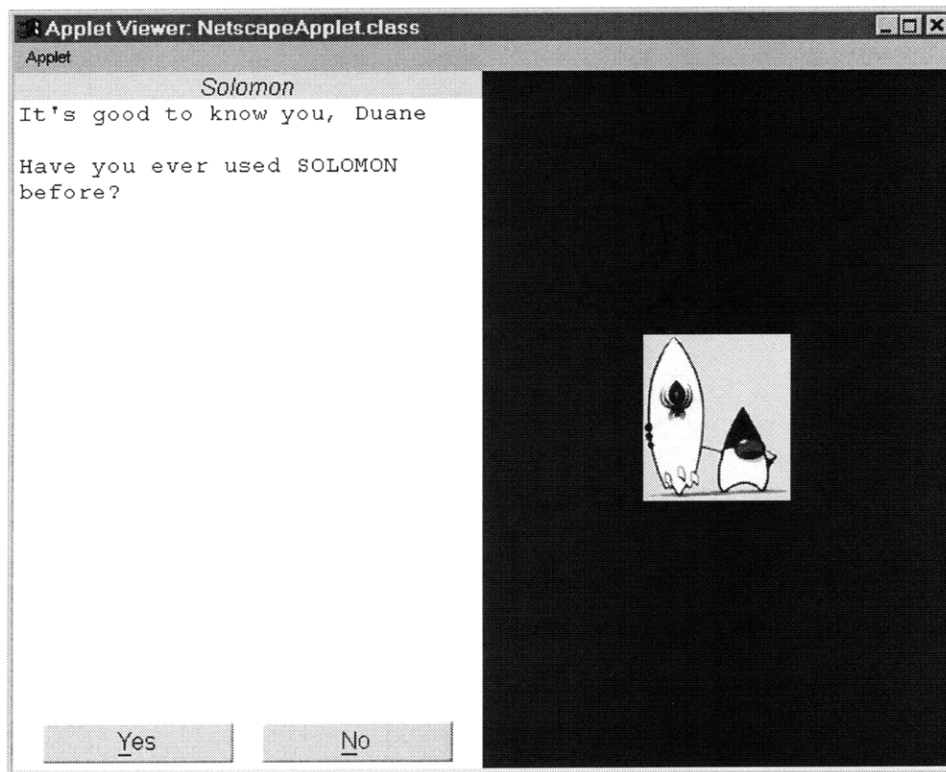
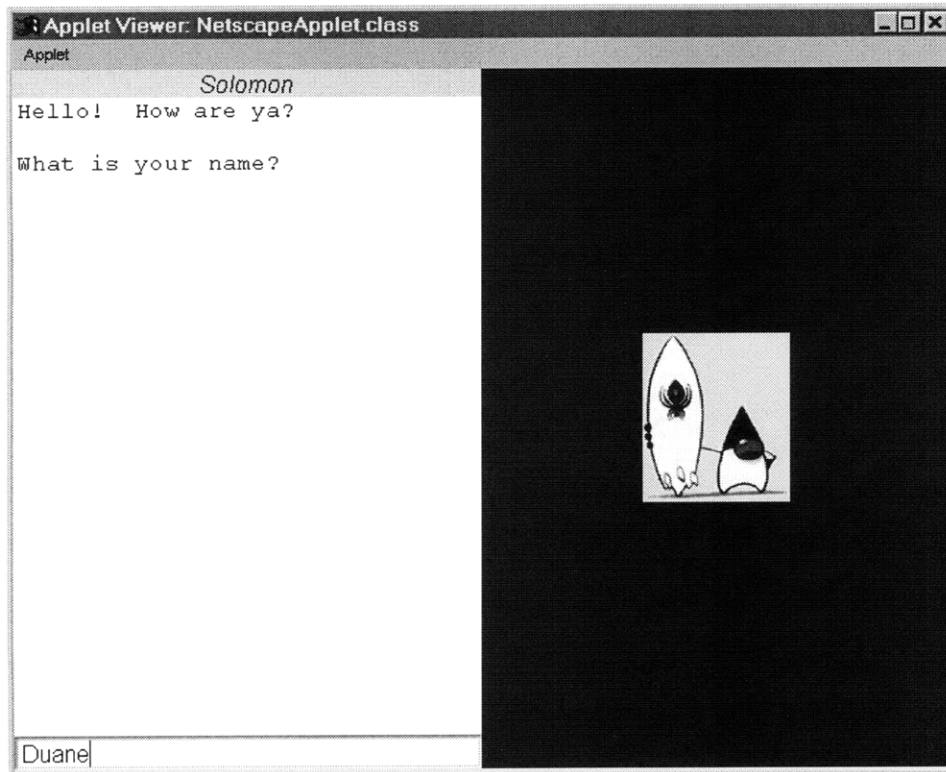
APPENDIX D: Individual IMQOL Models

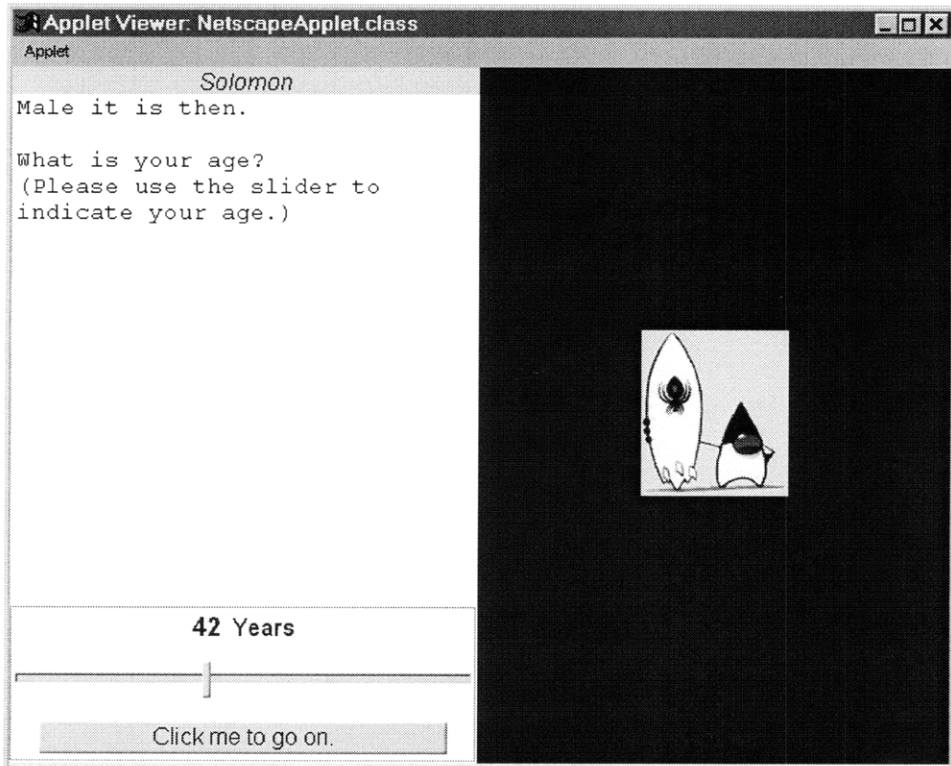
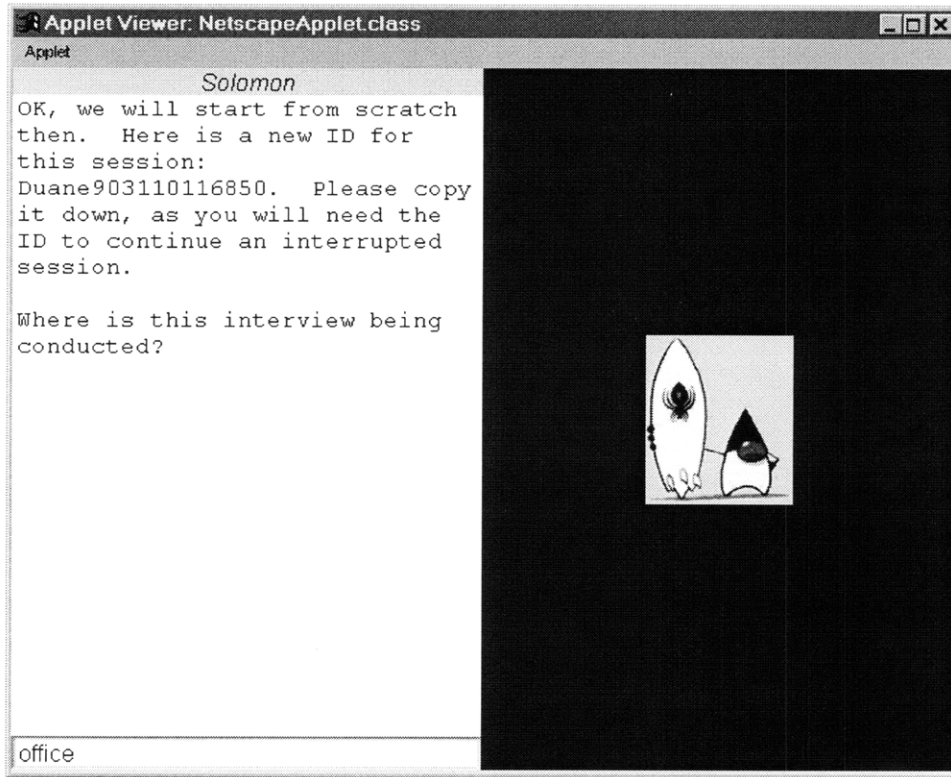
Table 11.14 IMQOL Model for Participant #14.

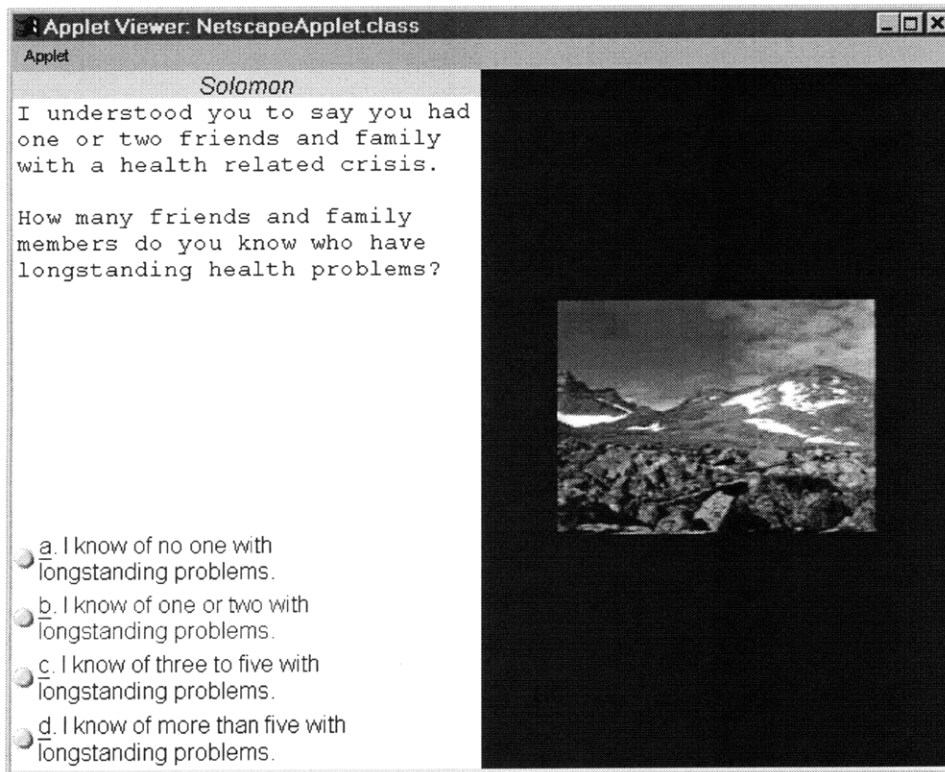
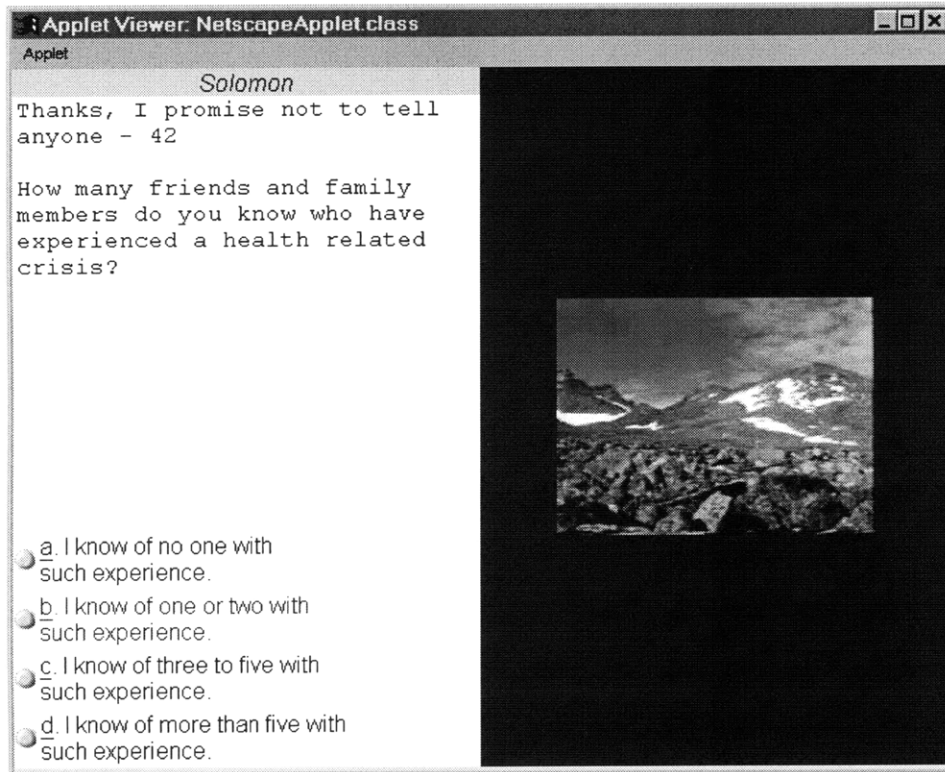
Coeff	Attribute levels	Value
0.435	no health issues, very good or good	1
	mediocre	0.97
	bearable health issues	0.83
	terrible health issues	0
0.145	happy, total tranquility, or fairly at ease, comfortable, (Satisfaction with lifestyle, ease of mind, e.g. doesn't work full time but volunteers implying less stress, e.g. have children or other people who depend upon you)	1
	satisfied but not necessarily ecstatic	0.98
	somewhat anxious	0.97
	very anxious	0.83
	extreme stress, suicidal	0
0.145	support ing family and friends to help through health issues and stress	1
	no support, nobody to depend upon to back you up	0
0.145	thin, right weight for height or 10% outside the parameter	1
	20% out side parameter of right weight for height	0.97
	40% out side parameter of right weight for height	0.83
	> 40% out side parameter of right weight for height	0
0.058	Athlete,2x/week exercise, walks shopping mall	1
	at home its TV, & their job, that's all they do	0.98
	spends all day in bed	0
0.058	having enough vim & vigor, can walk with kids or dog, easily climb up stairs with laundry, would not rather sleep than get up in AM, feel like going out to dinner or have energy to do what you want, not get tired more than you like	1
	push to get things done but its never easy	0.8
	not wanting to do because energy and motivation are lacking	0
0.014	seeing doctor, e.g. cholesterol in order, paying attention to diet, in a condition of knowledge of health	1
	don't see doctors, don't have knowledge of their own health	0

