Working Paper No.9
17 May 2004

ENGLISH ONLY

STATISTICAL COMMISSION and UN ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

WORLD HEALTH ORGANIZATION (WHO)

Joint UNECE/WHO/Eurostat Meeting
on the Measurement of Health Status
(Geneva, 24-26 May 2004)

Session 3– Invited paper

WORK DONE TO ACHIEVE INTERNATIONAL COMPARABILITY: RECENT CANADIAN EXPERIENCE.

Submitted by Statistics Canada[*]

A Joint U.S./Canada Health Survey: Improving the International Comparability of Health Survey Statistics[**]

I. OVERVIEW

1. The Joint Canada/U.S. Health Survey (JCUSH) is a collaborative effort of Statistics Canada and the U.S. National Center for Health Statistics to conduct a telephone survey in both countries, using the same questionnaire. About 3,000 interviews of adults will be conducted in Canada and 3,000-5,000 will be conducted in the U.S. The questionnaire will cover chronic health conditions, functional status, smoking, height and weight, cancer screening, dental visits, and demographics. Staff members from the two organizations are already working closely together on the survey design and questionnaire testing.

[*] Paper presented by Jean-Marie Berthelot. This document contains 3 brief articles prepared by individuals working on the Joint Canada/US Survey of Health (JCUSH) and on the Classification and measurement of functional health (CLAMES) projects. They are provided for information only.

[**] Prepared by the survey team

2. JCUSH is expected to produce a body of comparable data for Canada and the U.S. that researchers can use to study Canada-U.S. differences in health status and health care. Because the two countries are similar in many ways, but different with respect to the structure of their systems for providing and financing health care, previous research comparing similar survey data from the two countries has provided many useful policy insights. The fully comparable data from JCUSH will enhance such research.

3. JCUSH is testing the ability of two national statistical offices to integrate their survey design and analysis activities at all staff levels to achieve a greater degree of international comparability in health statistics than has ever before been achieved. If successful, the JCUSH will be a model for future collaborations between national statistical offices, not only in Canada and the U.S., but in other countries as well. This will widen the international comparability of health data that is so much needed.

II. THE PROBLEM AND ITS IMPORTANCE.

4. The problem of the lack of international comparability in statistics on health and functioning has been noted by several international organizations, including the United Nations, the World Health Organization, the Organization for Economic Cooperation and Development, and the European Commission. Comparable statistics are required to assess and compare the performance of national health systems. Such assessments and comparisons will lead to improvements in those systems and better health care for the populations they serve.

5. There are two basic approaches to "harmonizing" international health statistics: "pre-harmonization," and "post-harmonization." (Harmonization is the preferred term in international discussions of data comparability.) Pre-harmonization refers to efforts to standardize data in different countries before they are collected, and includes using the same questions and the same field work procedures. Post-harmonization refers to efforts to standardize data after they have been collected, and includes coding data with the same coding scheme and applying item response theory to identify common underlying scales.

6. JCUSH is an example of pre-harmonization, but it goes further than previous efforts. Whereas previous efforts have achieved some success in standardizing data collection instruments used by survey organizations in different countries JCUSH will standardize all components of survey design and analysis in the two countries, including questionnaires, questionnaire testing, data collection, data processing, and data analysis. That is because the survey is being planned and conducted jointly by staff from Statistics Canada and the National Center for Health Statistics.

7. The probability of success of the JCUSH is enhanced by the relationship between the two countries: they share an open border, an official language, a high standard of living, and a cultural heritage. Success is not guaranteed, however, because there are differences between the two nations' statistical and health systems that create obstacles and because of funding uncertainties. If those obstacles can be overcome, JCUSH may form the basis for a model of international collaboration in the collection of health data, contributing to the harmonization of health statistics and the comparison of national health systems

III. PROJECT DESCRIPTION

8. The project will be conducted in three phases, as follows:

Phase 1, Questionnaire development, cognitive testing, and sample design. Questionnaire design staff from Statistics Canada (STC) and the National Center for Health Statistics (NCHS) will meet in person several times and by telephone at least monthly, and will communicate by e-mail to design a single telephone questionnaire for the survey (to be administered using Computer Assisted Telephone Interviewing (CATI)), to cognitively test the questionnaire, and to conduct small scale field tests in Canada and the U.S.

