Eliciting Canadian population preferences for health states using the Classification and Measurement System of Functional Health (CLAMES)

Cameron N McIntosh, Sarah Connor Gorber, Julie Bernier and Jean-Marie Berthelot

Abstract

A major objective of the Population Health Impact of Disease in Canada (PHI) research program was to obtain Canadian-specific preferences for health states associated with various diseases, in order to estimate the morbidity component of summary measures of population health embodying the Canadian experience of disease. In this study, preferences for health states were elicited from lay panels (N = 146) in nine Canadian communities (Vancouver, Edmonton, Saskatoon, Toronto, Ottawa, Montréal, Québec, Moncton and Halifax); the study was conducted from January to June of 2003. Information on health states was presented to raters using the **CLA***ssification and* **ME***asurement* **S***ystem of Functional Health (CLAMES), which assesses functional capacity using 11 health status attributes, each with four to five levels ranging from normal to severely limited functioning. Preferences for 238 health states classified by CLAMES were elicited using the standard gamble (SG) technique in both individual and group exercises. Mean preferences for these health states were then used to estimate the parameters of a log-linear scoring function for CLAMES. The function provides a convenient method of computing preference scores for any health state classified by CLAMES, without the need for direct measurement in surveys. Further, the SG appears feasible in group settings.*

Key words: *Canadian health state preferences, classification and measurement system of functional health, Population Health Impact, preference-based scoring function, standard gamble*

Introduction

A major objective of the Population Health Impact of Disease in Canada (PHI) research program was to obtain Canadian-specific preferences for health states associated with various diseases, in order to estimate the morbidity component of summary measures of population health (SMPH) embodying the Canadian experience of disease. Health state preference scores quantify the perceived desirability of particular health states, typically in terms of a continuum bounded by 0 (i.e., death) and 1 (i.e., full health).[1,2]

Within the context of burden of disease research, health state preference scores are used to weight the time spent in sub-optimal health states, in order to compute SMPH that integrate information on both mortality and morbidity.[3]

Thus far, however, the health state preferences used in burden of disease studies have been largely those of medical experts,[4] who may not constitute a representative sample of the general population.[5] If health state preferences are to form part of the evidence base for broad health care policy and planning, then the preferences of those ultimately affected by any decisions in the health sector should figure into the process.[5-8]

The current article describes three methodological steps required to obtain Canadian-specific preferences for health states linked to different diseases: 1) the use of a generic tool—the CLAssification and MEasurement System of Functional Health (CLAMES)—for communicating information about health states to raters; 2) the implementation of standard gamble (SG) protocols for measuring health state preferences in panels of lay Canadians; and 3) the use of a preference-based scoring function to compute a tariff (i.e., a summary health-related quality of life [HRQoL] value) for all health states classified by the CLAMES instrument.

Materials and methods

The CLAMES instrument for classifying health states

To represent the impact of various diseases on physical, mental and social functioning and to convey information on health states to raters, a standardized tool was created, namely the CLAssification and MEasurement System of Functional Health (CLAMES; see Table 1). CLAMES contains 11 health status attributes borrowed and adapted from three leading generic health status instruments: the Health Utilities Index Mark III (HUI3),[9] the Medical Outcomes Study Short-Form 36 (SF-36)[10] and the European Quality of Life

Author References

Cameron N McIntosh, Sarah Connor Gorber, Julie Bernier, Jean-Marie Berthelot, Health Analysis and Measurement Group, Statistics Canada, Ottawa, Ontario, Canada
Correspondence: Cameron N McIntosh, Health Analysis and Measurement Group, Statistics Canada, 100 Tunney's Pasture Driveway, 24th floor, RH Coats Building, Ottawa, Ontario, Canada K1A 0T6; fax: (613) 951-3959; e-mail: cameron.mcintosh@statcan.ca

TABLE 1
The CLAssification and MEasurement System of Functional Health (CLAMES) Instrument

Attribute	Level	Description
Pain or discomfort*	1	Generally free of pain and discomfort
	2	Mild pain or discomfort
	3	Moderate pain or discomfort
	4	Severe pain or discomfort
Physical functioning**	1	Generally no limitations in physical functioning
	2	Mild limitations in physical functioning
	3	Moderate limitations in physical functioning
	4	Severe limitations in physical functioning
Emotional state*	1	Happy and interested in life
	2	Somewhat happy
	3	Somewhat unhappy
	4	Very unhappy
	5	So unhappy that life is not worthwhile
Fatigue**	1	Generally no feelings of tiredness, no lack of energy
	2	Sometimes feel tired and have little energy
	3	Most of the time feel tired and have little energy
	4	Always feel tired and have no energy
Memory and thinking*	1	Able to remember most things, think clearly and solve day-to-day problems
	2	Able to remember most things but have some difficulty when trying to think and solve day-to-day problems
	3	Somewhat forgetful, but able to think clearly and solve day-to-day problems
	4	Very forgetful, and have great difficulty when trying to think or solve day-to-day problems
Social relationships**	1	No limitations in the capacity to sustain social relationships
	2	Mild limitations in the capacity to sustain social relationships
	3	Moderate limitations in the capacity to sustain social relationships
	4	Severe limitations in the capacity to sustain social relationships
	5	No capacity or unable to relate to other people socially
Anxiety***	1	Generally not anxious
	2	Mild levels of anxiety experienced occasionally
	3	Moderate levels of anxiety experienced regularly
	4	Severe levels of anxiety experienced most of the time
Speech*	1	Able to be understood completely when speaking with strangers or friends
	2	Able to be understood partially when speaking with strangers but able to be understood completely when speaking with people who know you well
	3	Able to be understood partially when speaking with strangers and people who know you well
	4	Unable to be understood when speaking to other people
Hearing*	1	Able to hear what is said in a group conversation, without a hearing aid, with at least 3 other people
	2	Able to hear what is said in a conversation with 1 other person in a quiet room, with or without a hearing aid, but require a hearing aid to hear what is said in a group conversation with at least 3 other people
	3	Able to hear what is said in a conversation with 1 other person in a quiet room, with or without a hearing aid, but unable to hear what is said in a group conversation with at least 3 other people
	4	Unable to hear what others say, even with a hearing aid
Vision*	1	Able to see well enough, with or without glasses or contact lenses, to read ordinary newsprint and recognize a friend on the other side of the street
	2	Unable to see well enough, even with glasses or contact lenses, to recognize a friend on the other side of the street but can see well enough to read ordinary print
	3	Unable to see well enough, even with glasses or contact lenses, to read ordinary newsprint but can see well enough to recognize a friend on the other side of the street
	4	Unable to see well enough, even with glasses or contact lenses, to read ordinary newsprint or to recognize a friend on the other side of the street
Use of hands and fingers*	1	No limitations in the use of hands and fingers
	2	Limitations in the use of hands and fingers but do not require special tools or the help of another person
	3	Limitations in the use of hands and fingers, independent with special tools and do not require the help of another person
	4	Limitations in the use of hands and fingers, require the help of another person for some tasks
	5	Limitations in the use of hands and fingers, require the help of another person for most tasks