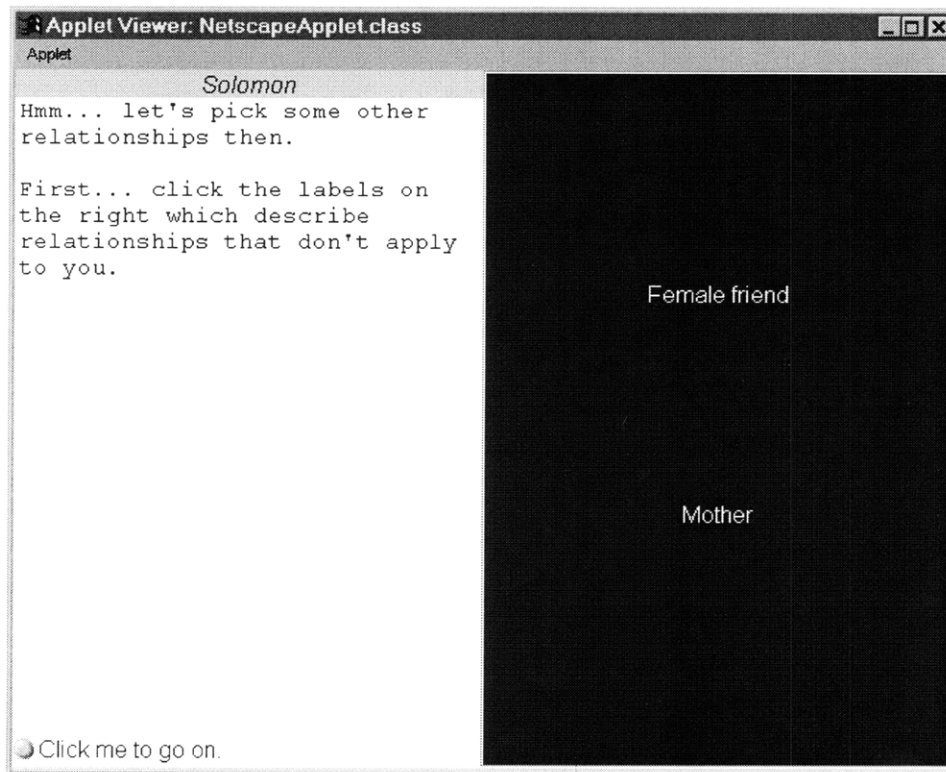
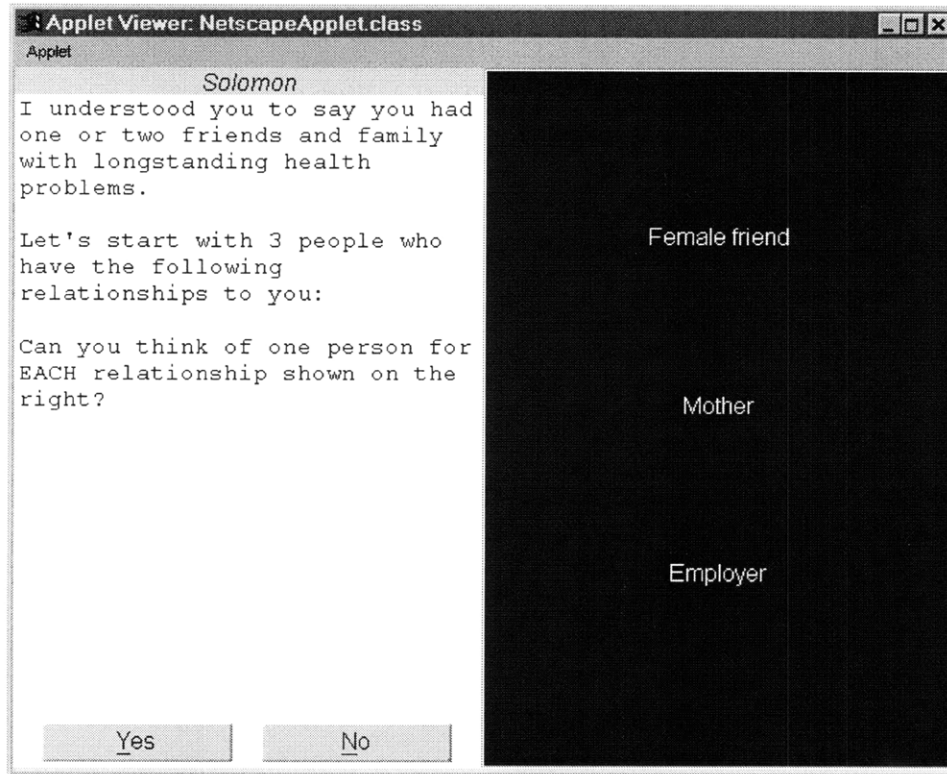
Table 11.15 IMQOL Model for Participant #15.

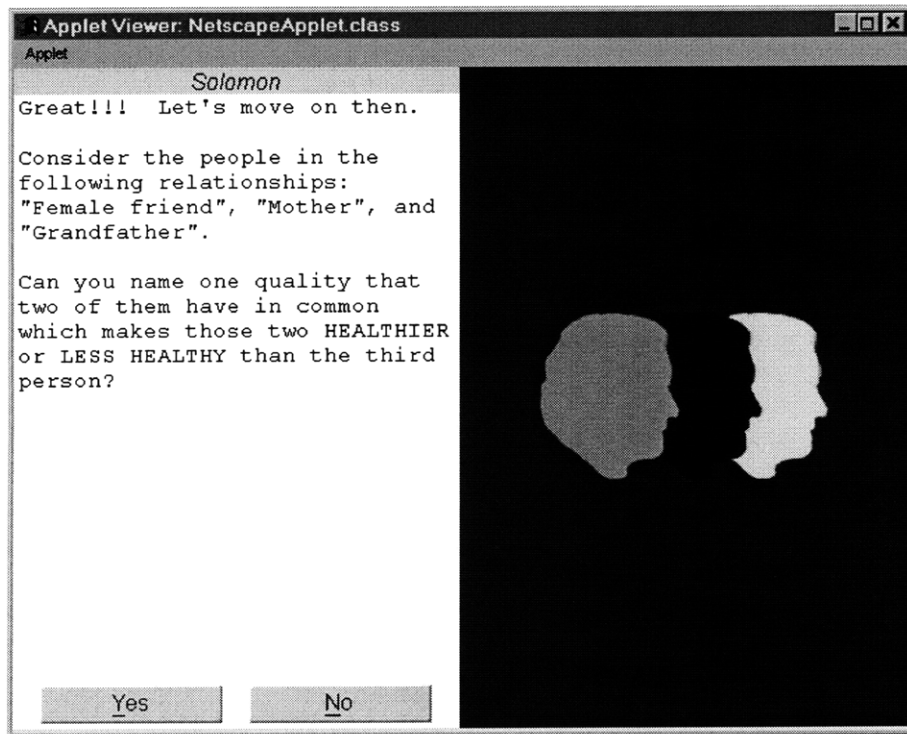
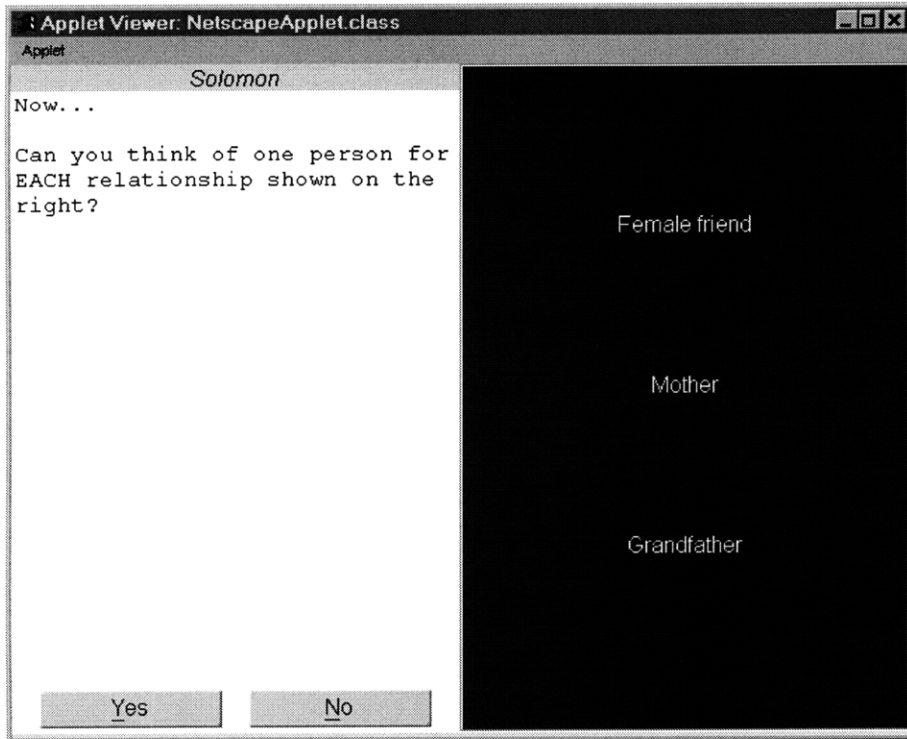
Coeff	Attribute levels	Value
0.194	no medical treatment, no medical problem or infection once in a while, get sick then get healthy again	1
	sick, under a big treatment or long term	0
0.123	the way they eat, good diet	1
	eat regular diet, a lot of fat food today and a lot of vegetables next day	0.95
	eat a lot of fat foods	0.82
	not eating at all	0
0.123	exercise a lot	1
	exercise enough, some or little bit	0.999
	exercise once in a while	0.983
	no exercise at all	0
0.165	little stress, not a lot of responsibility, or doesn't take on too much problems	1
	not all the time stress or normal with kids, work, bills	0.999
	stress, in charge of a business a lot of kids, take on lot of problems even for other people	0
0.135	good childhood, enough food, home, love from parents, play around	1
	some years of bad childhood but get better or started work early, troubles in childhood, poor and had to work, lots of brothers and sisters so had to help	0.917
	totally bad childhood, e.g. abused child, no parents or live in streets	0
0.129	strong	1
	mixed person, sometimes strong – sometimes weak	0.96
	weak	0
0.132	part time job or less; plenty of rest	1
	full time job but enough time to rest, no interference with rest	0.999999
	work - does a lot by hand, means less time to rest	0.99
	don't rest at all	0
0.165	less kids	1
	not too much, but have some kids	0.983
	Woman who miscarried	0.95
	More kids	0