9. Survey methodologists from the two agencies will consult by telephone to design a sample that meets the needs of both nations. Sample sizes of 3,000 completed cases are planned for each country, for a total of 6,000 cases. If sufficient funding is available, the U.S. sample will be expanded to 5,000 to permit more reliable and precise estimates, including estimates for the African American and Hispanic populations.

10. Phase 2, CATI instrument development, data collection, and post-processing. The CATI program will be written by STC and tested jointly by STC and NCHS. The questionnaire will be translated and administered in three languages, English, French, and Spanish. Data collection in both countries will be conducted by the STC telephone survey center, which will also do initial editing and produce raw data files. STC and NCHS will jointly plan and program additional data cleaning procedures and file structures, and will jointly develop and produce documentation for the final data files.

11. Phase 3, Analysis and data release. Analysts from STC and NCHS will meet in person and by telephone, and communicate by e-mail to plan a series of project reports. They will then jointly analyze the data and co-author reports for publication. Public use versions of the final data files will be made available for downloading from each agency's Web site.

12. Accomplishments to date: The survey has been completed and it will be released on June 2, 2004, simultaneously by Statistics Canada and by the National Centre for Health Statistics. The objective of the dissemination report is to provide a first look at the results from the JCUSH survey. The findings focus primarily on the overall similarities and differences between our two countries in a manner not possible before. The report will also provide an overview regarding the methods and processes used to conduct the survey.

The Classification and Measurement System of Functional Health: A New Approach to Health Status Measurement[***]

I. INTRODUCTION

13. Although there has been considerable growth in quality of life research and health status measurement in recent years, a lack of consensus remains on the most appropriate way to classify the health status of the general population. Identifying the main dimensions of health is a first step in developing appropriate generic measures to classify health. Such measures can then be used to measure and monitor population health, and to permit comparisons among different diseases or the experience of illness in different socio-demographic groups. This paper introduces a new measurement tool designed to describe limitations in functional health that are associated with diseases that occur in Canada.

14. The World Health Organization recently published a series of domains of health that they recommend be used to describe the multi-dimensional nature of this concept [1]. These domains include self care, usual activities, affect, pain, mobility and cognition as well as some additional domains that are either directly or indirectly related to health such as interpersonal relations, breathing or dexterity. Recent work at Statistics Canada [Bernier et al, submitted for publication] compared these domains of health to those that are currently being measured in the Canadian context with the National Population Health Survey[1] (NPHS) and made recommendations for the series of domains to be included in a Canadian classification system of health status. This study identified eighteen distinct dimensions of health across different age and gender subgroups including: psychological well being, stress, functional limitations, chronic conditions, disability days, sensory impairment (or communication), depression/distress and chronic breathing problems, breathing, hearing, energy expenditure, depression, social support, speech, incontinence, vision, emotion and cognition. Together these 18 dimensions are recommended as a core set from which health status should be measured.

15. Statistics Canada has used these 18 domains as a guide to developing a new classification system that can describe the range of health states that are prevalent in the developed world. The Classification and Measurement System of Functional Health (CLAMES), the result of this work, is a classification and description tool that measures health status and health related quality of life associated with disease. It is designed to provide generic health status descriptions for which preference scores can then be elicited.

16. CLAMES was developed because there was a need for a classification system that could cover the spectrum of health-related functioning, including all significant aspects of health, which results from diseases commonly experienced in Canada. An instrument that could embed the concept of health as seen by the population and that could be aggregated into a single index was required. Detailed reviews of existing instruments were conducted to determine if an existing classification would be suitable. These reviews focused mainly on three of the most commonly employed indices for measuring health status including the Health Utilities Index (HUI3) [2], the EuroQol five dimensions index [3] and the Short Form 36 (SF-36) Health Status Questionnaire [4] as these instruments had been tested and validated in past Canadian studies.

[***] Prepared by Sarah Connor Gorber, Julie Bernier and Jean-Marie Berthelot, Statistics Canada.

17. This review concluded that although as a group these instruments covered a broad range of categories of functioning none was on its own sufficient to describe the range of illness and injury to be studied. The HUI lacked attributes to describe social limitations/ functioning associated with health states and the EQ5D had too few attributes with too few levels to make the fine distinction between stages and severity of disease that was required. In addition, factor analysis work using Canadian survey data examined the independence of the dimensions of the EQ5D and found that its attributes were highly inter-correlated [5]. This analysis revealed that the EQ5D consists of only two independent dimensions (one common factor that includes mobility, self-care, pain, and usual activities and one unique factor to describe anxiety and depression). Similar analysis of the 36 questions in the SF-36 measure indicated that only two independent dimensions of health - physical and mental were present.