* Adapted from HU13

** Adapted from SF-36

***Adapted from EQ-5D

Five-Dimensions Index Plus (EQ-5D).[11,12] CLAMES focuses on individuals' capacities (i.e., what they are able to do) with respect to the various attributes, each of which has four or five levels ranging from normal to severely limited functioning. A complete health state is represented by an 11-tuple or list of attribute levels; thus, 10,240,000 health states are possible within the system.

The HUI3 was adapted (see Table 1) to broaden its scope by using attributes from the SF-36 and EQ-5D. The attribute "Social Relationships" was added to help classify health states in which limitations in the ability to maintain social relationships are a defining feature (e.g., Asperger's syndrome, schizophrenia). The HUI3 ambulation attribute was expanded to include a broader range of physical limita-

tions resulting from disease (e.g., stroke). The addition of attributes "Anxiety" and "Fatigue" also assisted in the classification of disease-related limitations. Focus groups with members of the lay population, as well as consultation with experts in multi-attribute health status instrumentation, assisted in refining the content of the CLAMES framework.

The health states

Since it was infeasible to directly measure preferences for all possible health states generated by CLAMES, a subset of 238 health states was taken into the field in order to obtain data for building a scoring function. Twelve of these states were "marker states"[9] to be tested by all participants. These states were chosen to span the intermediate range of morbidity between full health and death. An additional 189 states consisted of health states associated with actual diseases, as well as some hypothetical health states created to ensure that all levels of all attributes appeared at least once. These health states permitted an econometric (or statistical) approach to developing a scoring function for CLAMES.[13,14] Another 37 states in which all attributes were at the best level—except one attribute, which was set at its worst level (forming a "corner state") or an intermediate level (forming a "pure state")—allowed for the use of a decomposed approach to modeling the observed preference scores.[9,13]

Laminated cards (see Table 2) were used to present the classification of functional limitations for each of the 238 health states to raters. The health states were identified by a randomly allocated two-letter code, rather than disease labels, in order to reduce the influence of participants' idiosyncratic experience with or knowledge of the diseases on the preference measurement exercises. Further, to minimize the cognitive load imposed on participants,[15] the cards did not always explicitly present all 11 attributes. The cards always contained six core attributes (i.e., Pain or Discomfort, Physical Functioning, Emotional State, Fatigue, Memory and Thinking, and Social Relationships) that were expected to be most commonly affected by the various health states under study. For these attributes, a blank space beside the attribute name denoted no limitations on that attribute. For the remaining five supplementary attributes (i.e., Anxiety, Speech, Hearing, Vision, and Use of Hands and Fingers), an attribute was included on the card only if it was affected by the health state. Participants were instructed that the absence of information about limitations meant there were no limitations; they were provided with reference booklets on CLAMES that contained all the attributes.

Participants

Lay panels consisting of 8 to 11 participants each were assembled for the preference measurement exercises. Recruitment was carried out through market research agencies in the following nine Canadian communities: Vancouver, Edmonton, Saskatoon, Toronto, Ottawa, Montréal, Québec, Moncton and Halifax. Participants were selected using a combination of pre-existing research databases, random digit dialling and advertising in local newspapers. In all, 146 individuals participated in 14 panels nationwide.

Screening questionnaires and quota sampling were used to help ensure that each group included a mixture of socio-demographic and other characteristics (i.e., age, sex, education, income, marital and immigrant status, rural versus urban dwellers, and activity limitations). The market research agencies also worked with contacts in other organizations (e.g., student, senior and immigrant associations) in order to help fill the quotas. Some of the study activities were carried out on weekends in order to facilitate representation of the working population. (More information on the recruiting strategies is available from the authors upon request).

