

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

NOTE TO USERS

The original manuscript received by UMI contains pages with slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI



Université d'Ottawa • University of Ottawa

**TITLE: A PILOT PROJECT TO EVALUATE PATIENT
PREFERENCES IN OSTEOPOROSIS**

AUTHOR: © Ann B. Cranney MB, FRCP C

**Submitted in Partial Fulfilment of the Requirements
for the Master of Science Degree in Epidemiology
University of Ottawa**

May 8, 1998



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Voire référence*

Our file *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-36678-2

Canada

TABLE OF CONTENTS

LIST OF TABLES	4
LIST OF FIGURES AND APPENDICES.....	5
ABSTRACT	6
ACKNOWLEDGEMENTS.....	8
1. INTRODUCTION	
1.1 Background on Osteoporosis	9
1.2 Outcome Measures for Cost-effectiveness Analyses.....	11
1.3 Cost-effectiveness Studies in Osteoporosis	14
1.4 Current status of Preferences in Osteoporosis	18
1.5 Elicitation of Preferences.....	20
1.6 Methodological Issues in Elicitation of Patient Preferences	25
1.7 Source of Preferences.....	26
1.8 Psychometric Properties of Preferences in Osteoporosis	27
2. OBJECTIVES	29
3. METHODS	
3.1 Study Design.....	30
3.2 Consent.....	31
3.3 Measures.....	32
3.4 Assessment Schedule.....	32
3.5 Description of Health Status Measures.....	32
3.6 Development of Scenarios for Elicitation of Preferences for Osteoporosis	33
4.0 Statistical Analysis.....	35
4.1 Sample Size	35
4.2 Statistics and Data Analysis.....	35
5. RESULTS	
5.1 Feasibility	38
5.2 Demographics.....	38
5.3 Reliability of the Feeling Thermometer and Standard Gamble	38
5.4 Comparison of Feeling Thermometer and Standard Gamble	39

5.5 Valuation of Different Marker States	39
5.6 Difference in the Preferences for the Marker States	40
5.7 SF-36 Scores for Different Groups.....	40
5.8 Construct Validity of Preference Assessment.....	41
5.9. Sensitivity to Change.....	41
5.10 Transformation of Feeling Thermometer Scores into Standard Gamble Scores.....	43
 6. DISCUSSION	
6.1 Reliability	43
6.2 Comparison of the Preferences Elicited by the Feeling Thermometer and Standard Gamble	44
6.3 Valuation of Different Marker States.....	45
6.4 Differences in Preferences for the Markers States Elicited from the Feeling Thermometer	45
6.5 SF-36 Scores for the Different Subgroups.....	46
6.6 Construct Validity	46
6.7 Sensitivity to Change	48
6.8 Transformation of Feeling Thermometer into Standard Gamble Scores.....	49
6.9 Limitations.....	50
6.9 Potential Biases.....	51
 7. CONCLUSION	52
 8. TABLES	54
 9. FIGURES.....	67
 10. REFERENCES	71
 11. APPENDICES.....	78
 11. GLOSSARY	94

<u>LIST OF TABLES:</u>		<u>Page</u>
Table 1.	Economic Evaluations of Therapeutic Interventions for Prevention and Treatment of Osteoporosis	16
Table 2.	Demographics of Patients	54
Table 3.	Reliability Coefficients (Test-retest) for Marker States and Current Health	55
Table 4.	Baseline Preferences for Current Health according to Group	56
Table 5.	Follow-up Preference Scores for Current Health	57
Table 6.	Baseline Preference Scores for Marker States Elicited with the Feeling Thermometer	58
Table 7.	Preference Scores for the Marker States at Baseline and Follow-up	59
Table 8.	Differences in Preference Scores for the Marker States (Elicited by the Feeling Thermometer)	60
Table 9.	Mean SF-36 Scores for each Group at Baseline	61
Table 10.	Correlation Coefficients between Baseline Preferences and Health Status Outcomes	62
Table 11.	Relationship of HUI and SF-36 with Feeling Thermometer and Standard Gamble	63
Table 12.	Sensitivity to Change of Preference Measures	64
Table 13.	Correlation Coefficients between Change in Preferences and SF-36 Outcomes	65
Table 14.	Summary of Psychometric Properties of Feeling Thermometer and the Standard Gamble	66

