The Generalizability of Multi-Attribute Utility Functions

David Feeny, PhD
Senior Investigator, The Center for Health Research,
Kaiser Permanente Northwest;
University of Alberta; Health Utilities Incorporated

Ethical Issues in the Measurement of Health and the Global Burden of Disease, Harvard University, April 24-25, 2008

Center for Health Research

The Generalizability of Multi-Attribute Utility Functions

Outline

I. Introduction

II. Group- and Individual-Level Agreement

III. Comparing Multi-Attribute Utility Functions for the Same System Estimated Using Preference Scores from Different Groups/Countries

IV. Selecting the Right Functional Form

V. Are Differences in Preferences for Health States Related to Demographic and Other Characteristics?

VI. Minority View? Differences in the Structure of Preferences for Health States: States Worse than Dead

VII. Summary and Conclusions

Center for Health Research

Group-Level Agreement Between Standard Gamble and Multi-Attribute Utility Function Scores for HUI2 and HUI3

Group	Mean Utility Score	SD
ELBW, n = 140		
HUI2	0.89	0.14
Standard Gamble	0.90	0.20
HUI3	0.80*	0.22
Controls, n = 124		
HUI2	0.95	0.09
Standard Gamble	0.91	0.16
HUI3	0.89*	0.13
Combined, n = 264		
HUI2	0.92	0.13
Standard Gamble	0.91	0.16
HUI3	0.84*	0.19

Note: ELBW = extremely low birthweight. Source: Feeny et al. 2004.
*Statistically significant difference between mean HUI3 and SG score, $p<0.05$.

Center for Health Research

Individual-Level Agreement Between Standard Gamble and Multi-Attribute Utility Function Scores for HUI2 and HUI3

ICCs and results from two-factor ANOVA

Cohort	Scoring method comparisons	ANOVA factors	ICC	
			Point estimate	95% CI
Directly measured SG scores for subjectively defined current health state and HUI scores for current health				
ELBW	SG vs. HUI2	R* & M	0.13	0.00–0.25
	SG vs. HUI3	R* & M*	0.28	0.00–0.40
	HUI2 vs. HUI3	R* & M*	0.64	0.01–0.74
Controls	SG vs. HUI2	R* & M*	0.14	0.00–0.26
	SG vs. HUI3	R* & M*	0.24	0.00–0.36
	HUI2 vs. HUI3	R* & M*	0.56	0.00–0.68
Combined	SG vs. HUI2	R* & M	0.15	0.12–0.25
	SG vs. HUI3	R* & M*	0.29	0.00–0.37
	HUI2 vs. HUI3	R* & M*	0.63	0.01–0.73
Directly measured SG scores for four hypothetical health states and HUI2 scores for the same health states				
ELBW		R* & M*	0.11	0.00–0.26
Control		R* & M*	0.10	0.00–0.26
Combined		R* & M*	0.10	0.00–0.22

M: scoring method, R: respondents; $p<0.05$.

Source: Feeny et al. 2004.

Center for Health Research

Group- and Individual-Level Agreement, Elective Total Hip Arthroplasty

Standard Gamble (SG), HUI2, and HUI3 Scores at Assessments 1 and 2

Assessment period	No.	Mean	SD	Agreement	ICC
A1					
SG	103	0.62	0.31	--	--
HUI2	103	0.62	0.19	SG-HUI2	0.09
HUI3	103	0.52	0.21	SG-HUI3	0.09
A2					
SG	84	0.67	0.30	--	--
HUI2	84	0.68	0.18	SG-HUI2	0.06
HUI3	84	0.58	0.22	SG-HUI3	0.09

Note: SG means standard gamble utility score for subjectively-defined current health state; HUI = Health Utilities Index Mark; No. = number of patients; SD = standard deviation. The utility score for dead = 0.00; the utility score for perfect health = 1.00.

Center for Health Research

Overall Gain in Utility (Post-Surgery minus Pre-Surgery) Associated with Elective THA

SF-6D	0.10
Standard Gamble	0.16
HUI2	0.22
HUI3	0.23

Source: Feeny et al. 2004.

Center for Health Research

Types of MAUT Models

Let $u_j(x_j)$ = the single attribute utility function for attribute j.