TABLE 2
Sample health state cards

Health state: UF	
You have problems with the following:	
Pain or discomfort	Moderate pain or discomfort
Physical functioning	Severe limitations in physical functioning
Emotional state	Very unhappy
Fatigue	Most of the time feel tired and have little energy
Memory and thinking	Very forgetful and have great difficulty when trying to think or solve day-to-day problems
Social relationships	Severe limitations in the capacity to sustain social relationships
Anxiety	Mild levels of anxiety experienced occasionally
Speech	Unable to be understood when speaking to other people
Vision	Unable to see well enough, even with glasses or contact lenses, to read ordinary newsprint but can see well enough to recognize a friend on the other side of the street
Use of hands and fingers	Limitations in the use of hands and fingers, require the help of another person for some tasks
Health state: ML	
You have problems with the following:	
Pain or discomfort	Moderate pain or discomfort
Physical functioning	Mild limitations in physical functioning
Emotional state	
Fatigue	Sometimes feel tired and have little energy
Memory and thinking	
Social relationships	
Anxiety	Mild levels of anxiety experienced occasionally

The preference measurement exercises were conducted from January to June of 2003. Four of the sessions were conducted in French (two in Quebec, one in Ontario and one in New Brunswick), while the remaining ten were conducted in English. Each session, which lasted approximately six hours, included both group and individual measurement exercises. In order to minimize variance due to facilitator effects, an experienced bilingual facilitator from Statistics Canada's Questionnaire Design Resource Centre led each session using a standardized script; addi-tional support was provided by one of the study team members (SCG or JB).

The preference measurement exercises

After an introduction about the purpose and implications of the research program, the Visual Analogue Scale (VAS)[16] was used as a training exercise. Specifically, the VAS—a thermometer-like instrument marked in single, equal interval units ranging from 0 to 100 (i.e., from the least to the most desirable health state)—was used to rank-order the twelve marker states in terms of desirability. For assigning rankings to the health states, participants were asked to imagine living in those states for the rest of their lives, as well as to think about the impact of the health states on their lives in terms of their current family and work situations, usual activities such as social roles, leisure activities and life-style. Further, they were asked to consider the health care services and social support that were currently available to them. This strategy was aimed at facilitating full consideration of how the health states would affect one's personal circumstances, in order to help ensure completeness of preferences. This exercise, while not directly providing the cardinal measures of utility necessary for scaling the CLAMES instrument, served to familiarize panellists with the health state terminology and classification system used in the study, and the concept of expressing personal preferences regarding health states.[17]

Preferences were then elicited for the twelve marker states in a group exercise using the standard gamble (SG) technique, which is based on expected utility theory.[18-21] In the SG procedure, preferences for a given health state are assessed in terms of participants' willingness to undergo a specific treatment, which has a probability of either restoring them to full health or causing death. A ping-pong approach is used to vary the probability of treatment success (see Appendix for further details). A paper-and-pencil variant of the standard gamble was adapted from protocols developed at McMaster University[19] and the University of York.[20] A member of the McMaster team provided consultation regarding the modified protocols; the protocols were also refined in accordance with the results of earlier qualitative pre-testing.

The SG was first conducted as a group exercise for the 12 marker states. During this exercise, participants were asked to carefully consider how the health states described on the cards would impact their own lives in terms of their current family and work situations, usual activities, social roles and social support.

After assigning a preference score to each marker state, participants were encouraged to present their initial preference scores on individual whiteboards and share within a group discussion the reasons for their choices. Participants were given the opportunity to change their initial preference ratings after the discussion. Consensus was not required; the purpose of the discussion session was to ensure common interpretations and understandings of the health states. In order to provide balance to the discussion and ensure that dominant personalities did not take over the conversation, the facilitator made sure that everyone had equal time and opportunity to talk. Further, participants' seats were moved during breaks in order to help control for any possible undue influences associated with sitting in what might be considered more "powerful" positions (i.e., at the end of the table). In order to assess the effect of the discussion sessions, paired sample *t*-tests were conducted on the pre- and post-discussion mean preference scores for the 12 marker states.

Following the SG group exercise, the preference scores for the other health states were elicited in two individual exercises, using the same procedures (described in the Appendix). For the first individual exercise, each participant was assigned a series of 10 additional health states randomly generated from a pool of 193 states (the 189 health states noted previously plus four marker states from the group exercises). For the second individual exercise, participants were randomly assigned a series of four health states from the pool of 37 corner and pure states. The number of preference ratings obtained for each health state in the individual exercises ranged from 6 to 20.

Data cleaning: Inconsistency checks

The data for participants having higher than expected numbers of inconsistent responses were removed prior to analysis. Ten pairs of health states having an obvious severity ordering were identified,[22] and participants' scores were examined to identify their rates of inconsistency, defined as the proportion of pairs for which they rated the less severe health state as more severe. A natural cutoff point was established based on the frequency distribution of the inconsistencies (i.e., the point at which a sharp drop in the number of participants occurred). Having a total number of inconsistencies above this point was considered to be a sign of more serious misunderstanding or misinterpretation of the preference elicitation exercise, and all responses for these participants were removed from the analysis. A total number of inconsistencies at or below this point was viewed as representing a more natural amount of measurement error.

Test-retest exercise

One panel (N = 10) was reconvened to repeat the marker state portion of the preference measurement exercise one month later, in order to assess the test-retest reliability of the measurement protocols. Paired sample *t*-tests of the difference between the mean preference scores on each of the marker states from Time 1 to Time 2 were used to determine

the stability of the estimates between the first and second measurements.

Developing a preference-based scoring function

Mean scores based on directly measured preferences for the 238 health states were used to estimate the parameters of a log-linear scoring function that would transform scores on the 11 CLAMES attributes into a single score.

As a preliminary step, a linear regression model was used to estimate mean preference values for each level of each CLAMES attribute, in order to verify that the ordering of the values was consistent with the severity of the attribute levels. In this analysis, the mean preference scores for the 238 health states were regressed onto 37 dummy independent variables, each corresponding to an attribute set at a specific, less-than-best level. For this analysis, each health state was weighted in accordance with the number of preference ratings it received. The weighting was applied to reflect the fact that preferences for some states were measured with better precision than others, due to their being rated by a larger number of participants.