<u>LIST OF FIGURES:</u>		<u>Page</u>
Figure 1	Feeling Thermometer for ranking of Health States	22
Figure 2.	Baseline Preferences for Current Health	67
Figure 3.	Follow-up Preferences for Current Health for each Group	68
Figure 4.	Marker States Elicited using Feeling Thermometer at Baseline	69
Figure 5.	SF-36 Scores for each Group	70

LIST OF APPENDICES

A.	Consent Form	78
B.	Data Collection Form	80
C.	Script for Marker states	81

ABSTRACT:

Statement of the problem: Osteoporosis is a chronic disease which results in fractures with significant morbidity: the goal of treatment is to prevent these fractures and improve quality of life. Recently, quality of life measures have been incorporated into osteoporosis clinical trials. Preference-based measures are now recommended for the evaluation of the cost-effectiveness of new therapeutic interventions. There has not been any work evaluating the validity of preference measures in osteoporosis. The study objective was to develop and evaluate the psychometric properties of preference-based measures in osteoporosis.

Methods: Preference scenarios were constructed for marker health states associated with osteoporosis using the feeling thermometer. The reliability, validity and sensitivity to change of these measures and the standard gamble were evaluated in 42 women from four different patient subgroups. The four subgroups of women included women commencing hormone replacement, and women with recent wrist, vertebral or hip fractures. The reason for looking at different groups of women was to establish the feasibility of eliciting preferences in women with varying degrees of osteoporosis. Also, for cost-effectiveness analyses in osteoporosis, estimates of quality of life are required for the four different subgroups. The validity and sensitivity to change of the feeling thermometer and standard gamble were compared with the Health Utilities Index (HUI) and SF-36. All subgroups were reassessed 2-3 months following their first interview.

Results: Preference measurement was feasible in women of different age groups. The reliability coefficients for health states ranged from 0.65- 0.87. The preference scores

decreased according to the severity of the marker state which demonstrated content validity. The baseline preferences for current health using the feeling thermometer and the HUI were lower than the standard gamble in the fracture subgroups. The preferences elicited for current health using the feeling thermometer demonstrated a gradient according to severity, with the highest scores for the hormone replacement subjects and the lowest for the hip fracture subjects.

Convergent validity of the feeling thermometer was supported by a significant correlation with the HUI ($r=0.38$), and the physical health summary of the SF-36. The standard gamble did not correlate with the HUI ($r=0.15$) or the feeling thermometer ($r=0.09$), but correlated with two domains of the SF-36. The preference measures were sensitive to change with the efficiency scores ranging from 0.80-1.10.

Conclusion: The use of preference measurements in the evaluation of osteoporosis is feasible, reliable and these instruments can be used to obtain measurements of preferences for marker states associated with osteoporosis. The feeling thermometer and standard gamble appear to be related to different aspects of health-related quality of life. More work on the assessment of validity and sensitivity to change of the different instruments is required, prior to the use of these instruments in clinical trials.

ACKNOWLEDGEMENTS:

I thank the supervisors of this thesis project; D. Coyle, Dr. P. Tugwell and Dr. G. Wells for their assistance, time and patience. I would also like to thank Ba Pham for his statistical advice.

I thank Jackie Tetroe for her conscientious work in the study interview phase.

I would like to thank the following physicians: Drs. Elaine Jolly, Malcolm Son-Hing, B. Powers, E. Mierins and the orthopedic division of the Ottawa General Hospital for their assistance in the recruitment of patients.

I thank the Arthritis Society for their commitment of resources to allow me to pursue this degree.

Above all I would like to thank the women who volunteered their time, and allowed me to complete this project.

1. INTRODUCTION:

1.1 Background on Osteoporosis:

In Canada, osteoporosis is a debilitating disease, which results in a significant burden to society. The total acute health care costs in Canada attributable to osteoporosis are estimated to be as high as \$ 1.3 billion per year (1). In addition, 15 out of 100 women will fracture their hip during their lifetime (2,3). It is estimated that 8-10% of these women die within the first six months of their hip fracture and only one-third will return to their pre-fracture level of functioning. Postmenopausal women are particularly susceptible to the complications of osteoporosis which is characterized by the presence of fractures of the wrist, spine and hip. The number of women that will be at risk for this disease will increase as our populations ages.