Table 1 Sources from which CLAMES attributes were adapted

Pain or Discomfort	HUI 3[1]
Physical Functioning	SF-36[2]
Emotional State	HUI 3
Fatigue	SF-36
Memory and Thinking	HUI 3
Social Relationships	SF-36
Anxiety	EQ-5D[3]
Speech	HUI 3
Hearing	HUI 3
Vision	HUI 3
Use of Hands and Fingers	HUI 3

[1] Health Utilities Index- Mark III
[2] Medical Outcomes Study Short Form 36
[3] Euroqol Five Dimensions Index

18. We therefore opted to develop a new classification system using these existing instruments and past work on the key dimensions of health as a guide to creating its content. Dimensions from past work that fit with our conceptual framework (see below), and could be objectively measured without requiring detailed lists of question were considered. We then selected the most appropriate attributes and concepts from each of the three instruments and modified them as required (Table 1). We have also concluded that the system must be made up of two sets of attributes to accurately capture functional status: core attributes to describe the main domains of functioning and supplementary attributes to describe aspects of functioning that are only relevant to specific states of health. This resulted in an 11 attribute classification system (6 core and 5 supplementary attributes).

II. CORE ATTRIBUTES

19. Effort was made to ensure that the six core attributes were structurally and statistically independent, validated and coherent. Pain/discomfort, physical functioning, emotional state, fatigue, memory and thinking and social relationships all make up the core attributes. Each attribute is described as one of 4 or 5 levels (Table 2).

Table 2 Attributes used in CLAMES

Core Attributes	
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. Somewhat forgetful, and have some difficulty when trying to think or solve day-to-day problems 5. Very forgetful, and have great difficulty when trying to think or solve day-to-day problems
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
Supplementary Attributes	
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 newsprint 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

20. Statistical independence was examined through direct (head to head) comparisons of the attributes making up the EQ5-D and those on the HUI3 using data from the NPHS [6]. Similar analysis with the Canadian Community Health Survey[2] was undertaken to compare the attributes in the HUI3 to those in the SF-36 questionnaires [7]. Statistical independence is based on the correlation of attribute levels reported for subjects in the population. Most items that were considered redundant (where correlations were above 0.25) were eliminated unless there was a conceptual reason for them to remain.

21. Structural independence refers to the independence of the constructs being measured. If attributes are structurally independent it signifies that changes to one attribute do not necessarily have effects on the other attributes.

22. Structural and statistical independence were used as guides in selecting attributes and eliminating overlap in the concepts. In not all cases are the attributes completely independent, but this information has been used to assist in determining which attributes should be included and which could be excluded. For instance, each of the instruments had at least one attribute that measured concepts related to physical functioning and mobility yet physical functioning, mobility, self care, ambulation, and usual activities were all highly correlated attributes [6-7]. As a result only one of these dimensions (physical functioning) was retained in CLAMES and now encompasses many of these related concepts. In cases where overlapping attributes did exist in each of the instruments we selected the most appropriate, widely applicable, easily understood source.

23. Most items that were considered redundant (where correlations were above 0.25) were eliminated unless there was a conceptual reason for them to remain. Pain for instance, is a dimension that was statistically correlated with mobility and self care [6] but is considered to be conceptually different as well as structurally independent from these attributes. Emotion and cognition are also attributes that were shown to be statistically correlated [6] yet are viewed as distinct and necessary concepts in gaining a complete understanding of a health state.

III. SUPPLEMENTARY ATTRIBUTES

24. The five supplementary attributes include anxiety, vision, speech, hearing, and dexterity. The majority of supplementary attributes were also obtained from existing, validated sources, however, in some cases modifications were made to the original items. In some instances the language was simplified to ensure comprehension by the general population (which is required if the tool is to be used in population surveys or preference measurement exercises) and in others the number of levels within the attributes were collapsed. The NPHS was used as a guide to determine which levels of the HUI3 to group together (based on prevalence) and to ensure that the integrity and original meaning of the items were preserved. Both the general and institutional component of the NPHS were examined, although the levels were rarely collapsed in the same way in each sample. Nonetheless, it was a useful guide in understanding the prevalence of the impairments in different populations and final decisions were made based on the levels that were thought to be most meaningful to the health states being defined and for the participants in the study.