Next, the following log-linear function was used to estimate the parameters:

$$\ln(p) = \sum_{i=1}^{11} \sum_{j=1}^{5} I_{ij} x_{ij} \quad [1]$$

or

$$p = \prod_{i=1}^{11} y_i$$

where p represents a health state preference score, I_{ij} is an indicator that takes a value of 1 if attribute i is at level j (0 otherwise), x_{ij} represents the parameter or utility weight associated with a specific level of a given attribute and y_i is the appropriate parameter estimate obtained via regression analysis. The multiplicative form of this model assumes that the contribution of a specific level on a given attribute to the overall preference for a health state is relative to one's standing on the other attributes, as opposed to being absolute. (Other functional forms were tested, such as a decomposed model and an additive statistical model with interaction terms, but these are beyond the scope of this paper. Additional information is available from the authors upon request.)

One further adjustment was made because the preference scores for health states ranged from 0 (death) to 1 (full health), and by construction a log-linear model has an asymptote that prevents having a score of 0. Values were "stretched" downwards towards 0 using a scaling parameter, in this case, the lowest possible value estimated by the function or the preference score for the health state where each attribute is set at its worst level of severity. Formally, one would calculate an adjusted or rescaled preference score as follows:

$$p_{adj} = \frac{\prod_{i=1}^{11} y_i - \lambda}{1 - \lambda} \quad [2]$$

where λ is the scaling parameter.

The function was evaluated in terms of its ability to reproduce the mean health state preference scores, using the following global indices of goodness-of-fit: Mean Error (ME), Mean Squared Error (MSE) and the Weighted Mean Squared Error (WMSE).

Results

Participants: A socio-demographic profile

The socio-demographic profile of the panel participants, alongside that of the Canadian population in 2003 (based on Cycle 2.1 of the 2003 Canadian Community Health Survey; CCHS),[23] is shown in Table 3. Most of the participants (65%) were under 50 years of age and there were more women than men. Between one fifth and one quarter of participants had an activity limitation, lived in a rural area or had immigrated to Canada. Each panel had at least one participant from a rural area; one panel included only rural dwellers. For the most part, the socio-demographic profile of the sample was reasonably similar to that for the Canadian population in 2003. The sample had somewhat lower income, was younger and had higher education when compared with the general population.

Inconsistency rates

Table 4 shows the distribution of inconsistent responses for the ten pairs of health states having an obvious ranking in terms of severity. A natural cut-point was observed between those with four and five inconsistent responses (i.e., 60% or more responses deemed inconsistent), and resulted in the removal of the responses of seven individuals (i.e., 5% of the total sample); all subsequent analyses were therefore based on data obtained from 139 participants.

Descriptive statistics: Health state preference scores

Table 5 summarizes the results of the standard gamble group exercises, based on responses from 139 raters across all nationwide focus groups. Pre-discussion mean scores ranged from 0.98 (YD) to 0.29 (UF). The highest scores (associated with the least severe functional limitations) imply that participants would be willing to risk very little to avoid these health states. The standard errors of the pre-discussion mean scores for the marker states were quite small (≤ 0.02).

For each marker state, some of the preference scores were revised following the discussion. The number of changes ranged from 8 (NW) to 50 (BZ); a larger number of changes tended to be associated with the states showing more severe functional limitations. For five of the more severe marker states, the *t*-tests indicated a statistically significant, though small, impact of the discussion; the post-discussion mean preference scores were lower than the pre-discussion mean preference scores. The standard errors of the mean preference scores did not change as a result of the discussion sessions.

Figure 1 displays the mean preference scores for the remaining 226 health states plotted against their standard errors. Since these states received fewer preference ratings than the marker states (i.e., 6 to 20 versus 139), the standard errors are

TABLE 3
Demographic characteristics of participants and Canadian population (2003)

	Participants (%)	Canadian population* (%)
Age		
18-29	22	19
30-39	24	17
40-49	19	19
50-59	16	15
60-69	13	10
70 and above	6	10
Sex		
Male	45	49
Female	55	51
Income		
< $20,000	19	11
$20,000 - $39,999	29	21
$40,000 - $49,999	12	10
$50,000 - $59,999	14	10
$60,000 - $79,999	14	17
$80,000 +	13	31
Education		
Some high school or graduation	31	45
Some college or diploma	28	35
Some university or degree	34	15
Post university	6	5
Activity limitation	21	18
Rural resident	24	19
Immigrant to Canada	20	21

Note: Numbers may not add up due to rounding

*Source: Canadian Community Health Survey, Cycle 2.1, 2003

TABLE 4
Distribution of inconsistent responses

Number of errors*	Number of individuals
0	3
1	13
2	53
3	41
4	29
5	*4*
6	*1*
7	*1*
8	*0*
9	*1*
10	*0*

*Higher preference score given to the health state with the logically lower score

TABLE 5
Results of group exercise for twelve marker states

Marker state	Classification		Initial mean and standard error		Post-discussion score change		Final mean	*t*-statistic	*p*-value
					#	%			
YD	211111	11111	0.98	0.00	9	6.5	0.98	-0.91	0.36
NW	211211	21111	0.96	0.00	8	5.8	0.97	-1.56	0.12
ML	321211	21111	0.93	0.01	14	10.1	0.93	-1.80	0.18
GM	123222	21111	0.88	0.01	10	7.2	0.88	-1.34	0.18
IG	123223	31111	0.86	0.01	19	13.7	0.85	1.26	0.21
MV	332213	31111	0.85	0.01	10	7.2	0.85	-0.31	0.77
EK	333423	31111	0.73	0.01	20	14.4	0.72	1.88	0.06
FO	131254	21111	0.72	0.02	48	34.5	0.67	5.70	< .0001
VV	334323	31111	0.59	0.02	22	15.8	0.58	2.54	0.012
BZ	441314	31111	0.46	0.02	50	36.0	0.42	4.27	< .0001
NN	444444	31111	0.33	0.02	31	22.3	0.29	4.23	< .0001
UF	344354	24134	0.29	0.02	25	18.0	0.26	3.61	0.004

Note: Health states were presented with random alphabetic codes.

generally higher (≤ 0.18) than those obtained for the marker states.