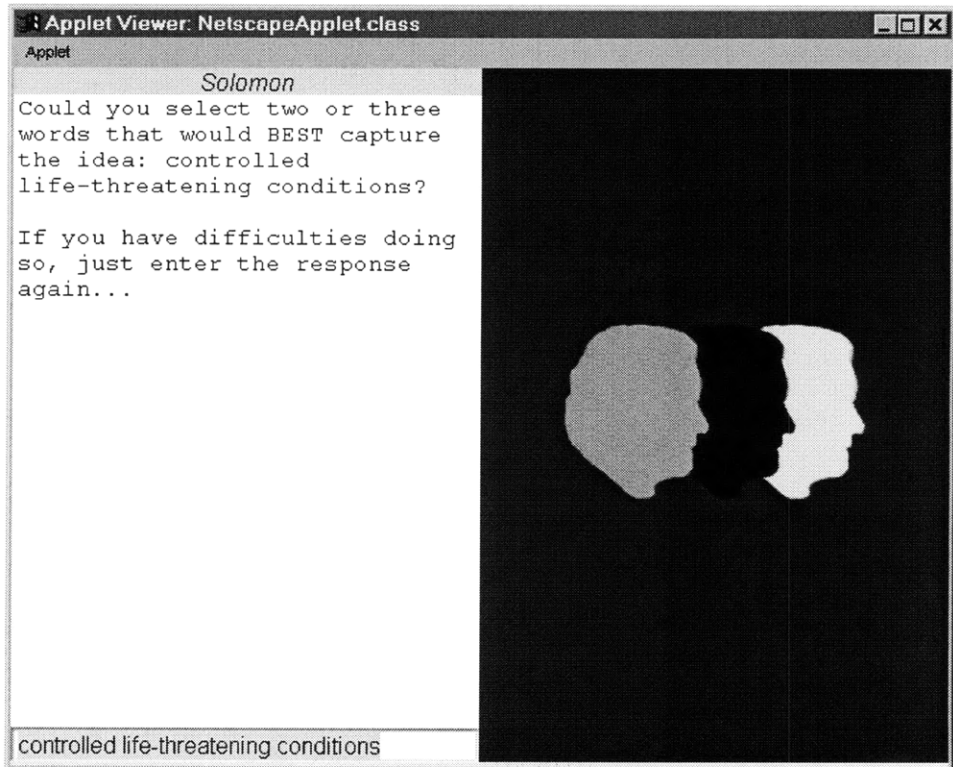
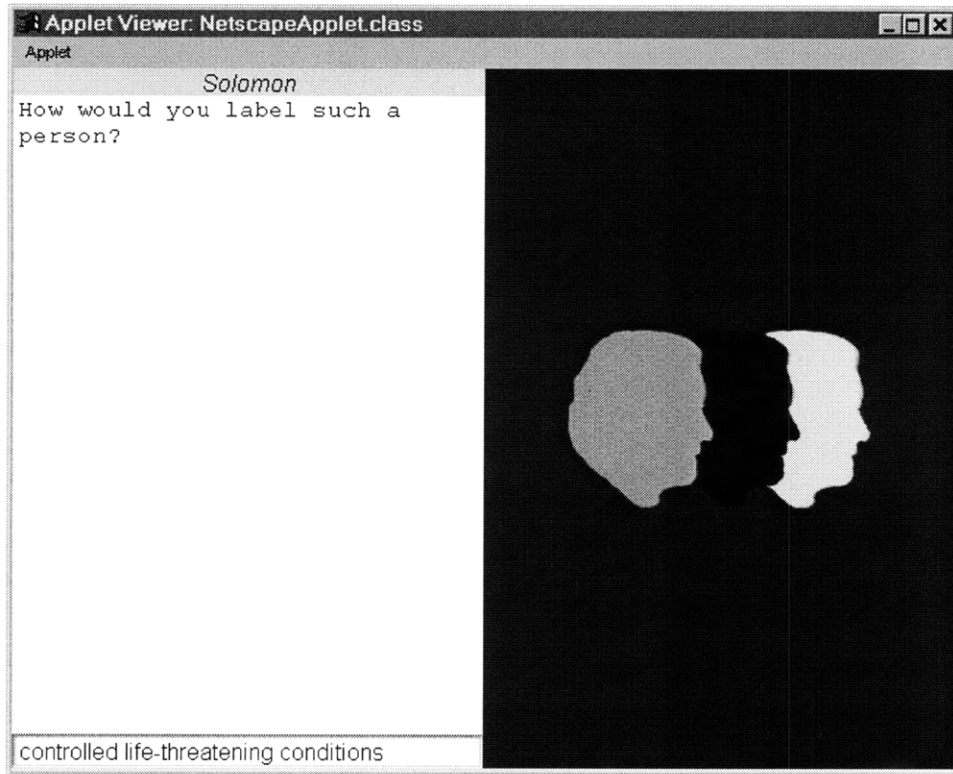


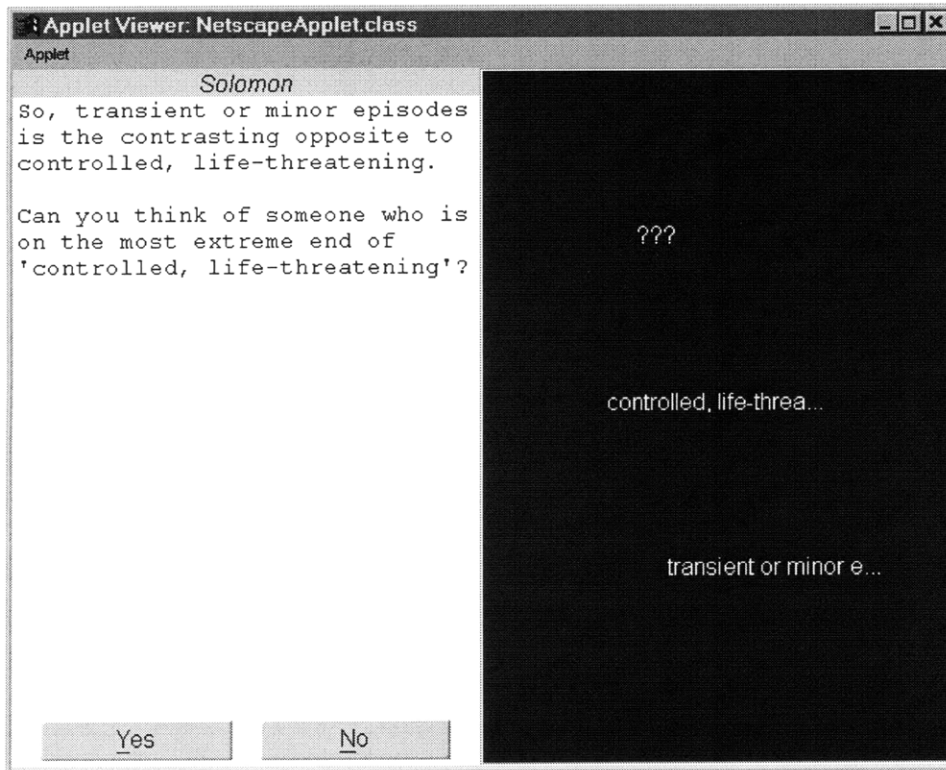
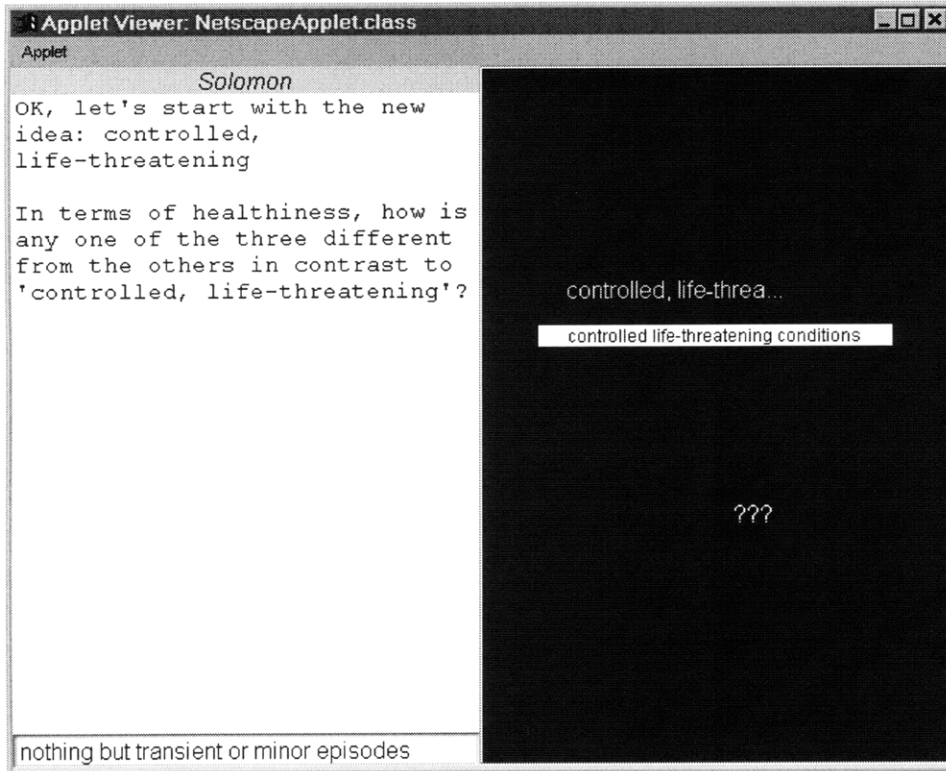


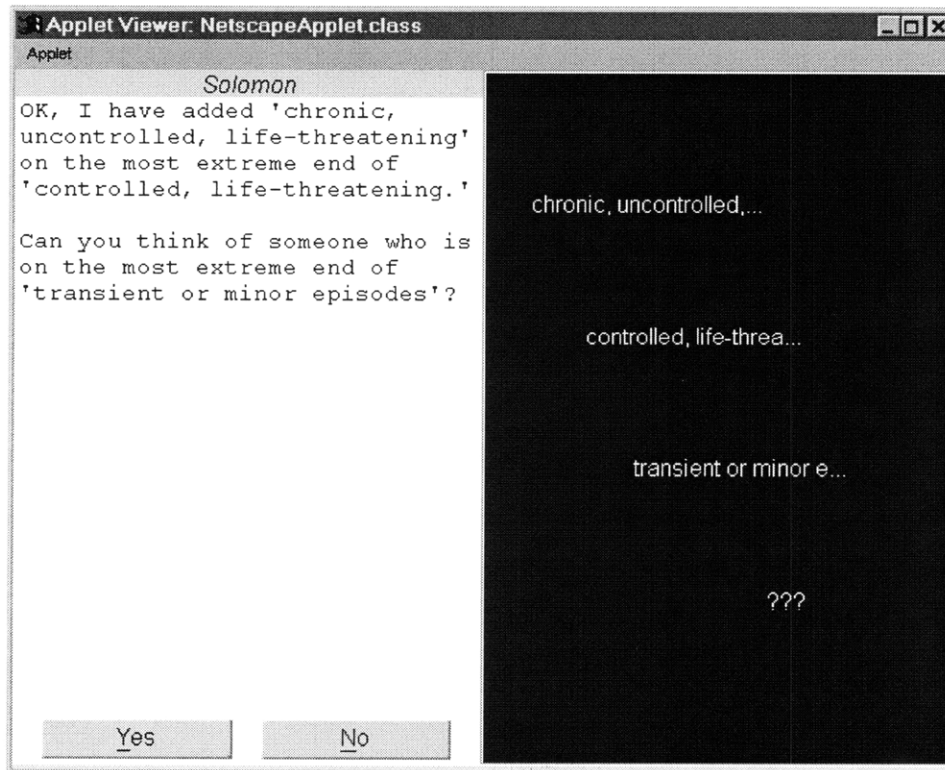
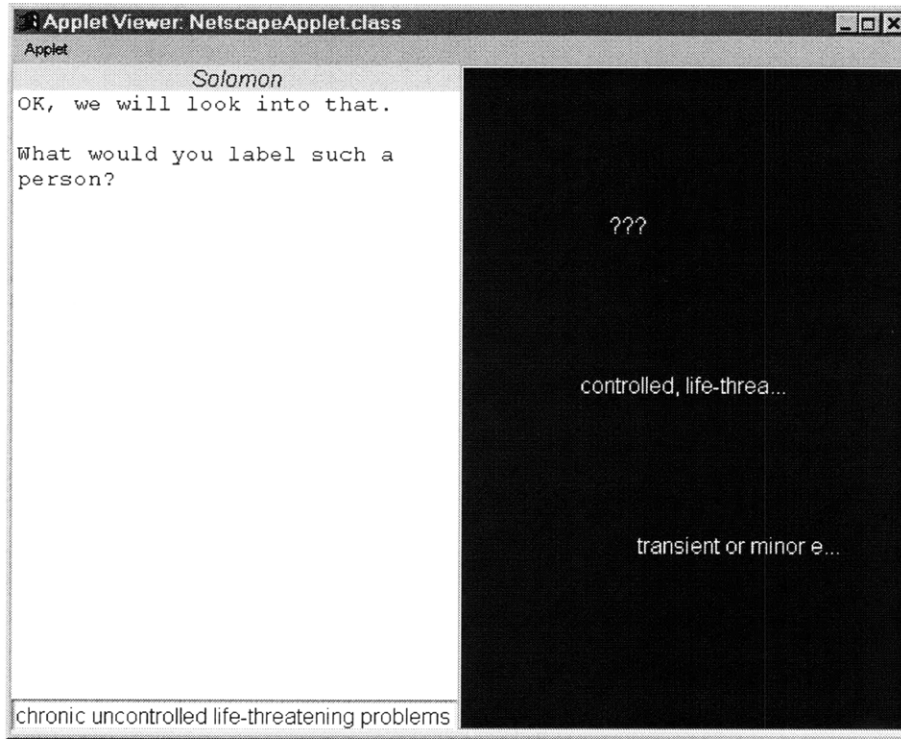