IV. CONCEPTUAL FRAMEWORK

25. Building on the concept of "capacity" from the International Classification of Functioning (ICF) framework developed by the WHO, each level reflects an individual's intrinsic capacities (what they can do and how they function) within an attribute as opposed to their performance [1]. For instance, the social relationships attribute assesses a person's internal **capacity** for developing and maintaining social relationships rather than measuring their opportunities for social relationships, which could be partially imposed on them by their larger environment. CLAMES has been refined subsequent to qualitative testing and peer review.

26. CLAMES is a generic health state classification system that builds on the strength of existing instruments to provide a standardized and coherent manner in which to describe a series of health states. This tool permits comparable descriptions and classifications of health status covering a wide range of severity levels and symptoms. It has recently been used to provide descriptions of approximately 300 health states that are prevalent in the Canadian population [8]. These health states were subsequently used to elicit preference scores, using the Standard Gamble, from panels of Canadians as part of the Population Health Impact Study (PHI). The PHI is designed to obtain an objective assessment of the relative health impacts of various disease, injury and risk factors on the Canadian Population using Summary Measures of Population Health. Once further validated, CLAMES could also be adapted for use on population surveys or clinical studies designed to measure and monitor health status in terms of functional limitations.

V. END NOTES

[1] The NPHS is a longitudinal survey that collects information about the health of the Canadian population. It began in 1994 and collects data at two year intervals.

[2] The CCHS is a cross-sectional survey that gathers health-related and socio-demographic data at the health region level. It began in 2000 and collects data from approximately 130 000 members of the Canadian population every two years.

VI. REFERENCES

1. Chatterji S, Ustün BL, Sadana R, Salomon JA, Mathers CD, Murray CJL. The conceptual basis for measuring and reporting on health. Global Programme on Evidence for Health Policy Discussion Paper No. 45. World Health Organization, 2002.

2. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. Multi-Attribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. Med. Care 2002; 40 (2): 113-128.

3. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37(1):53-72.

4. Ware JE Jr. *SF-36 Health Survey Manual and Interpretation Guide*. Boston: The Health Institute, New England Medical Centre, 1993.

5. Bernier J, Berthelot JM, Wolfson M. Comparison of Three Generic Health Status Measures. Paper presented to the Joint UN/ECE/ WHO Expert Meeting on Measuring Health Status. October 23-26, 2000. Ottawa, Ontario.

6. Belanger A, Berthelot JM, Guimond E, Houle C. A Head-to-Head Comparison of Two Generic Health Status Measures in the Household Population: McMaster Health Utilities Index (Mark 3) and the EQ-5D. Internal Documentation. Statistics Canada, Ottawa, 2000.

7. Bernier J, Berthelot JM, Wolfson M. Head to Head Comparison of three Generic Health Status Measures in Household Populations. Paper presented to the 22nd Annual Meeting of the Society for Medical Decision Making. September 24-27, 2000. Cincinnati, Ohio.

8. Gorber S. A New Classification and Measurement System of Functional Health. In *au courant*, newsletter of the Health Analysis and Measurement Group. Ottawa: Statistics Canada catalogue 82-005, September 2003:2-3.

Estimating a scoring function for the Classification and Measurement System of Functional Health (CLAMES)[****]

I. INTRODUCTION

27. Over the past century, advances in public health and population health have dramatically increased life expectancy. Canadians now live longer, but during these added years, they may be affected by disease or chronic conditions. For this reason, indicators used to monitor changes in population health and guide policy decisions need to include how health conditions affect the day-to-day functioning of Canadians over their lifetime.

28. In order to obtain a measure of functional health, we developed the Classification and Measurement System of Functional Health (CLAMES) to describe health states. CLAMES provides a standard set of eleven attributes, each with four or five levels to describe the level of functional capacity associated with each health state. We also conducted a study to derive a preference-based measure of health using CLAMES.

II. BACKGROUND

2.1 The measurement of Health State Preferences and the Multi-Attribute Approach:

29. Preferences for health state outcomes can be measured using values or utilities. The difference between the two measurements is that utilities are for applications that involve risk and are measured with questions that incorporate some notion of probability. The method we used to obtain utilities is the standard gamble, a method where participants reveal their indifference point between two alternatives, one containing uncertainty. (For details on this method, see Torrance, 1986.)