Test-retest reliability

Table 6 presents the results of the paired sample *t*-tests comparing the post-discussion mean preference scores between Time 1 and Time 2. Only the mean preference ratings for health states BZ and NN were significantly different between Time 1 and Time 2, at the 0.05 level.

Fitting the log-linear function

Table 7 displays mean preference values for each level of each CLAMES attribute, as derived from the linear regression. Some adjustment was required to accommodate the ordering of these values—a collapse of levels 1 and 2 for Emotion; levels 1 and 2 for Fatigue; levels 3 and 4 for Fatigue; levels 2, 3 and 4 for Memory and Thinking; levels 1 and 2 for Speech; and levels 3 and 4 for Speech—before estimating the parameters of the log-linear scoring function (see Equations 1 and 2). In addition, preliminary estimation of parameters yielded values greater than 1 for level 2 of the Social Relationships and Vision attributes. Therefore, these parameters were fixed at 1 and the model was re-estimated. The function provided a good overall fit to the mean preference scores (ME = -0.005; MSE = 0.005; WMSE = 0.002).

The scaling parameter λ, corresponding to the state in which each attribute is set at its worst level, is 0.115. With the scaling parameter applied, global model fit decreased slightly (ME = 0.024; MSE = 0.008; WMSE = 0.005). The final set of parameter estimates obtained for all attribute levels, as well as a practical and user-friendly version of the log-linear function, are displayed in Table 8.

Table 9 displays the directly measured mean preference scores for the 12 marker states, based on preferences elicited in both the group (final, post-discussion scores) and individual exercises, alongside those produced by the scoring function with the scaling parameter λ applied. For health states with a directly measured preference score above 0.8, the function fits the data very well. For health states with lower preference scores, the function tends to underestimate the preference score, due to the nature of the scaling adjustment.

FIGURE 1
Mean preference scores plotted against standard errors for 226 health states

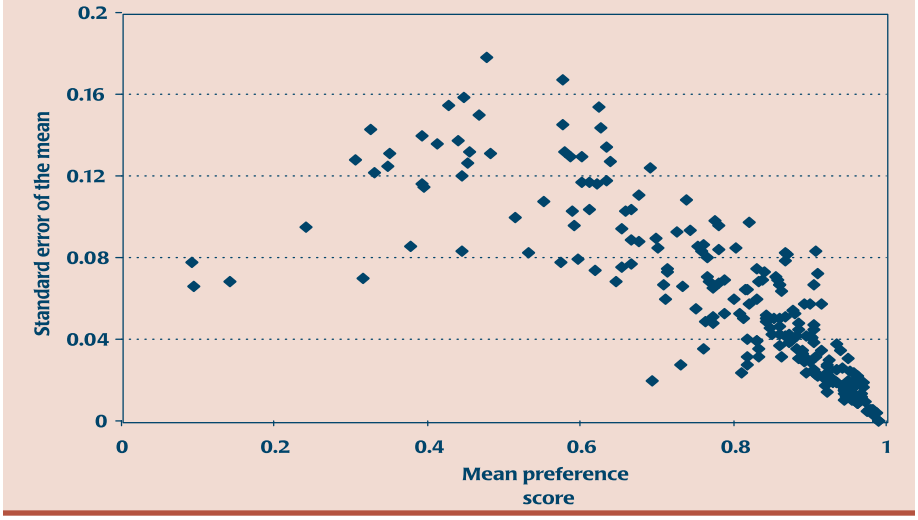

TABLE 6
Mean scores for test and re-test (paired samples *t*-test)

	Mean			
Health state	Time 1	Time 2	*t*-value	*p*-value*
YD	0.988	0.990	1.000	0.343
NW	0.983	0.974	-1.000	0.343
ML	0.960	0.965	0.434	0.675
GM	0.958	0.950	-0.550	0.596
IG	0.915	0.924	0.546	0.599
MV	0.889	0.899	0.294	0.775
RD	0.877	0.930	1.301	0.225
EK	0.790	0.774	-0.509	0.623
FO	0.764	0.788	0.654	0.529
VV	0.685	0.750	1.073	0.311
BZ	0.495	0.595	2.491	0.034
UF	0.431	0.575	1.942	0.084
NN	0.287	0.428	2.303	0.047

Note: These are the marker states and another health state, RD, which was considered by the group as an example.

*Two-tailed

Discussion

Preference scores for a subset of 238 health states classified by CLAMES were elicited from panels of lay Canadians using the standard gamble in both group and individual exercises. A log-linear function provided a good fit to the observed mean preference scores and can compute a preference score for any health state possible within the CLAMES framework.