Fortunately, there are currently many more treatments available to prevent and treat osteoporosis, including Hormonal replacement therapy, Bisphosphonates, Calcitonin and Vitamin D derivatives. New agents such as Raloxifene, a selective estrogen receptor modulator and an alternative to estrogen have recently been approved in the United States. Unfortunately, women may be required to take medications for many years, in order to prevent an osteoporotic-related fracture. Also, many of the therapies have unpleasant side-effects such as breast tenderness or may increase the risk of other diseases such as breast cancer. Some therapies have positive benefits on other diseases, such as reduction of cardiovascular disease and possibly Alzheimer's disease (4).

It is becoming increasingly difficult for both the practitioner and individual woman to decide which medication is best suited for them, given the complex risk-benefit profile. In order to make the decision a more evidence-based process, structured interventions such as decision aids

are being developed. Decision aids are used as an adjunct to physicians' counselling so that women can understand the probable risks and benefits of options and participate with their practitioner in the decision-making process.

Currently, our resources for health care treatments are limited and not all osteoporosis treatments that demonstrate efficacy can be supported. It is important that we not only identify cost-effective therapies that can prevent osteoporosis but that we incorporate quality of life/preference measures into these assessments, so that we do not undervalue potentially beneficial therapies.

Economic Evaluation in Clinical Decision-Making:

Economic evaluation is concerned with the most efficient allocation of available resources to gain the most benefit. Policymakers have advocated the use of economic evaluations to study the effectiveness of health care interventions and to assist in the allocation of resources.

Economic evaluations allow us to assess the cost-effectiveness of a new therapy/health care program in a systematic and unbiased fashion. The costs and outcomes of one therapy are compared to relevant alternatives and this information may then be used in the decision making process at various levels. When the methods of economic evaluation are applied to drug therapies, this is called pharmacoeconomics. Pharmacoeconomic studies play a role in a variety of decision-making situations: (a) Development and research of a new drug by a pharmaceutical company, (b) Pricing decisions by government regulators. (c) Formulary decisions in determining new listings of drugs (d) Post-marketing surveillance which monitors the performance of drugs including adverse effects and (e) Clinical guidelines for providers.

In the past, only the unit costs of the drug have been considered (5). It is essential that other

outcomes, such as the reduction of side effects, indirect costs, quality of life, toxicity and improved efficacy also be considered.

There is no standardised methodology for economic evaluations of health care, especially in comparison to the standards used to assess the efficacy of therapeutic interventions in randomized controlled trials (6,7). As a result, well-defined guidelines for economic analyses have been published (5,8-10). In Canada and Australia, economic evaluations are required prior to the approval of new pharmaceutical agents (10,11).

1.2 Outcome Measures for Cost-effectiveness Analyses:

The Canadian Guidelines for Economic Evaluation suggest that three types of clinical endpoints in addition to mortality rates be included: (5)

- (a) Generic quality of life outcomes
- (b) Disease-specific outcomes
- (c) Patient preferences for outcomes derived from measures such as the standard gamble and feeling thermometer.

(a) Generic instruments such as the SF-36 have been used to measure quality of life in osteoporosis patients. They allow comparisons across conditions but may fail to detect important differences between treatments (12). Health status measures such as the SF-36, describe a health state but not the preference for or valuation of that health state. The scores from the SF-36 and other generic measures are not measured relative to mortality, so they cannot be used to derive quality adjusted life years (QALYs) in economic analyses.

(b) Disease-specific instruments are designed to detect aspects of quality of life relevant to the particular condition. For example, the OPAQ- or Osteoporosis Assessment Questionnaire is

designed to assess the quality of life in vertebral fracture patients (13). Disease-specific instruments have the advantage of being more sensitive to clinically important differences. However, they do not allow for comparisons across conditions nor can they be used to derive QALYs in economic analyses.

(3) Preference measures assign values to particular health state to reflect the level of satisfaction that a person associates with a particular health state (14). Preference measures may be used in clinical trials where patients are requested to assign a value to a health state by balancing the positive treatment effects against the negative side effects. Preference measures may be less responsive to change than disease-specific instruments, although this may be related to the way in which we measure them. Preferences can be used for either individual decision making, or on a group level they can be aggregated across individuals to provide a group preference function.

Preference measures of quality of life are derived from economic and decision theory. In 1944 von Neumann- Morgenstern presented their theory of rational decision-making under uncertainty which was defined by fundamental axioms which are intended as a prescriptive model of how individuals should make decisions if they wish to act rationally (15). The original axioms include: (1) **independence** - if $a >$ (preferred or equivalent to) b , then the lottery offering a with probability p and c with probability $1-p >$ the lottery offering b with probability p and c with probability $1-p$. (2) **transitivity**: If $a >$ (preferred to) b , and $b > c$, then $a > c$. (3) **continuity of preferences**: if a is preferred to b , which is preferred to c , there is a probability p at which the individual is indifferent between outcome b with certainty or receiving the risky prospect made up of outcome a with probability p and outcome c with probability $1-p$ (16).