Additive:

$$u(\underline{x}) = \sum k_j u_j(x_j)$$
$$\text{where } \sum k_j = 1$$

Multiplicative:

$$u(\underline{x}) = (1/k)\left[\prod (1 + k\, k_j\, u_j(x_j)) - 1\right]$$
$$\text{where } (1 + k) = \prod (1 + k\, k_j)$$

Multilinear:

$$u(\underline{x}) = k_1 u_1(x_1) + k_2 u_2(x_2) + \ldots$$
$$+ k_{12} u_1(x_1) u_2(x_2) + k_{13} u_1(x_1) u_3(x_3) + \ldots$$
$$K_{123} u_1(x_1) u_2(x_2) u_3(x_3) + \ldots$$
$$+ \ldots$$
$$\text{where } \sum \text{All } k\text{'s} = 1$$

Center for Health Research

Mean Measured Values for Single-Attribute Levels, HUI2

Attribute	Level	PAT Mean	GP Mean
1. Sensation	1	1.00	1.00
	2	0.57	0.59
	3	0.37	0.36
	4	0.00	0.00
2. Mobility	1	1.00	1.00
	2	0.74	0.68
	3	0.34	0.34
	4	0.14	0.17
	5	0.00	0.00
3. Emotion	1	1.00	1.00
	2	0.52	0.58
	3	0.28	0.33
	4	0.15	0.18
	5	0.00	0.00
4. Cognition	1	1.00	1.00
	2	0.57	0.58
	3	0.32	0.38
	4	0.00	0.00
5. Self-care	1	1.00	1.00
	2	0.55	0.56
	3	0.26	0.29
	4	0.000	0.00
6. Pain	1	1.00	1.00
	2	0.73	0.72
	3	0.43	0.45
	4	0.16	0.21
	5	0.00	0.00
7. Fertility	1	1.00	1.00
	2	0.45	0.45
	3	0.00	0.00

Note: By definition, single-attribute scores for level 1 are equal to 1.00 and single-attribute scores for the lowest level are 0.00. PAT = parents of patients sample (n = 41); GP = general population parents sample (n = 203).

Source: Wang et al. 2002

Center for Health Research

Mean Measured Values and Utilities for Multi-Attribute Health States, HUI2

Health State		PAT Mean Value	GP Mean Value
1. Perfect health	(1,1,1,1,1,1,1)	1.00	1.00
2. Sensation 4/4	(4,1,1,1,1,1,1)	0.30	0.33
3. Mobility 3/5	(1,3,1,1,1,1,1)	0.52	0.54
4. Emotion 5/5	(1,1,5,1,1,1,1)	0.30	0.27
5. Cognition 3/4	(1,1,1,3,1,1,1)	0.57	0.59
6. Self-care and mobility 5/5	(1,5,1,1,4,1,1)	0.24	0.22
7. Pain 4/5	(1,1,1,1,1,1,4)	0.32	0.34
8. Fertility 3/3	(1,1,1,1,1,1,3)	0.67	0.61
9. Interior 1	(1,4,2,1,1,1,1)	0.39	0.34
10. Interior 2	(1,1,3,2,1,1,1)	0.45	0.41
11. Interior 3	(3,3,2,3,3,2,2)	0.24	0.21
12. Interior 4	(3,3,4,4,4,4,3)	0.09	0.09
13. Dead		0.04	0.02
14. All worst	(4,5,5,4,4,5,3)	0.00	0.00

Health State		Mean Utility	Mean Utility
1. Perfect health	(1,1,1,1,1,1,1)	1.00	1.00
3. Mobility 3/5	(1,3,1,1,1,1,1)	0.83	0.78
8. Fertility 3/3	(1,1,1,1,1,1,3)	0.89	0.88
9. Interior 1	(1,4,2,1,1,1,1)	0.79	0.76
11. Interior 3	(3,3,2,3,3,2,2)	0.46	0.51
13. Dead		0.00	0.00

Note: Value scores are measured using the Feeling Thermometer, and utility scores are measured using the standard gamble. Value scores are reported on the all-worst = 0.00 to perfect health = 1.00 scale. Utility scores are reported on the dead = 0.00 to perfect health = 1.00 scale. PAT = parents of patients sample ($n = 41$ for value scores, $n = 39$ for utility scores); GP = general population parents sample ($n = 203$ for value scores, $n = 194$ for utility scores).