Strengths of the protocols

The integrity of the measurement protocols, as evidenced by stability in health state preferences over time, could be due to several methodological strengths. First, the

TABLE 7
Adjusted mean for each attribute level*

Attribute	Level 1	2	3	4	5
Pain and discomfort	1.00	0.98	0.97	0.77	n/a
Physical functioning	1.00	0.97	0.93	0.83	n/a
Emotion	1.00	1.03	0.96	0.85	0.79
Fatigue	1.00	1.00	0.95	0.96	n/a
Memory and thinking	1.00	0.98	0.93	0.99	0.85
Social relationships	1.00	0.98	0.95	0.90	0.86
Anxiety	1.00	0.99	0.97	0.90	n/a
Speech	1.00	1.00	0.98	0.99	n/a
Hearing	1.00	0.95	0.93	0.88	n/a
Vision	1.00	0.97	0.99	0.92	n/a
Use of hands and fingers	1.00	0.97	1.00	0.94	0.90

Notes:

*The reference group for the adjustment is level 1 for all other attributes.

n/a - There is no level 5 on this attribute.

measurement protocols, which included introductory training exercises using the VAS, were developed on the basis of well-established methods[19,20] as well as expert consultation, and applied in a standardized manner across all participants and groups.

Second, the standard gamble (SG) method, considered by some experts as the "gold standard" preference measurement technique,[9] is the only preference elicitation method that produces true "utilities" (i.e., preferences measured under uncertain conditions) in accordance with von Neumann-Morgenstern expected utility theory.[21] Since the SG involves risk, it is regarded as highly appropriate in the context of health care decision making.[17] In addition, qualitative pre-testing indicated that participants preferred the standard gamble to the Time Trade-Off (TTO) technique because it was easier to understand and some participants considered the Person Trade-Off (PTO) to be ethically objectionable; one focus group member refused to do the PTO. (For further descrip-tion of these techniques, see Dolan et al.[22])

Third, a trained and experienced facilitator from Statistics Canada's Questionnaire Design Resource Centre was involved from the early stages of developing the protocols and conducted all sessions, both English and French, in order to eliminate variance due to facilitator effects.

Fourth, the health states were identified with randomly allocated two-letter codes rather than the name of the disease they represented (e.g., ML represented type II diabetes). This strategy may have minimized bias due to misunderstanding about particular diseases: Other studies have reported that different preference scores were obtained for the same disease when presented with and without labels.[24] The

TABLE 9
Observed and function-generated preference scores for twelve marker states

Marker state	Classification		Observed	Function
YD	211111	11111	0.98	0.98
NW	211211	21111	0.97	0.96
ML	321211	21111	0.93	0.91
GM	123222	21111	0.88	0.86
MV	332213	31111	0.85	0.85
IG	123223	31111	0.85	0.81
EK	333423	31111	0.72	0.70
FO	131254	21111	0.67	0.63
VV	334323	31111	0.58	0.52
BZ	441314	31111	0.42	0.33
NN	444444	31111	0.29	0.20
UF	344354	24134	0.26	0.17

Health states are presented in descending order of observed scores.

TABLE 8
Parameter estimates for log-linear model

Attribute level	Pain and discomfort y_1	Physical functioning y_2	Emotional state y_3	Fatigue y_4	Memory and thinking y_5	Social relationships y_6	Anxiety y_7	Speech y_8	Hearing y_9	Vision y_{10}	Use of hands and fingers y_{11}
1	1	1	1	1	1	1	1	1	1	1	1
2	0.98	0.983	1	1	0.985	1	0.985	1	0.958	1	0.985
3	0.954	0.949	0.919	0.952	0.985	0.955	0.982	0.956	0.938	0.93	0.985
4	0.704	0.681	0.719	0.952	0.985	0.915	0.833	0.956	0.897	0.884	0.985
5	n/a	n/a	0.663	n/a	0.784	0.821	n/a	n/a	n/a	n/a	0.784

Notes:

A preference score for any health state classified by CLAMES can be calculated using the following simplified functional form:
Padj = [(y1 * y2 * y3 * y4 * y5 * y6 * y7 * y8 * y9 * y10 * y11) − 0.115]/0.885, where y is the appropriate parameter estimate from Table 8.

n/a - There is no level 5 on this attribute.

removal of labels also avoided the presentation of unrealistic scenarios to participants (e.g., experiencing influenza or a heart attack for the rest of their lives).

Contribution to preference measurement

The present work demonstrates the potential for measuring health state preferences in small groups using facilitator-led and self-completion methods. Paper-based, self-completion approaches to the standard gamble used elsewhere have performed reasonably well; one study showed that health state preference scores derived from paper-based, self-completion methods were very highly correlated (R^2 = 0.88) with those obtained using a more sophisticated, computer-based version of the SG (which was similar to the interviewer-based approach).[25] However, to our knowledge, paper-based, self-completion SG techniques have not been implemented in group settings prior to the current study. Although the reliability of individual preferences have been found to be moderate to low over time, the reliability of group preferences has tended to be higher.[26]

Significant discrepancies between some of the pre- and post-discussion mean preference scores for the marker states suggest that preferences for some health states were developed further via discussion. The more severe health states were more likely to change after discussion, possibly because members of the general public are not likely to experience these health states. In line with the current findings, both Fischhoff[27] and Feeny[17] suggest that in the domain of health, people develop their preferences through a deliberative process, although a small study by Stein et al.[28] did not support this empirically.

The discussion was considered necessary in the present study because there were 11 attributes to consider and only six core attributes affected by a health state were shown on the laminated card, unless there was a limitation on other attributes. However, the mean preferences for severe health states were actually lower after discussion: The expected focusing effect where raters zero in on affected attributes and disregard those at normal functional levels was not observed. Further, it does not appear as though there was excessive bias brought about by the group discussion sessions, given that the standard errors did not change as a result of the group discussion sessions.