These axioms do not, however, allow for aggregation across individuals.

Health-related quality of life (HRQL) measures are being increasingly used as measures of quality of care and third party payers are beginning to use them in reimbursement decisions. Cost-effectiveness analyses require that HRQL is placed on a continuum, and that changes be followed for the duration of an individual's survival (14).

Summary measures that combine the effects of an intervention on quality and the length of life are useful way to compare therapeutic interventions. An outcome in cost-utility studies which incorporate quality of life is the Cost/Quality Adjusted Life Years gained, or cost/QALYs. Preference measures can provide a summary score (ranging from zero to 1) reflecting health-related quality of life and this score can then be used as a quality-adjustment factor in determining QALYs (14). For example, years of life gained are multiplied by the preference weights for the health status during those years in order to derive the number of quality adjusted life-years. An advantage of a QALY is that it captures gains from both reduced mortality and morbidity, and QALYs allow comparisons across diseases and programs in technology assessment, provided that the preferences are estimated in the same manner. The use of QALYs allows analysis of prevention programmes using a therapy that may have multiple positive and negative effects. For example, hormone replacement therapy has an effect on not only osteoporosis, but also heart disease and breast cancer. It is important to use measures in cost-effectiveness studies that are sensitive to change so that we can accurately assess the value of new medications on health-related quality of life.

An alternative to the QALY method is a more complicated method of measuring preferences for each unique health path - which is proposed in the *healthy -years equivalents* (HYEs)

method by Gafni et al. (17). This method focuses on the valuation of health paths instead of health states, and then converts the individual's preference for each health path into a healthy-year equivalent using a two-stage standard gamble technique (18). The HYE method has been criticised since it does not incorporate risk attitude and is a more complicated way of measuring of what is equivalent to time tradeoff values (19). More research is required to determine if the HYE approach is superior to the QALY model and whether it is feasible and valid.

1.3 Cost-effectiveness Studies in Osteoporosis

The U. S. National Osteoporosis Foundation is conducting a cost-effectiveness study comparing various therapeutic strategies for osteoporosis. In Canada, the Canadian Coordinating Office of Health Technology Assessment (CCOHTA) is also conducting a cost-effectiveness analysis of alternative therapies for the treatment of postmenopausal osteoporosis. One of the important outcomes of these analyses is to assess the impact of therapeutic interventions for osteoporosis on health-related quality of life.

There have been a number of cost-effectiveness studies of therapies dealing with the prevention or treatment of established osteoporosis (20, 21, 23-28,30,35). These were identified by a Medline search which was designed to identify published cost-effectiveness studies from 1975-1997. The MeSH search terms were osteoporosis, costs, cost-benefit analysis, and cost-effectiveness. Sixty potential references were retrieved. The author reviewed these references in addition to the bibliographies of each article. Of these only ten were actual cost-effectiveness studies, with the other articles dealing with costs/burden of osteoporosis. An outline of these ten analyses is presented in Figure 1. In particular, each study was classified according to the intervention, the effectiveness measure used in the cost-effectiveness ratio, the viewpoint, the

source of preferences and the discount rate. The majority of these analyses dealt with the intervention: hormone replacement therapy. The outcomes, methodology, and the use of preference measures were variable. Seven studies out of ten conducted a cost-utility analysis, with cost/QALY as the outcome. The view point was stated explicitly in six of the analyses. Five of the analyses considered discounting (20,24,27,26,28) and all of the studies performed sensitivity analyses.

Only one study elicited preferences from patients who had experienced a fracture (24). Andersen et al. interviewed small numbers of subjects, using the EuroQol and 15-D instruments and found considerable discrepancy between the EuroQol and 15-D estimates and chose to use the 15-D estimates in their analysis. They assumed that vertebral fracture patients had a quality of life of 0.80 the first year after the fracture and from the second year onwards had a quality of life equal to 1.0. This assumption could be criticized on the basis that many patients with vertebral fractures suffer chronic pain and are at much greater risk of having another vertebral fracture.

The result of reviewing the above cost-effectiveness studies highlights the need for more rigour and uniformity with respect to the methodology used for the estimation of QALYs.