30. When a large number of health states are under study, it is unfeasible to measure preferences for each health state. We thus developed an experimental design for selecting a sample to be measured in the field that properly covered the space represented by these health states. From that sample, a utility function can be defined for the multi-attribute health state classification system.

31. Two main approaches have been used when trying to construct a function to estimate health state preferences not measured in the field: the statistical approach and the decomposed approach (Farquhar 1977). The statistical approach uses a regression model applied to a large number of multi-attributes states. This method requires direct measurement of many combinations of attribute levels. This method has been used successfully by Brazier (Brazier 2002). In order to be sure that the space defined by the attributes is properly covered, one could include in the set of states to be measured a sub-set of states consisting of a Latin hypercube sample. This ensures that all possible levels of all attributes are measured at least once.

[****] Prepared by Julie Bernier and Jean-Marie Berthelot, Statistics Canada

32. The second approach, called the decomposed approach, requires the measurement of preferences for a pre-defined set of states called "corner states" and "pure states" along with a small set of multi-attributes states. A corner state is a state where one attribute is at the worst level and all other attributes at their best levels. A pure state is a state where one attribute is at an intermediate level and all other ones are at their best level. This approach is used by Torrance (1995) to calculate the Multi-Attribute Utility Function (MAUF) associated with the Health Utilities Index (HUI) developed at McMaster and can be used as well to model disutilities (1-utility) as proposed by Le Galès (2001).

2.2 The decomposed approach

33. The choice of standard gamble as the method to elicit individual preferences is based on the fundamental axioms of von Neumann-Morgenstern utility theory (1944). When a series of health states are described using a multi-attribute classification system, utility theory state that a utility function can be defined for the system as long as the attributes respect certain conditions. The basic approach is to measure the eleven single-attribute preference functions and to determine an equation that calculates the overall preference score as a function of these single-attribute scores. The theory even specifies alternative functional forms to be considered and conditions under which each would be appropriate. The approach is called decomposed because once you have chosen the functional form, each parameter is in theory the true preference score associated with a specific health state (one of the set of "pure states"). With this technique, the parameters are estimated one at a time, which makes it different from a statistical technique. For the CLAMES system, the appropriate form would be:

$$u(E) = \left(\frac{c+1}{c}\right) - \frac{1}{c}\left[\prod_{i=1}^{11}\left(1 + c\, c_i d_{ij}\right) - 1\right] \quad \text{where} \quad c = \prod_{i=1}^{11}\left(1 + c\, c_i\right) - 1 \qquad (1)$$

where u is the utility associated with a health state E, the c_i represents the disutility (1-u) associated with the worst level of the attribute i, all other attributes being at the best level (corner state). The $c_i d_{ij}$ represents the disutility associated with the jth level of attribute i, all other attributes being at the best level (pure state), and finally c is a scaling parameter.

2.3 The statistical approach

34. With the statistical approach, one must also choose a functional form to work with. This function is expressed in terms of a certain number of parameters. Once a set of health states has been measured, the set of parameters is estimated simultaneously, according to a fitting criteria. Usually, the criterion is to pick the set of parameters that will minimise the difference between the observed value and the estimated ones. The functional forms considered for a multi-attribute function are the linear model and the log-linear model. Some adjustments can be made to these functions in order to obtain a better fit.

III. METHODS AND ANALYSIS:

3.1 Data collection

35. The measurement was done through a series of focus groups across Canada. Each group was made of 10 or 11 participants. A session is a day long and consists of 4 exercises. In order to do these exercises, the health states were split into two groups: the anchor states and the regular states. The anchor states are 12 selected states that were measured in every group and were

chosen to cover the complete range of health states from full health to dead It is possible to use the anchor states to measure inter-group variability or for comparison within specific socio-demographic groups. In addition to the anchor states, 226 states were evaluated. The following exercises were performed by the groups:

36. Exercise 1: The participants rated on a thermometer-like scale the 12 anchor states, from full health to dead. The participants were allowed to rank some of the states as being worse than death, but scores for states worse than death were recorded as 0. This instrument is useful for rating health states but does not directly provide valid cardinal utility measures since it does not involve uncertainty. This exercise is a useful step in helping participants to familiarize themselves with the terminology and the definitions used in the study and ensures a higher quality of subsequent exercises.