Source: Wang et al. 2002

Center for Health Research

Parameter Estimates for Multiattribute Disutility Functions, HUI2

Parameter	Parent of Patients	General Population Parents
c	–0.966	–0.967
c_1	0.45	0.40
c_2	0.44	0.43
c_3	0.48	0.49
c_4	0.35	0.36
c_5	0.19	0.21
c_6	0.63	0.64
c_7	0.08	0.12
Σc_j	2.62	2.65

Note: c_i is the disutility version of k_j

Source: Wang et al. 2002

Center for Health Research

Agreement Between Scores Derived from Multiattribute Utility Functions and Directly Measured Utility Scores for 4 Validation States, HUI2, ICCs

Utility Scores Calculated by Function	Directly Measured Utility Scores: Parents of Patients	Directly Measured Utility Scores: General Population Parents
Parents of patients	0.89	0.87
General population parents	0.94	0.94

Source: Wang et al. 2002

Center for Health Research

Multiattribute Disutility Function: Standard Format on All-Worst – Perfect Health Scale, HUI3

Canada and (France)

	Vision		Hearing		Speech		Ambulation		Dexterity		Emotion		Cognition		Pain	
	x_1	$\bar{u}_1$	x_2	$\bar{u}_2$	x_3	$\bar{u}_3$	x_4	$\bar{u}_4$	x_5	$\bar{u}_5$	x_6	$\bar{u}_6$	x_7	$\bar{u}_7$	x_8	$\bar{u}_8$
1	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
2	0.05	(0.07)	0.14	(0.18)	0.18	(0.18)	0.17	(0.16)	0.12	(0.19)	0.09	(0.12)	0.14	(0.14)	0.08	(0.12)
3	0.27	(0.28)	0.29	(0.23)	0.33	(0.29)	0.33	(0.31)	0.27	(0.33)	0.27	(0.42)	0.08	(0.23)	0.23	(0.26)
4	0.41	(0.39)	0.52	(0.39)	0.59	(0.44)	0.64	(0.49)	0.55	(0.46)	0.67	(0.63)	0.30	(0.37)	0.52	(0.55)
5	0.62	(0.59)	0.69	(0.50)	1.00	(1.00)	0.84	(0.63)	0.80	(0.60)	1.00	(1.00)	0.69	(0.56	1.00	(1.00)
6	1.00	(1.00)	1.00	(1.00)	n/a	n/a	1.00	(1.00)	1.00	(1.00)	n/a	n/a	1.00	(1.00)	n/a	n/a

Parameter estimates for multiattribute disutility function

c = -0.991 (-0.990)

c1 = 0.40 (0.51)	c2 = 0.39 (0.34)	c3 = 0.33 (0.37)	c4 = 0.42 (0.43)
c5 = 0.44 (0.39)	c6 = 0.54 (0.55)	c7 = 0.58 (0.44)	c8 = 0.45 (0.50)

Sources: For Canada, Feeny et al. 2002 and Furlong et al. 1998; for France, Le Gales et al. 2001, 2002.

Center for Health Research

Per Cent of Respondents who Rated the All-Worst HUI3 State As Worse Than Dead

Canada: 91%

France: 93%

Note: For France person-mean and person-median functions very similar.

Sources: For Canada, Furlong et al. 1998; for France, Le Gales et al. 2001, 2002.

Center for Health Research

Additional Evidence That Multi-Attribute Utility Functions Estimated using Preferences from Different Groups/Regions Are Similar

Quality of Well-Being Scale

Comparisons of the original version

(general population in San Diego) with:

Arthritis patients in NE USA

General population in Oregon

Health Utilities Index Mark 3

Comparison of original version for Canada with:

General Population in the Netherlands

Center for Health Research

HUI2 Measured Values for Levels Within Attributes, Canada and UK

		VAS (n = 203) Canada	VAS (n = 176) UK
Attribute	Level	Mean	Mean
Sensation	1	1.00	1.00
	2	0.59	0.56
	3	0.36	0.36
	4	0.00	0.00
Mobility*	1	1.00	1.00
	2	0.68	0.74
	3	0.34	0.46
	4	0.17	0.30
	5	0.00	0.00
Emotion	1	1.00	1.00
	2	0.58	0.63
	3	0.33	0.38
	4	0.18	0.22
	5	0.00	0.00
Cognition	1	1.00	1.00
	2	0.58	0.60
	3	0.38	0.36
	4	0.00	0.00
Self-care	1	1.00	1.00
	2	0.56	0.60
	3	0.29	0.30
	4	0.00	0.00
Pain	1	1.00	1.00
	2	0.72	0.72
	3	0.45	0.45
	4	0.21	0.21
	5	0.00	0.00
Fertility	1	1.00	na
	2	0.45	na
	3	0.00	na

VAS = visual analogue scale.

*The mobility results have been corrected for confounding with self-care.

na = Not assessed.

Sources: UK Results from McCabe et al.2005; Canadian results from Torrance et al. 1996.

Center for Health Research

Multiattribute Utility Function for HUI2 Using Power Curve VAS-SG mapping Function, UK and (Canada)

Multiattribute Utility Function Using Power Function (Disutility Formulation)

Level	Sensation	Mobility	Emotion	Cognition	Self-Care	Pain	Fertility
1	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	n/a (0.00)
2	0.29 (0.13)	0.26 (0.08)	0.23 (0.14)	0.26 (0.14)	0.25 (0.15)	0.15 (0.05)	n/a (0.25)
3	0.51 (0.35)	0.75 (0.39)	0.49 (0.40)	0.52 (0.34)	0.59 (0.45)	0.40 (0.25)	n/a (1.00)
4	1.00 (1.00)	0.94 (0.66)	0.69 (0.63)	1.00 (1.00)	1.00 (1.00)	0.66 (0.58)	
5		1.00 (1.00)	1.00 (1.00)			1.00 (1.00)	

c	c1	c2	c3	c4	c5	c6	c7
-0.935 (-0.967)	0.38 (0.40)	0.27 (0.43)	0.47 (0.49)	0.32 (0.36)	0.13 (0.21)	0.65 (0.64)	n/a (0.12)
Death	0.96 (0.98)						

c and c1 to c7 are the paremeters in the multiattribute disutility function and are equivalent to k and kj in the multiattribute utility function.
n/a = not assessed.
Sources: UK results from McCabe et al. 2005; Canadian results from Torrance et al 1996.

Center for Health Research

Comparing British and Japanese EQ-SD Scoring Functions

	Japanese model[b]				British model			
	Weighted			Non-weighted	42-state model		17-state model	
	Coeff.	SE	*p*-value	Coeff.	Coeff.	*p*-value[c]	Coeff.	*p*-value[d]
Constant	0.148	0.006	0.000	0.148	0.081	0.000	0.075	0.000
*M*2	0.078	0.011	0.000	0.078	0.069	0.389	0.058	0.578
*M*3	0.418	0.016	0.000	0.416	0.314	0.000	0.292	0.000
SC2	0.053	0.010	0.000	0.053	0.104	0.000	0.110	0.000
SC3	0.101	0.014	0.000	0.101	0.214	0.000	0.201	0.000
UA2	0.040	0.009	0.000	0.041	0.036	0.677	0.039	0.923
UA3	0.128	0.015	0.000	0.130	0.094	0.029	0.043	0.000
PD2	0.083	0.010	0.000	0.083	0.123	0.000	0.110	0.009
PD3	0.189	0.013	0.000	0.190	0.386	0.000	0.348	0.000
AD2	0.062	0.009	0.000	0.063	0.071	0.305	0.067	0.577
AD3	0.108	0.012	0.000	0.110	0.236	0.000	0.279	0.000
*N*3	0.014	0.013	0.284	0.013	0.269	0.000	0.320	0.000
Adjusted R^2		0.400		0.401		0.46		Unavailable

[a] *Keys*. *M*: mobility dimension; *SC*: self-care dimension; *UA*: usual activities dimension; *PD*: pain/discomfort dimension; *AD*: anxiety/depression dimension; *N*3: dummy representing whether there is any dimension on level 3.
[b] Estimated using OLS, and weights to correct for sample representativeness.
[c] Based on *F*-tests on the null hypothesis: Japanese weighted coefficient = corresponding British coefficient based on the original 42 health states.
[d] Based on *F*-tests on the null hypothesis: Japanese weighted coefficient = corresponding British coefficient based on the 17 health states.