The scoring function

Log-linear models such as the one estimated here have performed well with preference data obtained using other instruments, most notably the EQ-5D in the Australian Burden of Disease study.[29] Although we tested other models for CLAMES, such as an additive statistical model and a multiplicative "decomposed" model,[9] they did not yield as good a fit to our standard gamble data as the log-linear model.

Limitations

Although the results of the current study are encouraging in a number of respects, some limitations should be noted. First, our panels were not fully representative of the Canadian population, though efforts were made to ensure that the sample was heterogeneous as to socio-demographic and health characteristics, so that the preferences would reflect a variety of personal and contextual factors.

Second, the levels of some attributes had to be combined before estimating the preference-based scoring functions, since the corresponding weights were not ordered as theoretically expected. It is possible that the small sample size provided an insufficient number of preference ratings to obtain clear empirical differentiation between attribute levels close in terms of actual impact on functional status. Third, it should be noted that CLAMES contains a larger number of attributes (i.e., 11) than are typically used on preference-based health status tools. The "magical number seven (plus or minus two)"[15] as a limit to the number of items individuals can process at once has been used to justify the limit of nine attributes for other multi-attribute classification systems.[17] However, we chose to provide more detailed information, as we felt that it was a justifiable trade-off because we did not use disease labels and therefore participants required more complete information on functional health to understand the health state.

As for the function, the scaling parameter may have introduced some downward bias when computing preferences for health states having low mean preference scores. This was intended to counteract the upward bias introduced by the inability of our preference measurement exercise to produce negative scores for states that may have been perceived as "worse than death."

Fourth, although the group SG exercise appeared to work well, we did not directly compare our results to the traditional, professional interviewer-administered one-on-one preference elicitation survey.[19] Efforts were made to preserve the integrity of the original method here (e.g., a member of the McMaster team reviewed the protocols). However, a specific objective of future research might be to examine the degree of convergence in preferences obtained from the group method and the traditional one-on-one approach.

Finally, due to constraints on resources, the current study did not attempt to replicate or validate the results of the log-linear modeling of the standard gamble scores with additional, directly measured preference data. Further work to assess both intra-survey and out-of-sample predictive validity of the function would thus strengthen this work.[9,30]

Contribution to policy decisions

The preference-based scoring function presented here allows for the convenient calculation of preference scores for any of the 10,240,000 health states possible within the CLAMES framework, providing wide coverage of health states that might be encountered in research and clinical practice. Within the PHI research program, the preference scores will contribute a comparable measure of severity of functional limitations across health states, which will serve as an important component of summary measures incorporating morbidity and mortality from specific diseases. The preference scores used in the construction of the function were obtained

from laypersons in the Canadian societal context, which is particularly desirable given the cultural and economic influences on health.[31] The preferences of the general public are appropriate for health state preferences that contribute to policy and priority setting in the health care sector.[6-8] As Dolan[32] notes, "we are all potential patients." The use of average preferences is conducive to fairness in health care decision making, since the scores can reflect the input of multiple perspectives (i.e., within a heterogeneous sample, as used here) but at the same time are not unduly biased in favour of particular sub-groups.

Conclusion

We obtained health state preference scores in a group setting using the SG technique. Provided that training of participants and standardized measurement procedures are in place, these methods appear to provide a viable and economical means of carrying out preference measurement. These observed preferences were used to build a preference-based scoring function for CLAMES, which was subsequently used to quantify health-related quality of life for numerous health states within the context of the Population Health Impact of Disease in Canada research program. Two related articles describe, respectively, how CLAMES was used to develop preferences scores for health states related to cancer[33] and how these preference scores were used in the calculation of health-adjusted life years (HALYs) lost to cancer in Canada in 2001.[34] Future work will use CLAMES to examine the impact on health-related quality of life of other diseases and health conditions.

Acknowledgments

This work is part of the Population Health Impact of Disease in Canada (PHI) research program, a collaboration of Statistics Canada, the Public Health Agency of Canada and researchers from McGill University, the University of Ottawa, the University of Manitoba, the Institute for Clinical Evaluative Sciences (ICES) and the Montérégie Regional Board of Health and Social Services. The PHI is funded by Statistics Canada and the Public Health Agency of Canada.

The authors acknowledge Benoit Allard for his role in developing and implementing the protocols; William Furlong for reviewing and providing constructive comments on the protocols; Stacey Todd and Bill Flanagan for assistance in preparing materials for the preference exercises; and Kathy White for significant contributions in preparing this manuscript.

APPENDIX

Paper-and-pencil version of standard gamble and description of search procedure

The standard gamble (SG) presents raters with a hypothetical scenario consisting of two alternatives. The first alternative is a lottery in which a treatment, sometimes referred to as a "magic pill", has a probability *p* of restoring participants to full health for the remainder of their lives and a corresponding probability 1 - *p* of killing them instantly. The second alternative is to simply remain with certainty in some other intermediate (i.e., less than full) health state under investigation (e.g., Health State X) for the rest of their lives. The probability *p* of succeeding at the lottery is systematically varied until the participant reaches his or her indifference point, that is, the point at which he or she cannot decide whether to gamble to escape Health State X or to stay in it for life. If values of 0 and 1 are assigned to immediate death and full health, respectively, then in accordance with the axioms of expected utility theory, the participant's preference for Health State X is simply *p* at the indifference point.

The standard gamble response sheet used in this study is displayed in Figure A. To minimize measurement bias resulting from reference or framing effects, an iterative, "ping-pong" approach was used to locate the indifference point. Specifically, starting at a 100% probability of gaining full health via the lottery, participants were instructed to ping-pong through the chances of gaining full health, moving toward the indifference point from both extremes of the continuum simultaneously (e.g., 100%, 0% … 2%, 98 %, etc.), until they rejected Choice A at probability *p* but accepted Choice A at probability *p* plus one unit; in other words, the point at which their answer changed from Choice A to Choice B. The response sheet has 2% intervals at the top and bottom to obtain finer elicitation for very high and very low utility values; the remaining intervals are each 5%.