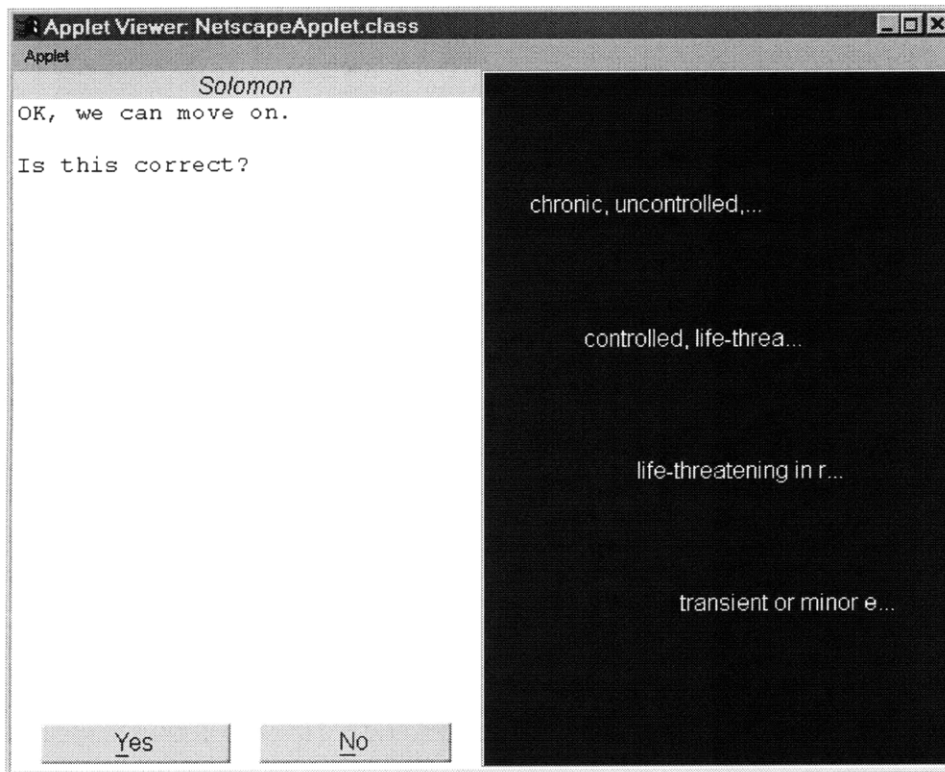
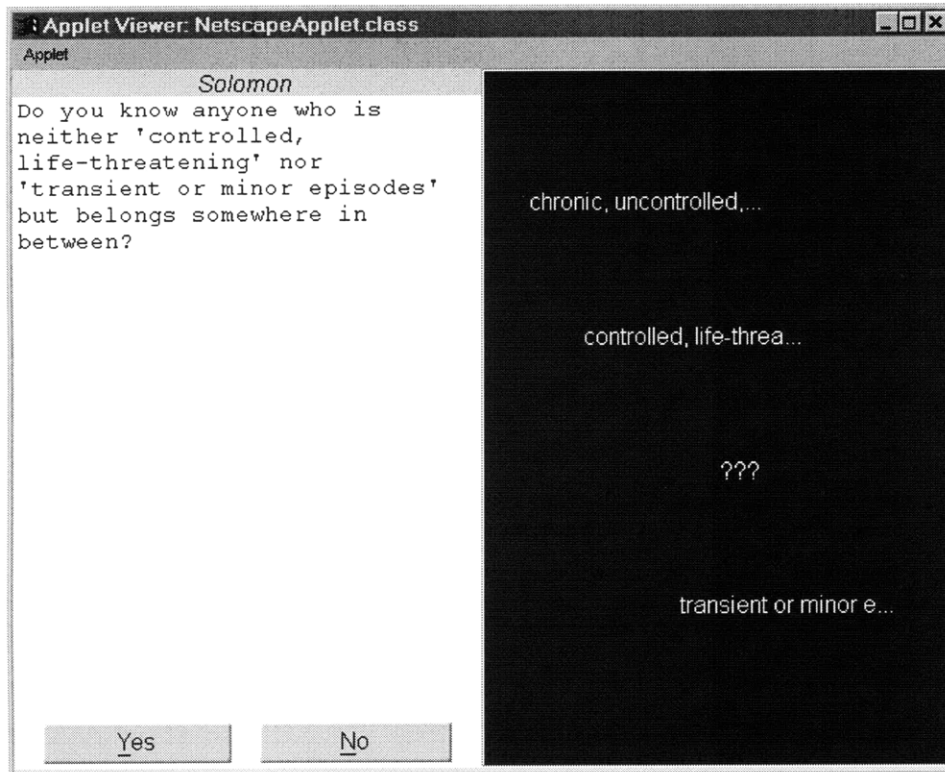


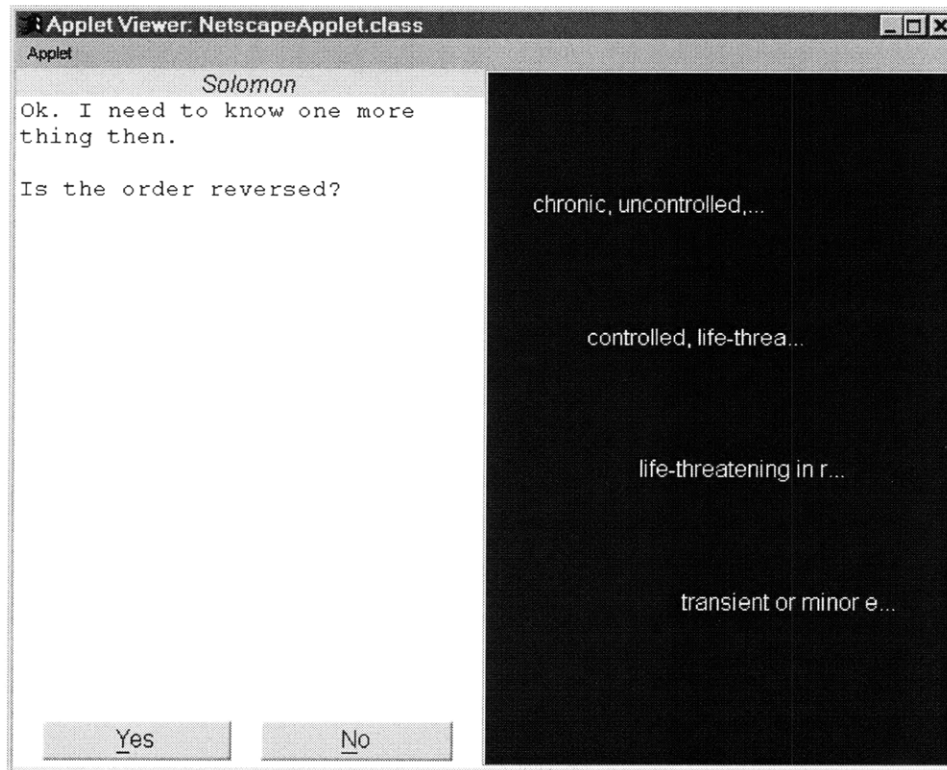
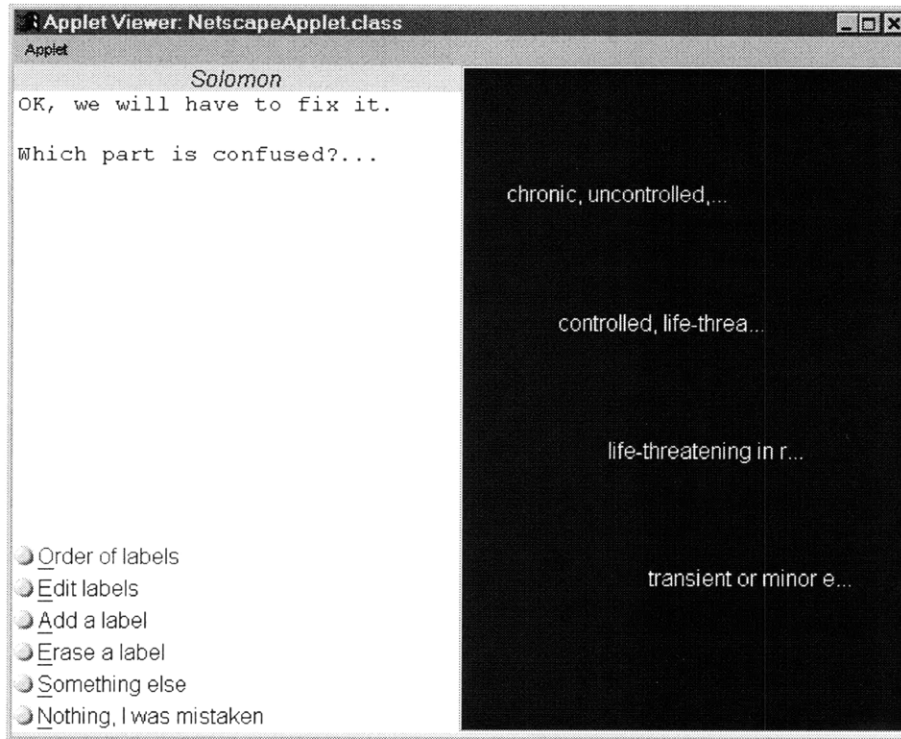


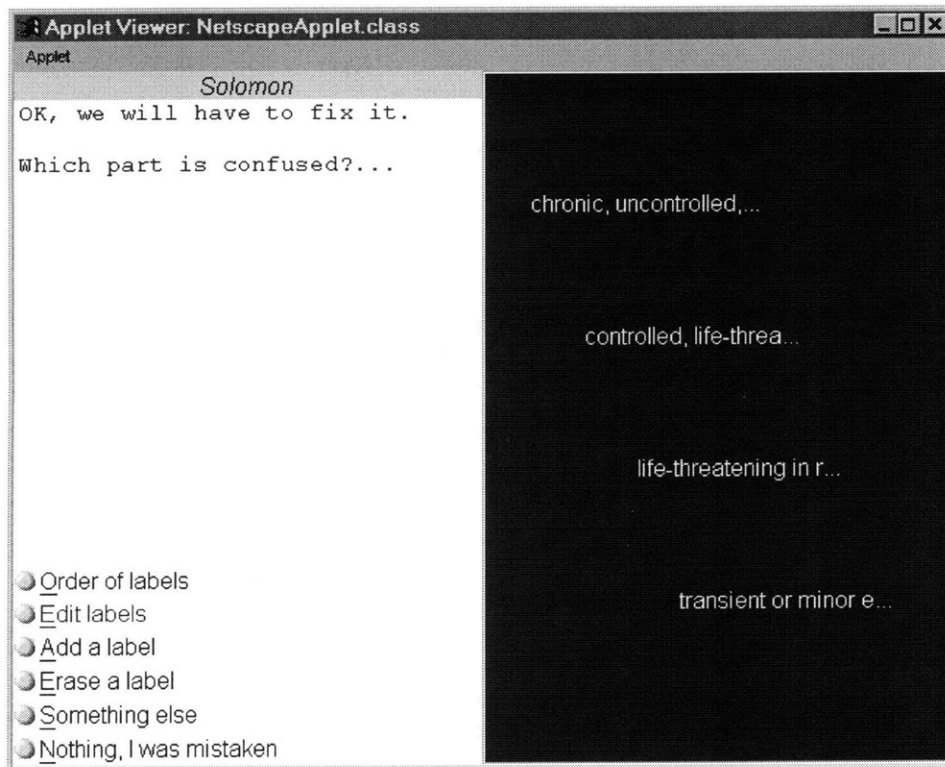
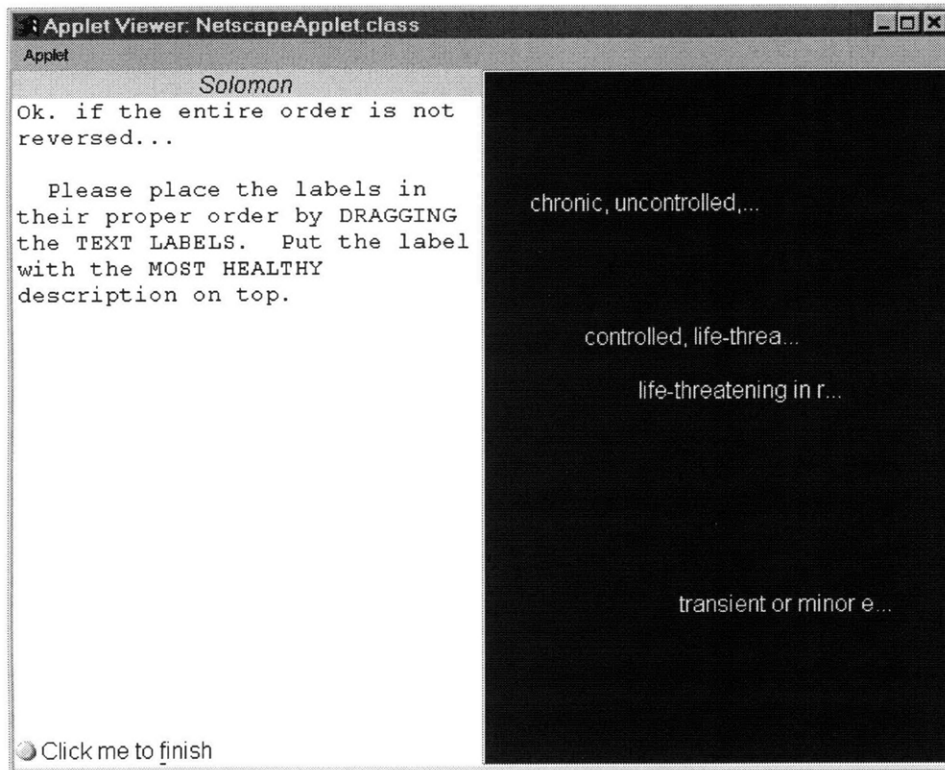