Source: Tsuchiya et al. 2002.

Center for Health Research

Additional Evidence That Multi-attribute Utility Functions Are Not Similar

EQ-5D

Comparisons of EQ-5D functions for the UK with:

Spain

Germany

United States

Center for Health Research

Thought Experiment, Part A

Part A.

Imagine that you go from perfect health in the EQ-5D system (1,1,1,1,1) to “I have some problems in walking about,” otherwise healthy (2,1,1,1,1)

What score do you attach to this (2,1,1,1,1) health state?

Please Think About the Decrement in Your Health-Related Quality of Life

Center for Health Research

Thought Experiment, Part B

Part B.
Imagine that you have the health state (1,2,2,2,2): no problems with mobility, but some problems with self-care, usual activities, pain and discomfort, and anxiety/depression.

What score do you attach to this health state?

Now imagine that you go from (1,2,2,2,2) to the health state (2,2,2,2,2), some problems in all 5 dimensions.

What score do you attach to this (2,2,2,2,2) health state?

Please Think About the Decrement in Your Health-Related Quality of Life

Center for Health Research

Thought Experiment, Part C

Part C.

Please compare the decrement in Part A

to the decrement in Part B

Center for Health Research

Types of MAUT Models

Let $u_j(x_j)$ = the single attribute utility function for attribute j.

Additive:

$$u(\underline{x}) = \sum k_j u_j(x_j)$$
$$\text{where } \sum k_j = 1$$

Multiplicative:

$$u(\underline{x}) = (1/k)\left[\prod (1 + k\, k_j\, u_j(x_j)) - 1\right]$$
$$\text{where } (1 + k) = \prod (1 + k\, k_j)$$

Multilinear:

$$u(\underline{x}) = k_1 u_1(x_1) + k_2 u_2(x_2) + \ldots$$
$$+ k_{12} u_1(x_1) u_2(x_2) + k_{13} u_1(x_1) u_3(x_3) + \ldots$$
$$K_{123} u_1(x_1) u_2(x_2) u_3(x_3) + \ldots$$
$$+ \ldots$$
$$\text{where } \sum \text{All } k\text{'s} = 1$$

Center for Health Research

Estimation of Multi-Linear Function for HUI3 System

Using 2^8 fractional factorial design

Estimable coefficients:

- all main effects, 26 of 28 two-factor interactions and 4 of 56 three-factor interactions
- others are confounded.

Estimated both unrestricted and

restricted through 0 and 1.

Center for Health Research

Results of Estimation of HUI3 Multi-Linear Function

Interaction terms were significant; additive model was not supported.

MLF unrestricted: all 8 main effects, 16 two-factor interactions and 1 three-factor were quantitatively important (>0.025) and stat. sig. ($p<0.05$), $R^2 = 0.70$.

MLF restricted: all 8 main effects, all 26 estimable two-factor interactions, and all 4 estimable three-factor were important and statistically significant, $R^2 = 0.77$.

Center for Health Research

Agreement Between Calculated (Multiplicative Model) and Directly Measured HUI3 Utility Scores: Predictive Validity

Survey	States	MD	MAD	Overall SD	ICC (95% CI)
HUI3	73 (not weighted by GP prevalence)	-0.008	0.087	0.1032	0.88 (0.49, 0.92)
HUI3	73 (weighted by GP prevalence, excluding PH)	+0.001	0.002	0.0061	--

Legend: CI = confidence interval

GP = general population (from 1991 General Social Survey)

ICC = intra-class correlation coefficient

MAD = mean absolute difference = [$\sum$ (|predicted - 10% trimmed mean|)/n]

MD = mean difference = [$\sum$(predicted - 10% trimmed mean)/n]

Overall SD = overall standard deviation

= $\sqrt{[\sum(\text{predicted} - 10\% \text{ trimmed mean})^{2)/(n-1)}]}$

Source: Feeny et al. 2002.