FIGURE A
Standard gamble response sheet

Health State: _____

Choice A		Choice B
Chances of Full Health (%)	Chances of Immediate Death (%)	Health State on Card (%)
100	0	100
98	2	100
95	5	100
90	10	100
85	15	100
80	20	100
75	25	100
70	30	100
65	35	100
60	40	100
55	45	100
50	50	100
45	55	100
40	60	100
35	65	100
30	70	100
25	75	100
20	80	100
15	85	100
10	90	100
5	95	100
2	98	100
0	100	100

INITIAL ANSWER: _____

FINAL ANSWER: _____

For example, a participant might choose the lottery (Choice A) at a 75% chance of gaining full health, yet elect to remain in the health state under study (Choice B) once the chances of gaining full health were reduced to 70%. Since the maximum indifference interval is the distance between these two values, the midpoint (i.e., 72.5%) is taken as a proxy for the true indifference point or utility for the health state of interest. All indifference points were calculated by the facilitator, who led the participants through the procedure for each marker state.

References

1. Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. Med Care 2000;38(9suppl):II138-II150.

2. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. J Health Serv Res Policy 1999;4(3):174-184.

3. Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002.

4. Murray CJL, Lopez AD, editors. The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Vol 1. Boston, MA: Harvard University Press; 1996.

5. Sommerfeld J, Baltussen RMPM, Metz L, Sanon M, Sauerborn R. Determinants of variance in health state valuations. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 549-579.

6. Gold ME, Russell LB, Siegal JE, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

7. Edgar A, Salek S, Shickle D, Cohen D. The ethical QALY: Ethical issues in health care resource allocations. Euromed Communications; 1998.

8. Hadorn DC. The role of public values in setting health care priorities. Soc Sci Med 1991;32:773-781.

9. Feeny D, Furlong W, Torrance GW et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. Med Care 2002; 40:113-128.

10. Ware JE, Jr. SF-36 health survey update. Spine 2000;25(24):3130-3139.

11. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQoL group. Ann Med 2001;33:337-343.

12. Krabbe PF, Stouthard MEA, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. J Clin Epidemiol 1999;52(4):293-301.

13. Dolan P. Modelling the relation between the description and valuation of health states In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 502-513.

14. Brazier J, Rice N, Roberts J. Modelling health state valuation data. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization, 2002. p. 529-547.

15. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psych Rev 1956;63(2):81-97.

16. Torrance GW, Feeny DH, Furlong W. Visual analog scales: Do they have a role in the measurement of preferences for health states? Med Decis Making 2001; 21:329-334.

17. Feeny DH. The utility approach to assessing population health. In Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 515-528.

18. Gafni A. The standard gamble method: What is being measured and how it is interpreted. Health Serv Res 1994; 29(2):207-224.

19. Furlong W, Feeny D, Torrance, GW, Barr R, Horsman J. Guide to design and development of health-state utility instrumentation. Hamilton, Canada: Centre for Health Economics and Policy Analysis, McMaster University; 1990.

20. Measurement and Valuation of Health Group, Gudex C, editor. Standard gamble user manual: Props and self-completing methods. York, England: Centre for Health Economics, University of York; 1994.

21. von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton, NJ: Princeton University Press; 1944.

22. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: A comparison of methods. J Health Econ 1996;15:209-231.

23. Statistics Canada. Canadian Community Health Survey (CCHS), Cycle 2.1. Available at ww.statcan.ca/cgi-bin/imdb/p2SV.pl? Function = getSurvey&SDDS = 3226&lang = en&db = IMDB&dbg = f&adm = 8&dis = 2/.

24. Mahapatra P, Nanda L, Rajshree KT. The 6D5L description system for health state valuation. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 349-367.

25. Ross PL, Littenberg B, Fearn P, Scardino PT, Karakiewicz PI, Kattan MW. Paper standard gamble: A paper-based measure of standard gamble utility for current health. Int J Technol Assess in Health Care 2003;19(1):135-147.

26. Rutten-van Mölken MPM, Bakker CH, van Doorslaer EKA, van der Linden S. Methodological issues of patient utility assessment - experience from two clinical trials. Med Care 1995;33:922-937.

27. Fischhoff B. Value elicitation: is there anything in there? American Psychologist 1991;46:835-847.

28. Stein K, Ratcliffe J, Round A, Milne R, Brazier JE. Impact of discussion on preferences elicited in a group setting. Health and Quality of Life Outcomes 2006;4:22.

29. Mathers C, Vos T, Stevenson C. The burden of disease and injury in Australia. Australian Institute of Health and Welfare (cat no. PHE17). Canberra: AIHW; 1999.

30. Shaw, JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. Med Care 2005;43(3):203-220.

31. Sadana R. Measurement of variance in health state valuations in Phnom Penh, Cambodia. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: Concepts, ethics, measurement and applications. Geneva: World Health Organization; 2002. p. 593-618.

32. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35(11):1095-1108.

33. Public Health Agency of Canada Web site. Preliminary results for cancers. Available at: www.phac-aspc.gc.ca/phi-isp/cancers.html

34. Boswell-Purdy J, Flanagan WM, Roberge H, Le Petit C, White KJ, Berthelot J-M. Population health impact of cancer in Canada, 2001. Chron Dis Can 2007; 28(1-2):42-55.