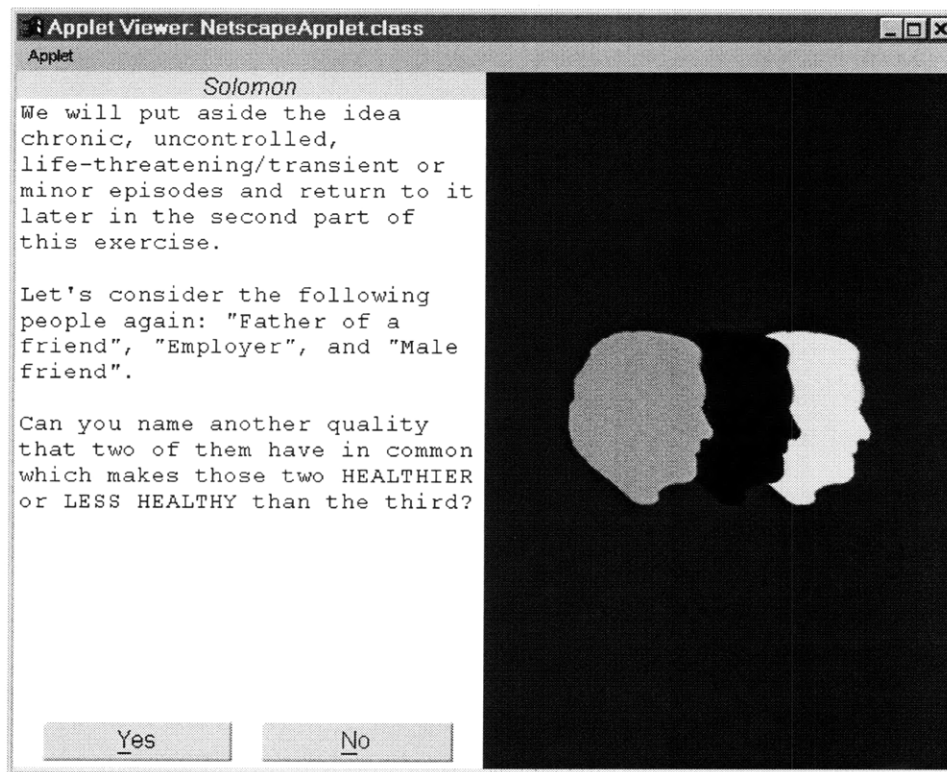
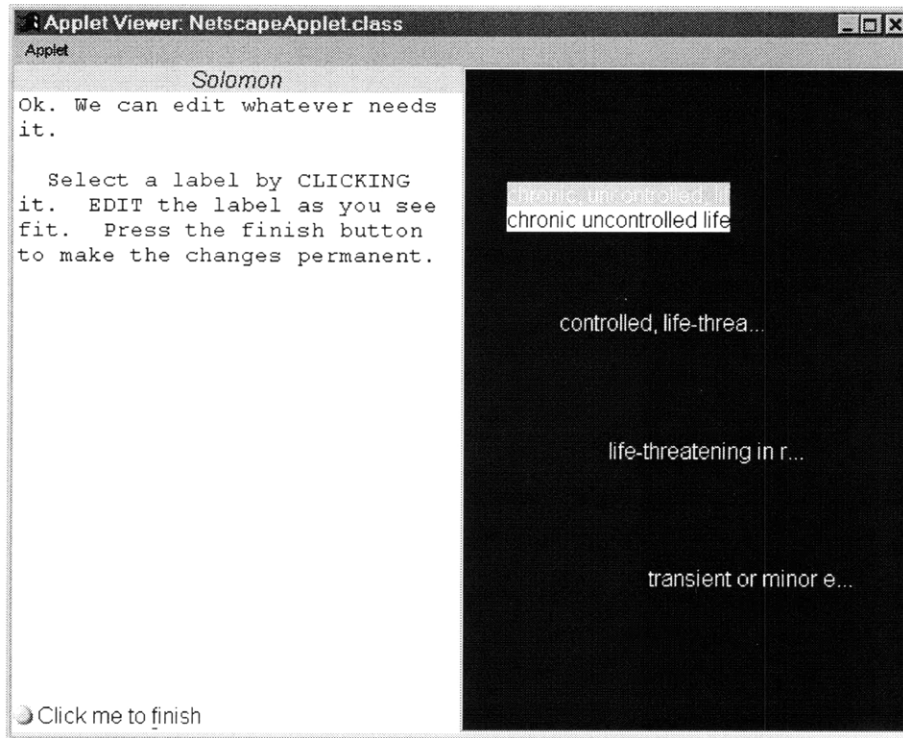


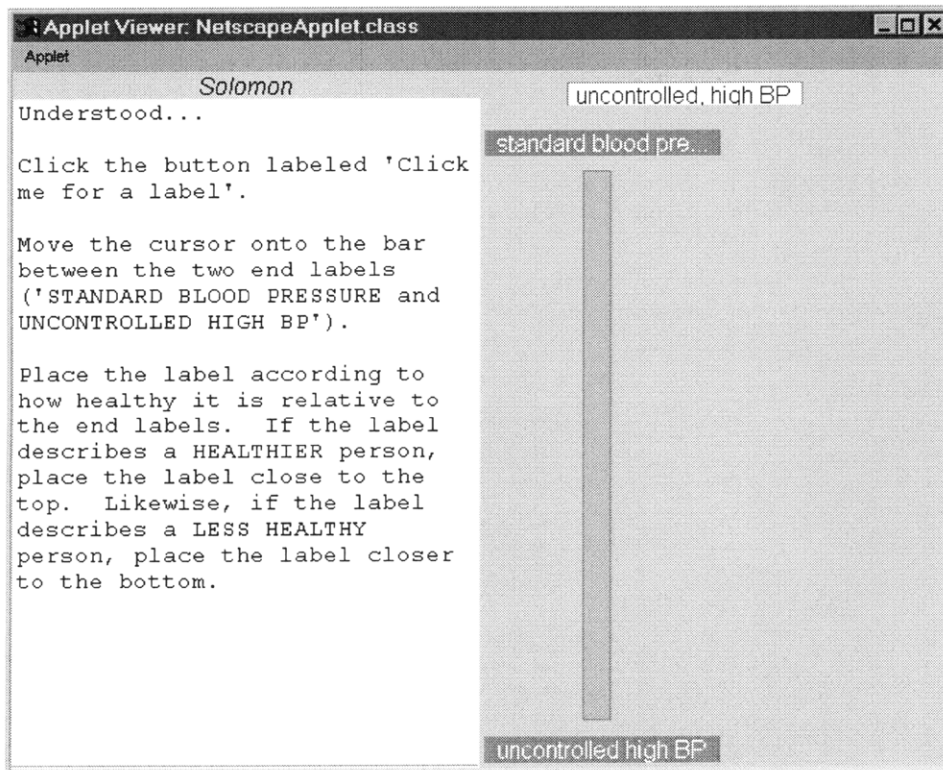
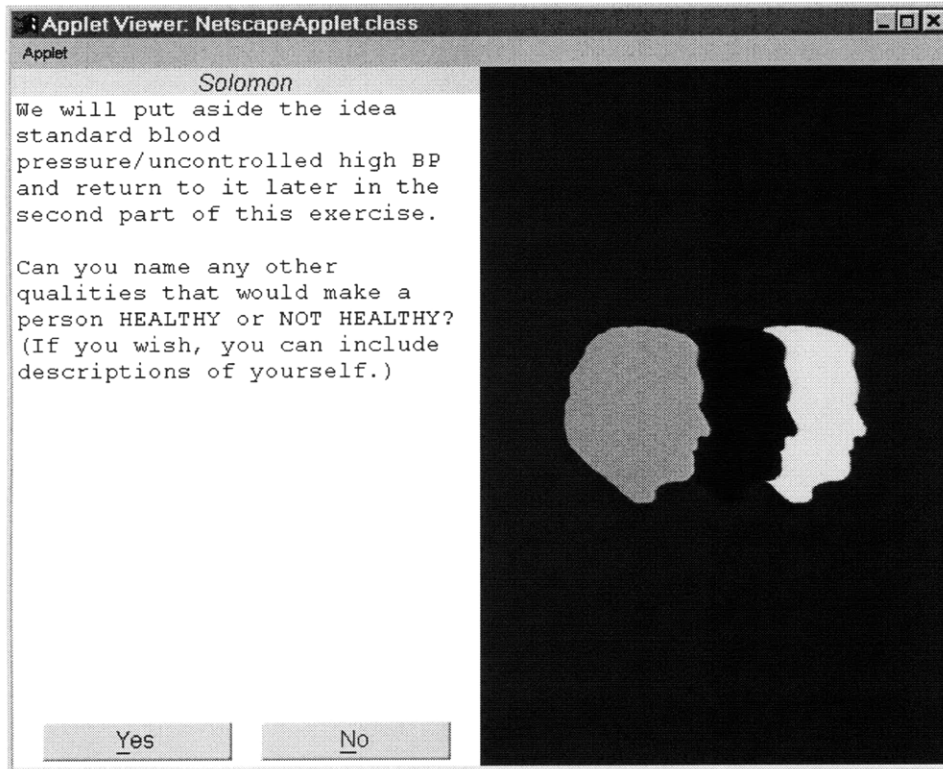