37. Exercise 2: We elicited preferences for the 12 anchor states using a standard gamble procedure adapted to a group setting. The protocol used for the standard gamble is a paper and pencil version adapted from a protocol developed by the McMaster Health Utilities Group (Furlong, 1990) and the University of York studies (Gudex, 1994). After having been taught the standard gamble using a ping-pong procedure (Protocol provided in Annex A), the participants indicated their preference for a particular health state. Scores were then shared with the group and the moderator led a short discussion among participants. After the discussion, the participants had a chance to modify their scores and then the group moved to the next anchor state. This discussion-type method is useful to ensure that the group has a common understanding of what the attributes mean. It also ensures that the participants take into account all the information provided in the definition while doing the rating. Both exercise 1 and 2 were conducted with anchor states presented in a random order that varied from one group to the other.

38. Exercise 3: Each participant received 10 additional health state descriptions. They individually did the standard gamble exercise without the discussion step. The selected states varied from one participant to the other to ensure a good coverage of the states under study and were allocated at random.

39. Exercise 4: Each participant received 4 special health state descriptions. They individually did the standard gamble exercise without the discussion step. These special cards represent corner states and pure states. The rationale for measuring these states is discussed in section 2.2.

3.2 Selection of participants

40. Canadian preference scores were measured in a series of 14 focus groups nation-wide. An effort was made to select participants from a variety of health and socio-demographic situations. Among other variables, range of illness experience, income, education, age and immigration status were taken into account. Each group was built by a recruitment agency following some inclusion guidelines and using a quota sampling. External consultants recruited participants either through random digit dialling or from existing research databases.

3.3 Experimental design

41. As mentioned previously, the number of health states that could be covered by CLAMES prevent us from doing direct measurement for all the states. The set of health states to be evaluated by participants was selected from a pre-established set that documented the impact on functional health of about 200 specific and prevalent diseases (including same disease, various

stages of severity). These 200 conditions were each mapped into an 11-tuple of CLAMES based on literature review and validation by an expert medical panels. These health states were used, without naming the disease, in the focus groups to elicit preferences using a standard gamble protocol. Therefore, a sampling strategy for selecting the states to be measured in the field was developed. This strategy is related to the method chosen for estimating the preference function. The study has been designed to allow for either statistical or decomposed approach: exercise 3 was done with a sample of states used in the statistical model and exercise 4 was performed on corner states and pure states. At the estimation stage, it was decided to use data from both of these exercises.

42. The sampling plan is related to the distribution of the health states to be evaluated among the participants participating in the focus groups and is made of two distinct experimental designs. The first design is for exercise 3. This exercise was done in 13 focus group, each group being made of 10 or 11 participants. The design was developed for an expected number of 135 participants: it consists of 135 series made of 10 different health states, each series to be used by one participant. In order to generate these series, 193 health states were used. Each state was measured by 7 participants except one that was measured only 6 times (some series were used an extra time because we ended up having 143 participants instead of the expected135). Three of these states were a subset of the anchor states studied during exercise 1 and 2 (to test for group effects). A systematic algorithm was used to generate the series. This algorithm generated 135 series of 10 numbers between 1 and 193, each number being repeated 6 or 7 times. The 193 health states to be studied were then randomly assigned a number between 1 and 193. Finally, the order in which the series were used was also determined at random.

43. The second design is for exercise 4. It was decided that this exercise would be conducted only when there was enough time. Because of that constraint, the design needed to be independent of the number of focus groups. This exercise is about measuring the preferences for corner states and pure states. With the list of attributes used to define the health states, there were 37 of these states. Seven supplementary health states were selected from the ones used in exercise 3. The design then consisted of randomly creating a series of 4 health states out of the 44 available (37 corner states and 7 regular states). Eleven series were generated and used in each focus group. The number of repetitions for each health state then depends on the number of focus groups that do this exercise. There was enough time to do the exercise with all the focus groups.

3.4 Estimation and some performance measures

44. In order to estimate a scoring function, we need to develop a function where a preference score is associated with any possible combination of levels for the 11 attributes. Therefore, a data point is made of one health state (defined as a 11-tuplet) and the corresponding preference score. Each health state under study was evaluated by several participants; the corresponding preference score has to be an aggregation of these scores. Some aggregation methods were tested and the simple mean was chosen. In addition, each health state was given a weight proportional to the number of participants who evaluated it. Based on the literature review, here are the functions tested with the statistical approach:

A) A log linear model

$$\ln(p) = \sum_{i=1}^{11} \sum_{j=1}^{5} I_{ij} x_{ij} \qquad \text{or} \qquad p = \prod_{i=1}^{11} \prod_{j=1}^{5} I_{ij} y_{ij}$$

Where I_{ij} is an indicator that takes value 1 if attribute i is at level j, value 0 elsewhere, x_{ij} (or y_{ij}) represent parameters associated with the dif ferent levels on each attribute.