Center for Health Research

Empirical Evidence on Functional Form

Linear additive rejected by results for:
HUI1
HUI2 [3 estimated functions]
HUI3 [3 estimated functions]
AQoL

Several disease-specific multi-attribute utility functions including
Rhinitis Symptom Utility Index
Asthma Symptom Utility Index

Functions that rely on linear additive or *ad hoc* modified linear additive functional form include:
EQ-5D
QWB
SF-6D

Center for Health Research

Are Differences in Preferences for Health States Related to Demographic and Other Characteristics?

Mixed evidence but in general a lack of consistent systematic relationship between preference scores for health states and a wide range of characteristics including:

Age
Gender
Income
Education
Religion
Frequency of participation in religious services
Current health state of respondent
Experience of respondent

Qualifications

Negative Studies:
Power

Positive Studies:
large sample sizes;
lack of *a priori* hypotheses;
multiple testing

Center for Health Research

Health States Worse Than Dead

Definitions of Perfect health (PH) and the all-worst HUI3 health state

	Level	Level Description
Perfect Health	1	Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses
	1	Able to hear what is said in a group conversation with at least three other people, without a hearing aid
	1	Able to be understood completely when speaking with strangers or friends
	1	Able to walk around the neighborhood without difficulty and without walking equipment
	1	Full use of two hands and ten fingers
	1	Happy and interested in life
	1	Able to remember most things, think clearly and solve day to day problems
	1	Free of pain and discomfort
All-Worst Health	6	Unable to see at all
	6	Unable to hear at all
	5	Unable to be understood when speaking to other people (or unable to speak at all)
	6	Cannot walk at all
	6	Limitations in the use of hands or fingers, requires the help of another person for all tasks (not independent even with the use of special tools)
	5	So unhappy that life is not worthwhile
	6	Unable to remember anything at all, and unable to think or solve day to day problems
	5	Severe pain that prevents most activities

Source: http://www.fhs.mcmaster.ca/hug/index.htm

Differences in disutility scores and parameter estimates between MAUF(A) and MAUF(B), HUI3

	Vision	Hearing	Speech	Ambulation	Dexterity	Emotion	Cognition	Pain
Level	$\overline{u}_{1(A-B)}$	$\overline{u}_{2(A-B)}$	$\overline{u}_{3(A-B)}$	$\overline{u}_{4(A-B)}$	$\overline{u}_{5(A-B)}$	$\overline{u}_{6(A-B)}$	$\overline{u}_{7(A-B)}$	$\overline{u}_{8(A-B)}$
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	-0.36	-0.03	0.07	0.07	0.05	0.01	0.10	0.07
3	-0.24	0.02	0.16	0.12	0.08	0.07	-0.30	0.00
4	-0.26	0.02	0.23	0.00	0.20	0.28	-0.17	0.06
5	-0.05	0.06	0.00	0.05	0.20	0.00	0.16	0.00
6	0.00	0.00	n/a	0.00	0.00	n/a	0.00	n/a

$\overline{u}_{j(A-B)}$: differences in disutility scores between Groups A and B (A-B) for attribute j

	c	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	$\sum c_j$
Δ(A-B)	0.0077	-0.13	-0.16	0.20	-0.28	-0.07	-0.40	0.01	0.08	-0.75

Δ(A-B): differences in parameter estimates between Groups A and B (A-B)

Note: Group A rated all-worst HUI3 state < dead.
Group B rated dead < all-worst HUI3 state.
Source: Asakawa et al. 2008; not for quotation.

Center for Health Research

Summary and Conclusions

- Evidence of substantial heterogeneity in the preferences for health states
- Some evidence on agreement between scores from multi-attribute utility functions and group-mean scores from patients and other groups
- Little agreement at the individual level
- Evidence on the generalizability of multi-attribute utility functions is mixed
- Lack of attention to justification for choice of functional form and widespread use of linear additive utility functions may be a dis-service
- Focus to date has been on person-mean: one person, one vote
- Little attention to representing minority preference viewpoints.
- A recommendation: large representative (random) samples to capture minority views with precision

Center for Health Research