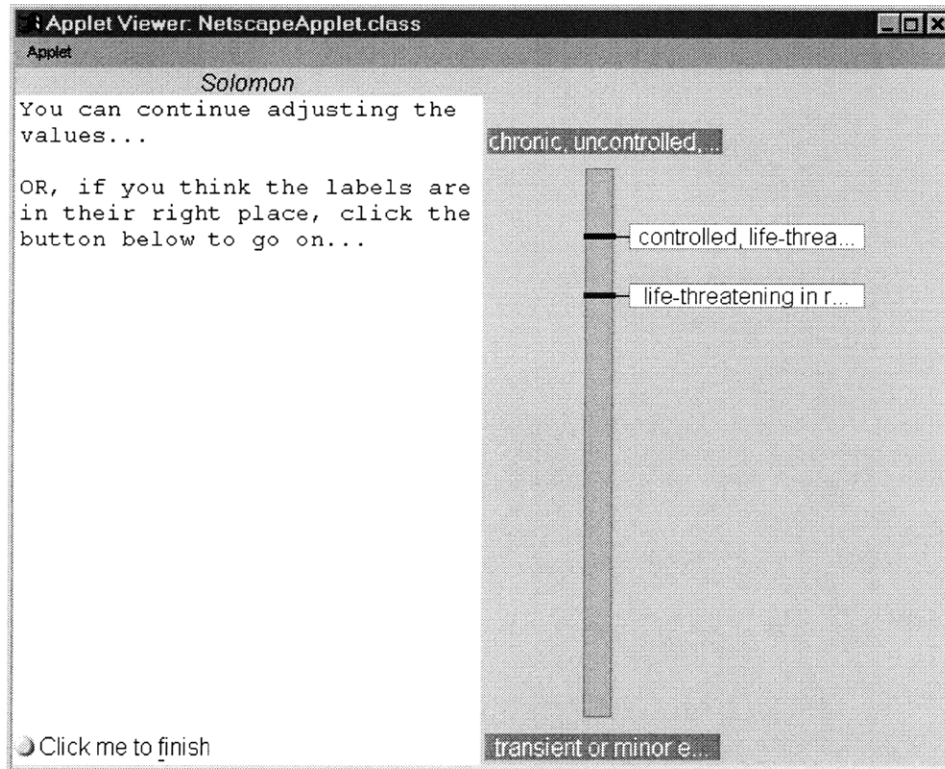
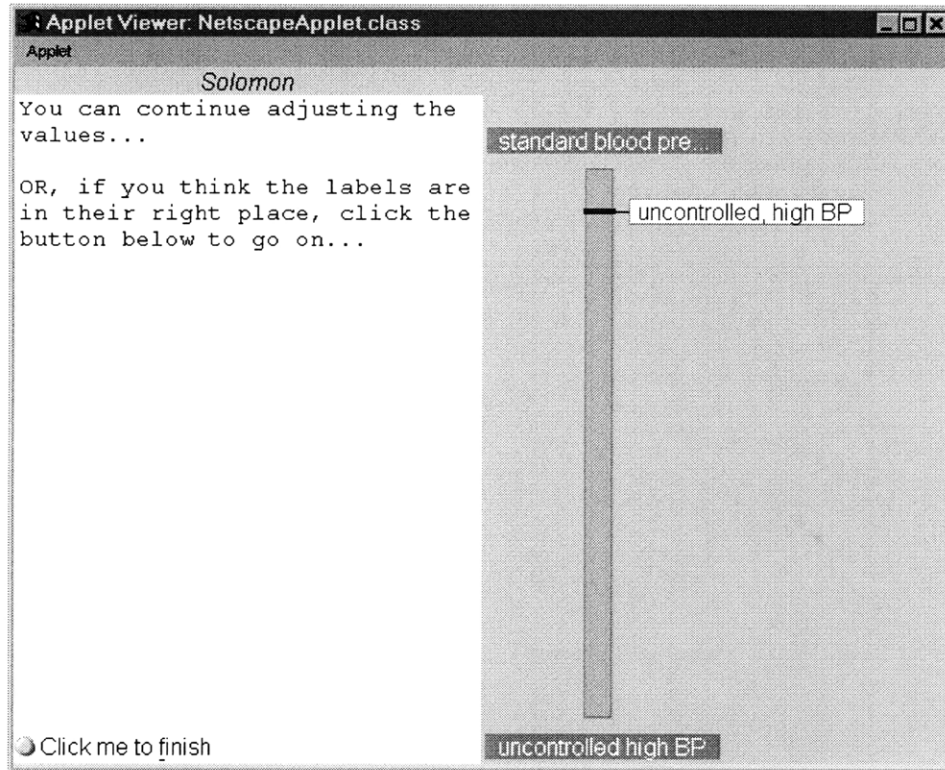


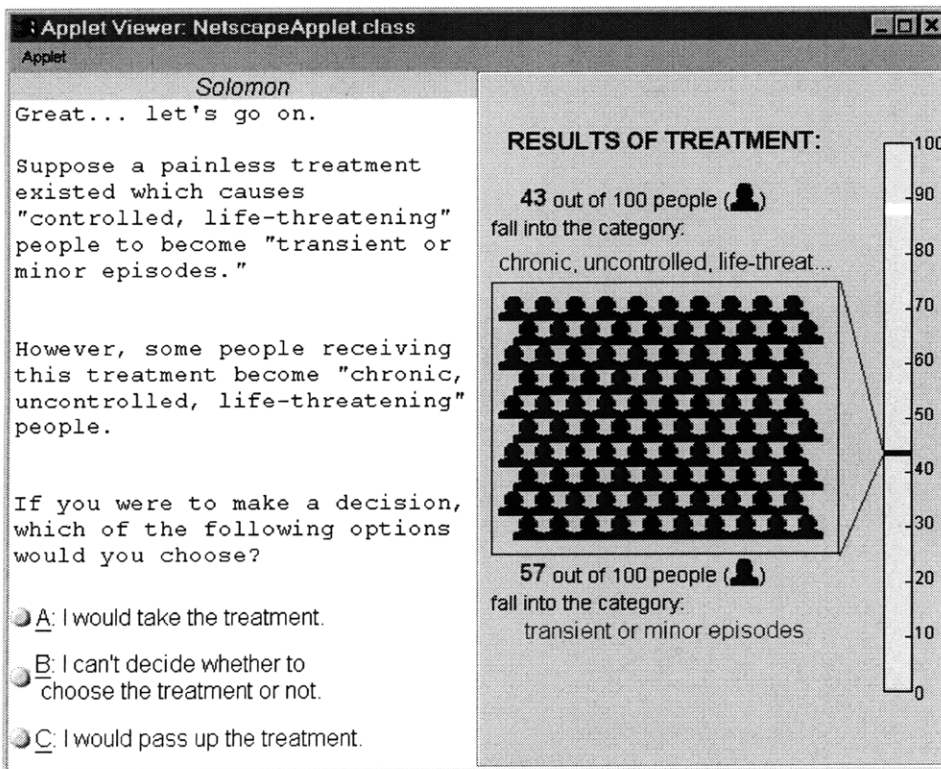
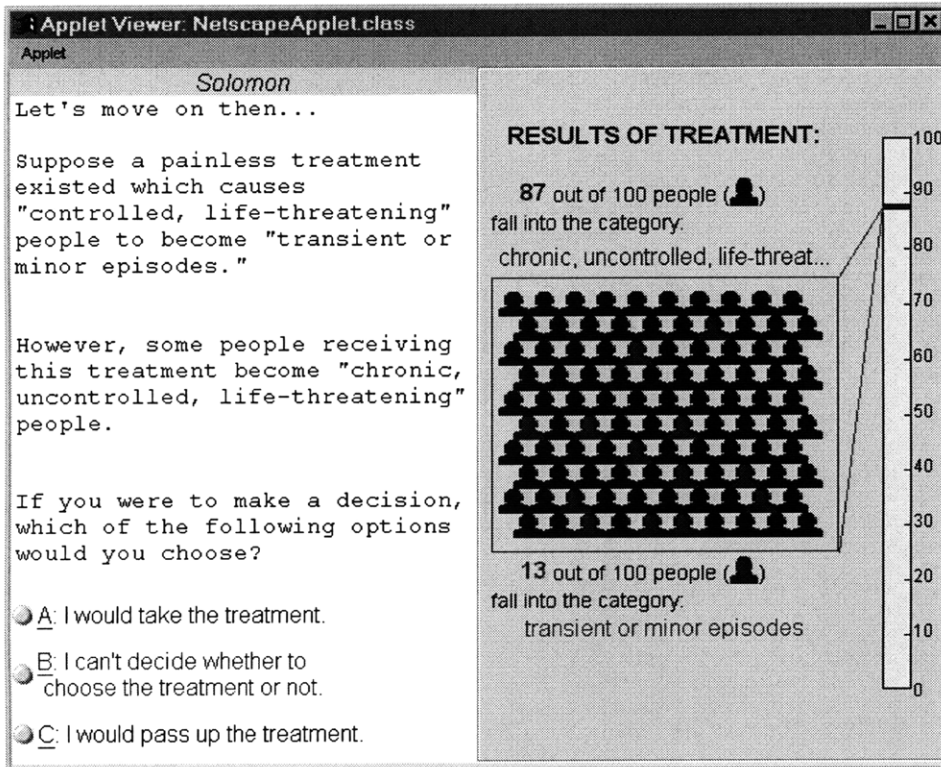












Applet Viewer: NetscapeApplet.class

Applet

Solomon

Great... let's go on.

Suppose a painless treatment existed which causes "controlled, life-threatening" people to become "transient or minor episodes."

However, some people receiving this treatment become "chronic, uncontrolled, life-threatening" people.


If you were to make a decision, which of the following options would you choose?

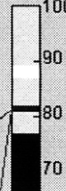
A: I would take the treatment.

B: I can't decide whether to choose the treatment or not.


C: I would pass up the treatment.

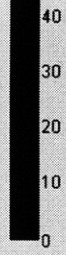
RESULTS OF TREATMENT:

81 out of 100 people () fall into the category: chronic, uncontrolled, life-threat...



You would NOT AGREE to the treatment if 68 people out of 100 are in the category chronic, uncontrolled, life-threatening and the rest are in the category transient or minor episodes. You would rather remain in the category controlled, life-threatening.

19 out of 100 people () fall into the category: transient or minor episodes



Applet Viewer: NetscapeApplet.class

Applet

Solomon

Great... let's go on.

Suppose a painless treatment existed which causes "controlled, life-threatening" people to become "transient or minor episodes."

However, some people receiving this treatment become "chronic, uncontrolled, life-threatening" people.


If you were to make a decision, which of the following options would you choose?

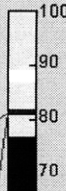
A: I would take the treatment.


B: I can't decide whether to choose the treatment or not.

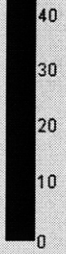
C: I would pass up the treatment.

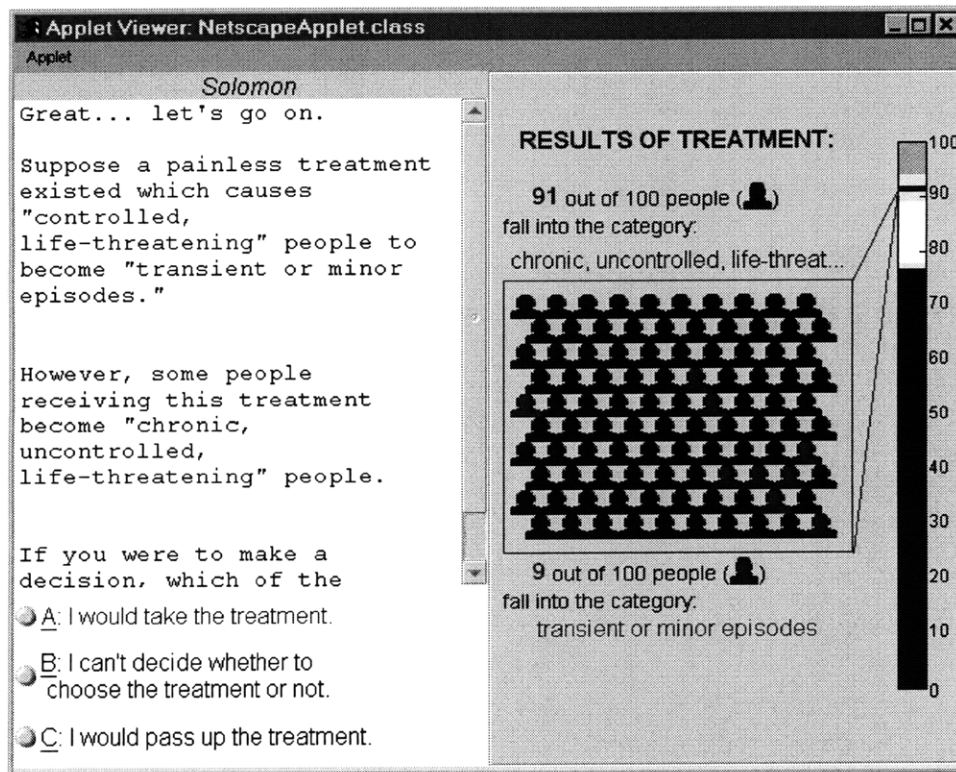
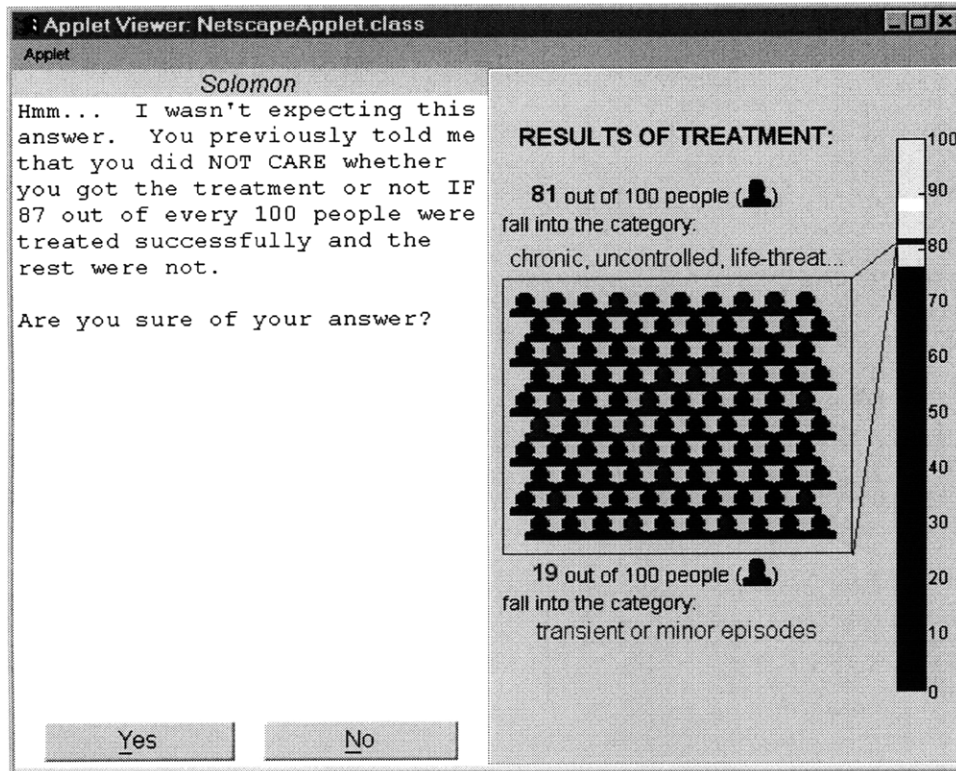
RESULTS OF TREATMENT:

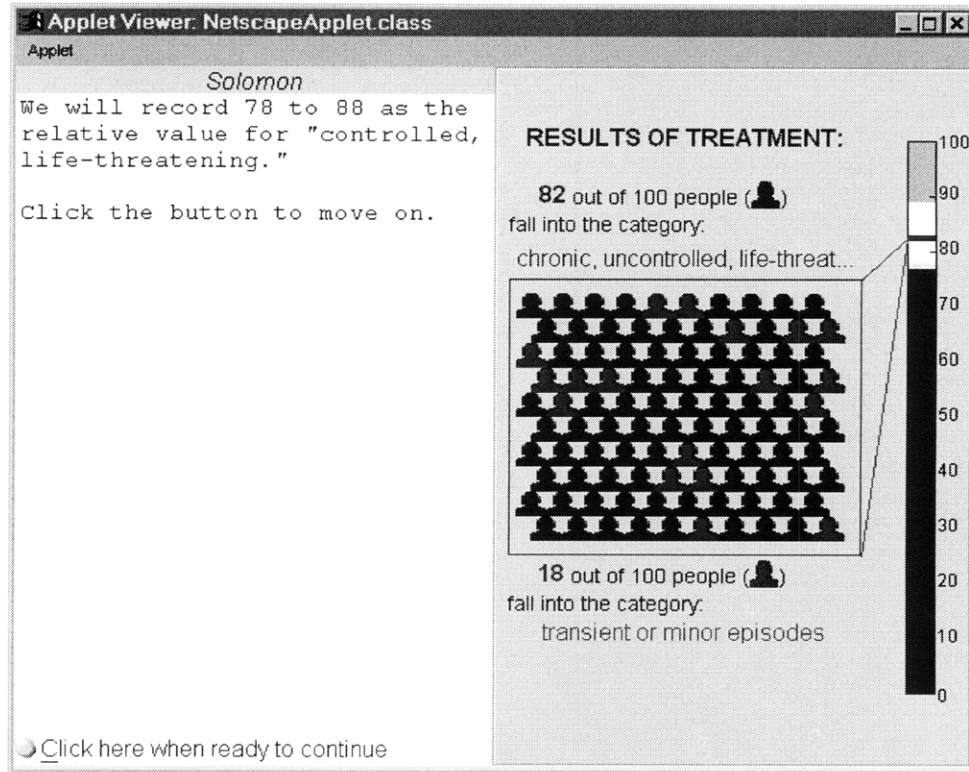
81 out of 100 people () fall into the category: chronic, uncontrolled, life-threat...



19 out of 100 people () fall into the category: transient or minor episodes







Applet Viewer: NetscapeApplet.class

Applet

Solomon

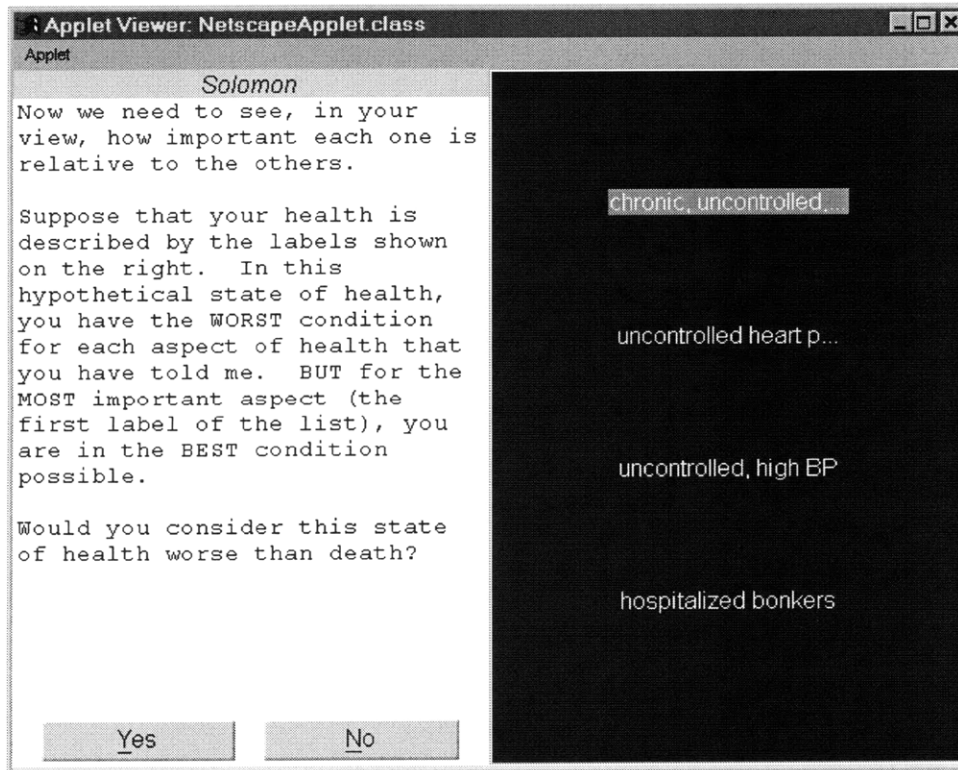
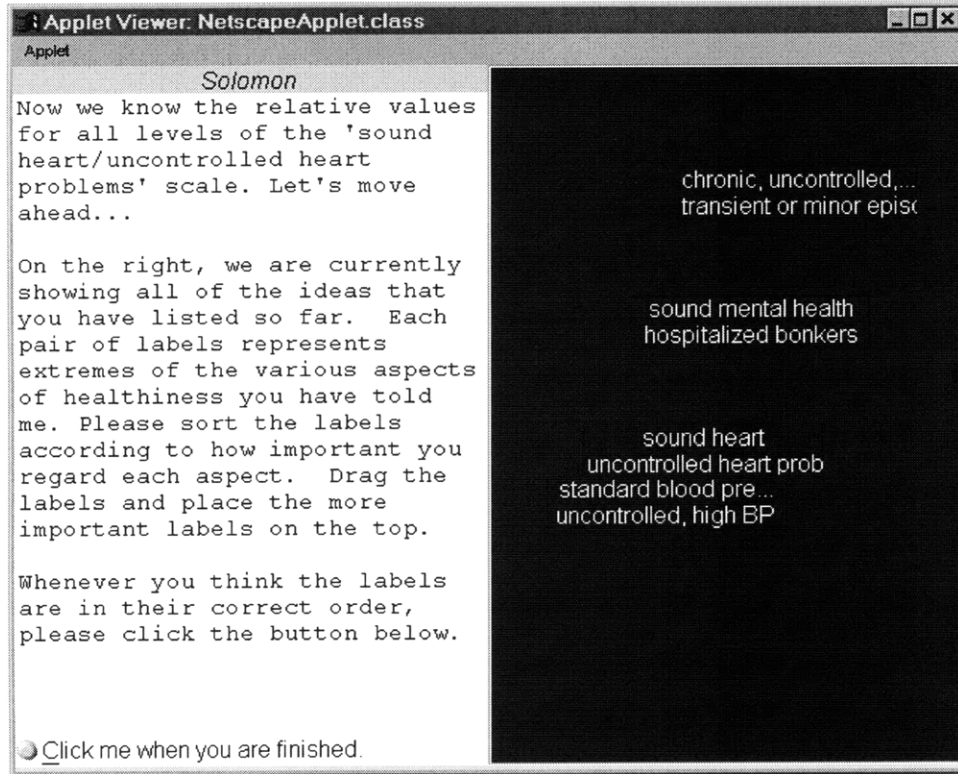
Now we know the relative values for all levels of the 'sound heart/uncontrolled heart problems' scale. Let's move ahead...

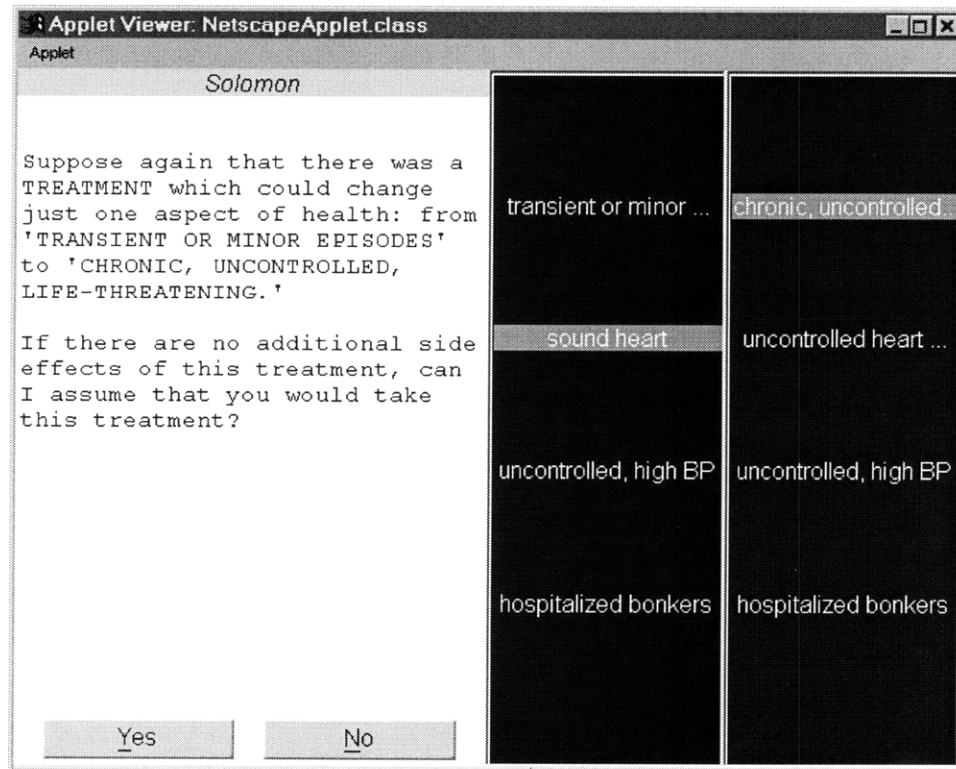
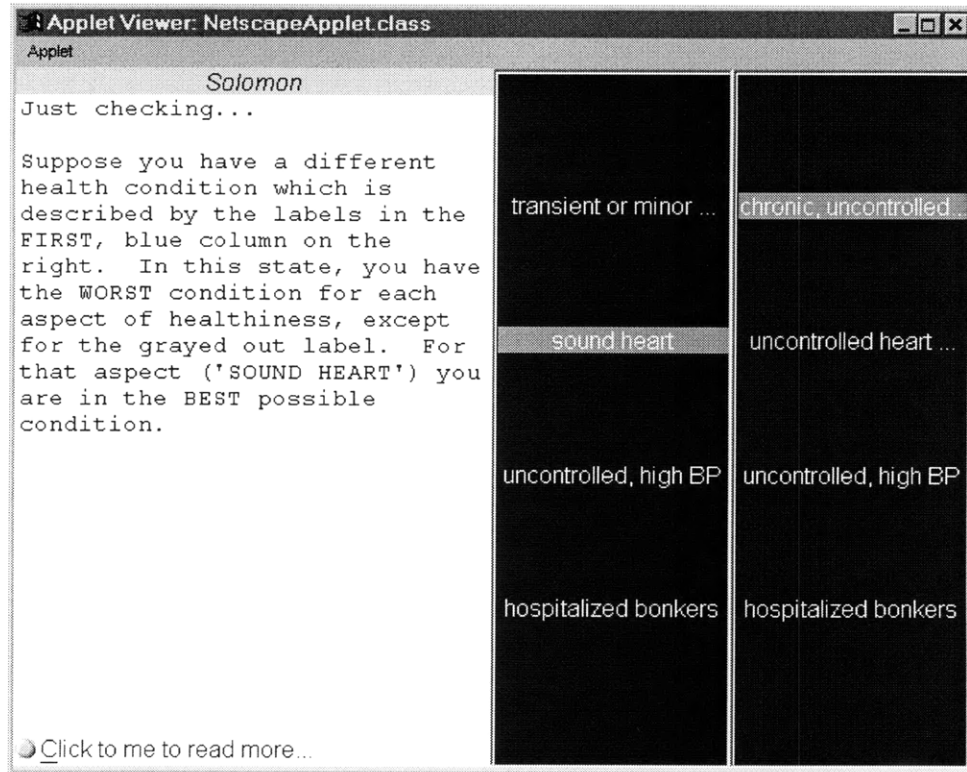
On the right, we are currently showing all of the ideas that you have listed so far. Each pair of labels represents extremes of the various aspects of healthiness you have told me. Please sort the labels according to how important you regard each aspect. Drag the labels and place the more important labels on the top.