B) An additive model with a correction at the bottom of the scale (Brazier et al 2002)

$$p = 1 - \sum_{i=1}^{11} \sum_{j=1}^{5} I_{ij} z_{ij} + J\ ad_1 + K\ ad_2$$

Where I_{ij} is and indicator that takes value 1 if attribute i is at level j, value 0 elsewhere, z_{ij} represent parameters associated with the different levels on each attribute, J is an indicator that takes value 1 if any of the seven first attributes is at its worse level and K is a similar indicator for the last four attributes.

45. We also built the model suggested in section 2.2 using a strict decomposed approach. The model didn't give a good fit to the data and we had a lot of concerns with it. First, the estimation by the decomposed technique made use of only a restricted set of health states (the 37 corner states and pure states). Second, these states were all measured at the end of the day. And last, some of these states were difficult to evaluate because they didn't seem realistic. For these reasons, it was decided to estimate the true value associated wit h the corner states and pure states by calculating adjusted means for the required states using the regression technique. One advantage of this method is to use the richness of the data collected since the full set of health state is used to estimate adjusted means.

46. Here are some measures of performance for the three models discussed:

Performance Measures for the Three Models

	Model log-linear	Model additive	Model decomposed
Mean error	-0.005	-0.009	-0.016
Mean square error (MSE)	0.005	0.004	.009
Weighted MSE	0.0025	.0022	.0044
R^2	0.97	.98	---------
Worst possible state	0.115	-0.10	-0.28
Best possible state	1	1	1

Where an error is, for a particular health state, the difference between the observed mean and the score estimated by the function.

IV. CONCLUSION:

47. Two of the estimated models provide a minimum score below 0. This is not unusual; in fact participants considered some of the tested health states as worse then death and these states were not at the very end of the scale. All the models have been specified so that their best possible state was worth 1 because the scale 0-1 is defined with 0 corresponding to dead and 1 to full health.

48. Intuitively, the first and third model make more sense, the multiplicative form meaning that any impairment that you add for a health state will take away a proportion of the functional health instead of a fixed amount. Our preference is the first model with a scaling adjustment to allow for a more realistic minimum.

49. Some validation still needs to be done; the plan for now is to do external validation by taking out a certain number of health states, re-estimating the model and then calculating a mean square error for the remaining states.

50. Another step required is the estimation of the variance associate with the estimates. Up to now the preferred option is to apply a Jackknife procedure to the groups.

V. REFERENCES:

1. Torrance George W,1986. "Measurement of Health-State Utilities for Economic Appraisal: A Review", J of Hlth Econ, vol.5:1-30.

2. Faquhar P.H., 1977. "A Survey of Multi-Attribute Utility Theory and Application", TIMS Studies in the Management Sciences, vol.6, pp.59-89.

3. Brazier J., Roberts J. Deverill M., 2002. "The estimation of a preference-based measure of health from the SF-36", Journal of Health Economics, no 21, pp. 271-292.

4. Torrance G.W., Feeny D., Furlong W. and Boyle M., 1995. "Multi-Attribute Health Status Classification Systems: Health Utilities Index", PharmacoEconomics, vol.7, no 6, pp. 490-502.

5. Le Galès C. et al 2001. " Développement d'un index d'états de santé pondéré par les utilités en population française : le Health utilities Index", Économie et Prévision, no 150-151, pp. 4-5.

6. von Neumann John, Morgenstern Oskar,1944. Theory of Games and Economic Behaviour. Princeton NJ: Princeton University Press.

7. Furlong, W., Feeny, D., Torrance, G. W., Barr, R., and Horsman, J. (1990). Guide to Design and Development of Health-State Utility Instrumentation. Paper 90-9, Centre for Health Economics and Policy Analysis: McMaster University, Hamilton, Ontario.

8. Measurement and Valuation of Health Group, Gudex, C (Ed). (1994). Standard Gamble User Manual: Props and Self-Completing Methods. Centre for Health Economics: University of York: York, England.