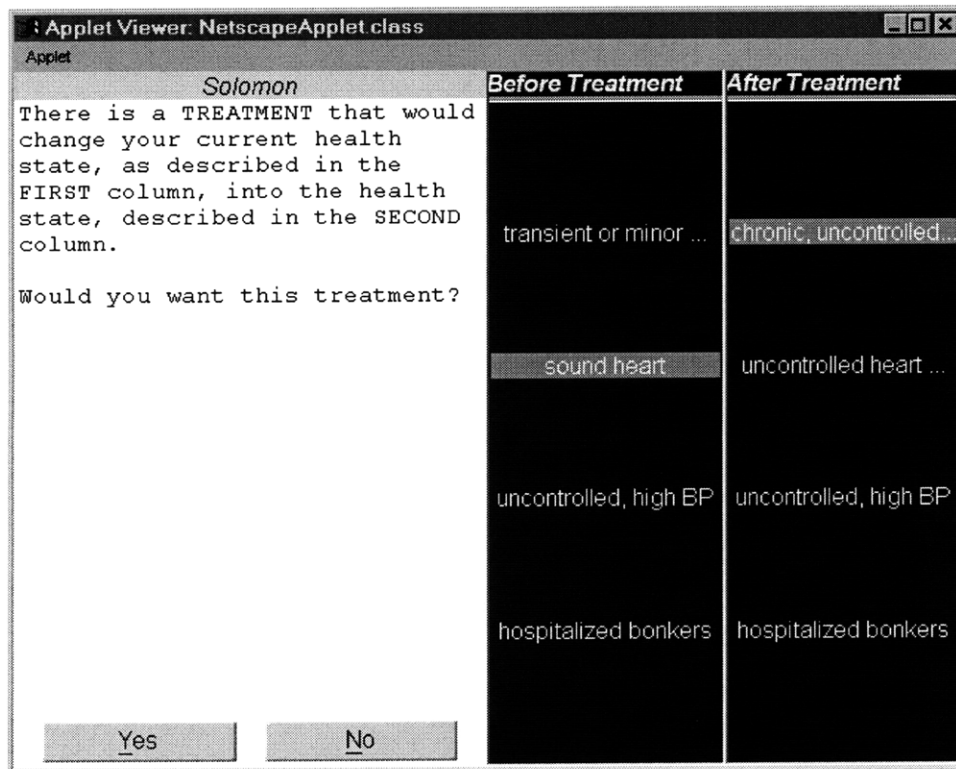
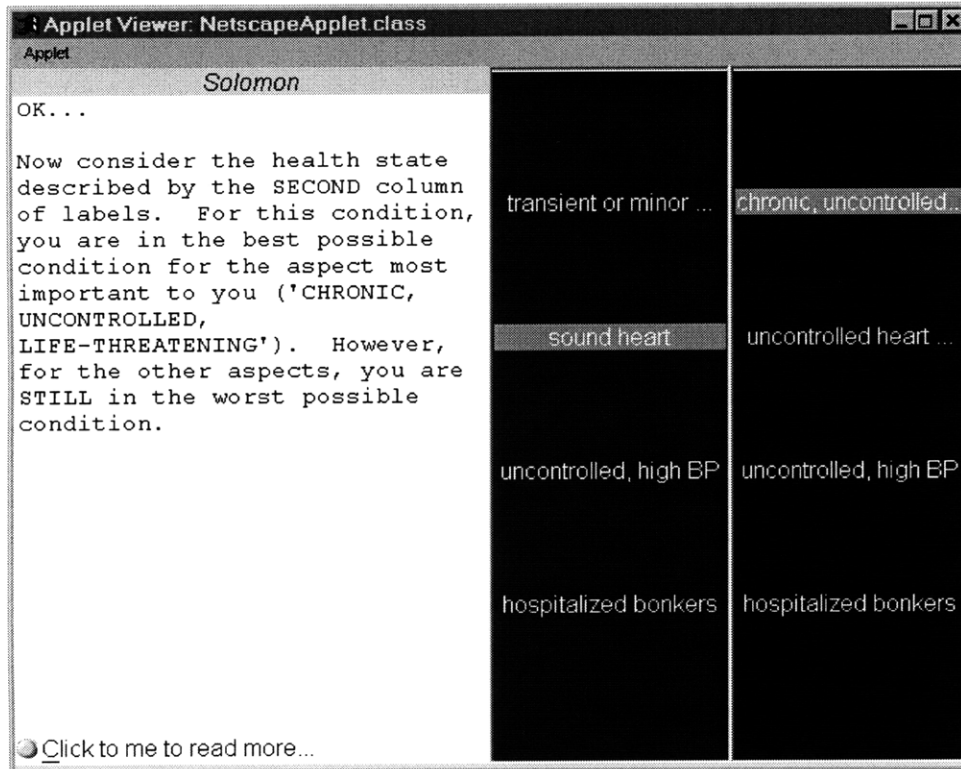
Whenever you think the labels are in their correct order, please click the button below.

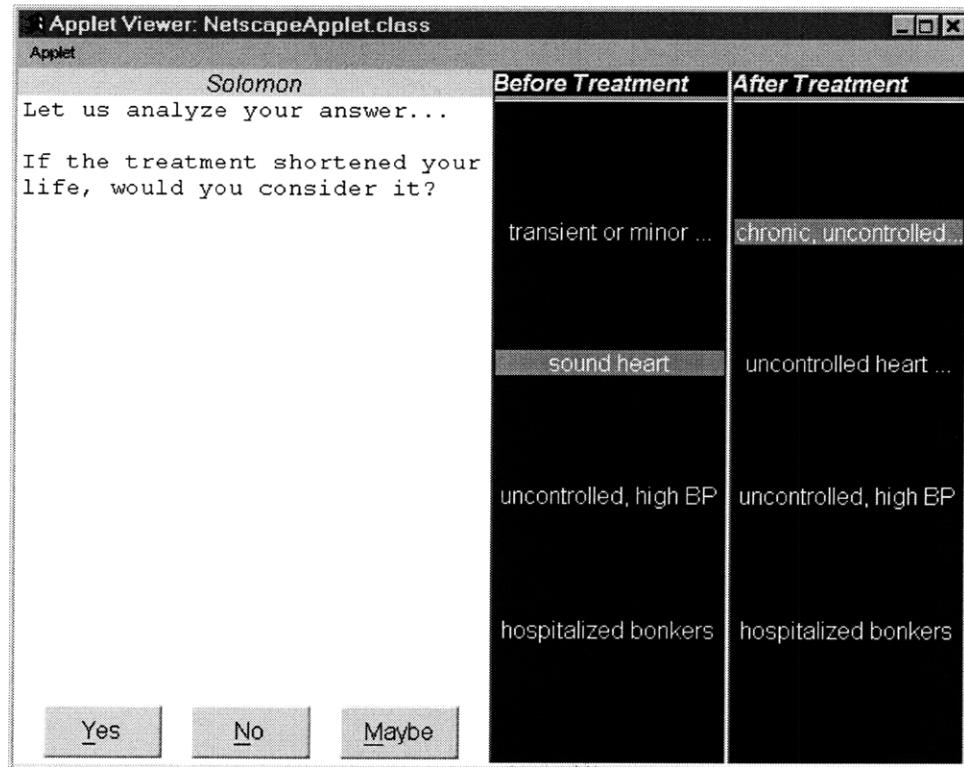
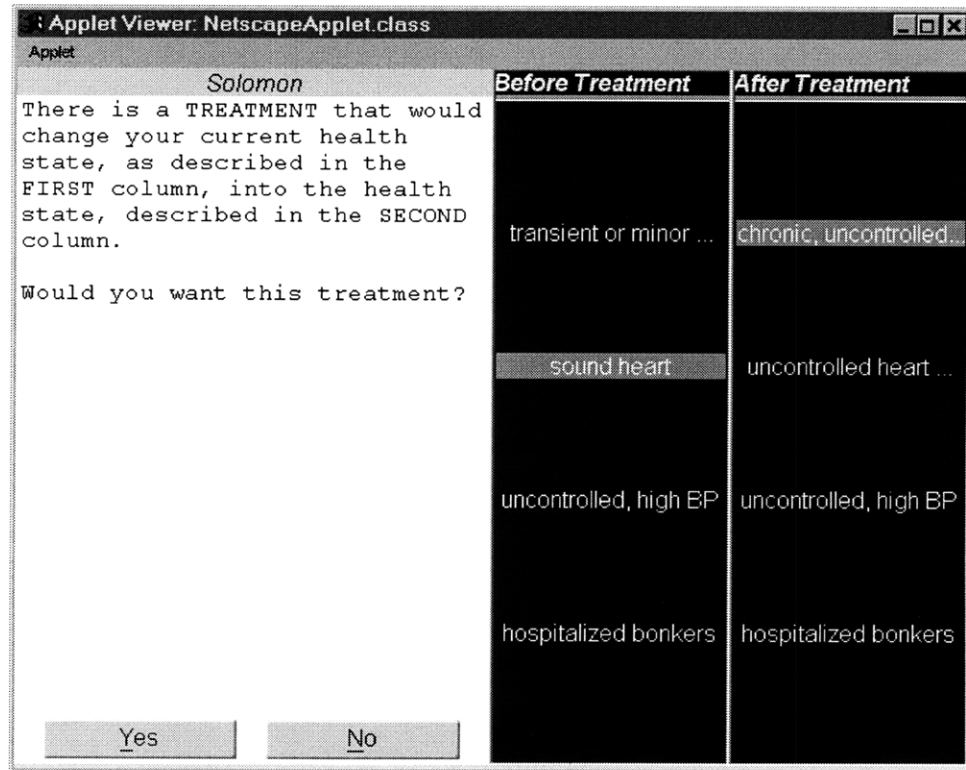
Click me when you are finished.

The right side of the applet window contains a dark vertical panel with several pairs of text labels. From top to bottom, the labels are: 'chronic, uncontrolled, ...' and 'transient or minor episic'; 'sound mental health' and 'hospitalized bonkers'; 'standard blood pre...' and 'uncontrolled, high BP'; and 'sound heart' and 'uncontrolled heart prob'.









Applet Viewer: NetscapeApplet class

Applet

Solomon

Let us analyze your answer...

Suppose, for the health state described in the FIRST column, the average life expectancy is 30 years.

Keeping in mind that AFTER the treatment your health would be as described in the SECOND column, please consider the following question carefully:

If the life span after the treatment is SHORTER than 30 years, at what point would you no longer have interest in the TREATMENT?

Life span: 7 Years

Click me to go on.

Before Treatment	After Treatment
transient or minor ...	chronic, uncontrolled...
sound heart	uncontrolled heart ...
uncontrolled, high BP	uncontrolled, high BP
hospitalized bonkers	hospitalized bonkers
Life Expectancy: 30 years	Life Expectancy: ???

Applet Viewer: NetscapeApplet class

Applet

Solomon

Just to make sure I hear you, let me rephrase what you told me:

You would PREFER to be treated so that your health is as described in the SECOND column, AND you live 7 years.

Is this correct?

Yes No

Before Treatment	After Treatment
transient or minor ...	chronic, uncontrolled...
sound heart	uncontrolled heart ...
uncontrolled, high BP	uncontrolled, high BP
hospitalized bonkers	hospitalized bonkers
Life Expectancy: 30 years	Life Expectancy: ???

Applet Viewer: NetscapeApplet.class

Applet

Salomon

I understand. Let's go on...

After the treatment, you now have a health state as described by the labels in the SECOND column. However, your life expectancy is now 3 years and 6 months.

Keeping in mind that the alternative is to live 30 years in a health state described by the FIRST column, would you accept this treatment?

Before Treatment	After Treatment
transient or minor ...	chronic, uncontrolled...
sound heart	uncontrolled heart ...
uncontrolled, high BP	uncontrolled, high BP
hospitalized bonkers	hospitalized bonkers
Life Expectancy: 30 years	Life Expectancy: 3 years and 6 months ??

Yes No Maybe

Applet Viewer: NetscapeApplet.class

Applet

Salomon

I will record 7 years as the MINIMUM life expectancy you would REQUIRE after the treatment.

In other words, you would PREFER to live 7 years in the health state described in the SECOND column MORE THAN living 30 years in a health state described in the FIRST column.

Do you agree?

Before Treatment	After Treatment
transient or minor ...	chronic, uncontrolled...
sound heart	uncontrolled heart ...
uncontrolled, high BP	uncontrolled, high BP
hospitalized bonkers	hospitalized bonkers
Life Expectancy: 30 years	Life Expectancy: 7 years

Yes No