Table 1 Economic evaluations of therapeutic interventions for osteoporosis prevention and treatment.

Author, Year of Publication, Country Reference	Intervention	Viewpoint	Outcomes/	Sensitivity Analysis	Elicitation of Preferences
Ross, 1988 Hawaii (35)	hypothetical treatment starting at age 50 that would slow bone loss by 50%	not stated	(1) fracture rates (2) cost/year	varied the value of bone density at which a fracture occurred	preferences not elicited
Tosteson, 1990 USA (27)	(1) no treatment versus (2) screening or universal treatment with HRT	societal	(1) incremental C/E ratios, (2) cost/QALY	varied discount rate 0- 10%, cost estimates and relative risk for hip fracture	preferences estimated
Weinstein, 1990 USA (30)	(1) estrogen/progesterone (2) estrogen alone in women aged 50 years for durations of 5 and 15 years	societal	(1) cost/QALY	varied effect of treatment on quality of life	preferences estimated
Daly 1992 UK (20)	(1) estrogen in hysterectomized women (2) estrogen/progesterone in non-hysterectomized women (3) estrogen in non-hysterectomized women	societal	(1) average life year gained (2) cost/QALY	varied the duration of treatment, drugs costs, discount rate	arbitrary preference values used - 0.95 for severe menopausal symptoms and 0.99 for mild symptoms.
Cheung, 1992 Australia (28)	(1) estrogen (2) estrogen/progesterone (3) non therapy for a duration of 5, 10, 15 years	not stated	cost/QALY	varied relative risks of myocardial infarction and discount rate	preference values based on expert opinion

Author, Year of Publication, Country Reference	Intervention	Viewpoint	Outcomes	Sensitivity Analysis	Elicitation of Preferences
Geelhoed, 1994 Australia (2-1)	(1) HRT for 50 years (2) HRT until age 65 (3) HRT starting at 65 (4) Calcium and exercise	not stated	(1) cost/QALY (2) cost/life year gained	-excluded the risk of breast cancer -eliminated benefits on cardiovascular system	arbitrary preference values used
US Technology assessment 1995 (23)	(1) screening/HRT treatment at age 50, 65 years versus (2) universal treatment;	societal	(1) cost per year of life added	- varied relative risks of breast cancer, MI, and the degree to which HRT prevents bone loss	patient preferences not used
Jonsson, 1995 Sweden (26)	5 year treatment with an agent which reduces fracture by 50% in a 62 year old woman with osteoporosis	not stated	(1) cost per life year gained (2) cost per QALY gained	(1) duration of treatment (2) discounting	arbitrary definition of preferences: 0.8 for the first year after fracture, 0.9, 0.4 for those returning to normal function and those severely disabled
Andersen, 1996 Norway (24)	Treatment for 5 years with alendronate versus no intervention in a 65 year old osteoporotic woman	societal	(1) cost/life year saved (2) cost/QALY (3) cost per hip fracture avoided	varied bone density, and discount rate 5-7%	preferences derived indirectly using the EurQoI and the 15D
Ankjaer-Jensen 1996 Denmark (25)	(1) calcium x 5 years (2) calcitonin x 5 years (3) etidronate x 5 years	societal	(1) cost per hip fracture avoided	varied costs, compliance with treatment	preferences not used

1.4 Current Status of Preferences in Osteoporosis

Osteoporosis differs from many disease states in that individuals are often asymptomatic for 20-25 years prior to experiencing an osteoporotic-related fracture, which includes fractures of the hip, vertebrae or wrist. Mortality, morbidity and the economic burden related to hip fractures has been well described in the literature (1). Wrist fractures are usually self-limiting and have only temporary effects on quality of life, whereas vertebral and hip fractures can result in significant deformity, chronic pain and disability. Detailed knowledge about the epidemiology and the impact of vertebral and hip fractures on quality of life is lacking.

A Medline search (1975-1997) for studies involving preference measures in this area, revealed that little work has been conducted in osteoporosis. Researchers have elicited preferences in women on HRT, both with and without osteoporosis (20,36,37) but there are no published studies on preferences in patients on other therapeutic regimens. Andersen et al. conducted interviews using the indirect methods including the EuroQol and 15-D instruments in a small sample of women with hip fractures (n=11), and vertebral fractures (n=10) as part of a cost-effectiveness study for alendronate (24). Using the 15-D data, they found similar quality of life for women with hip or vertebral fractures. Hillner calculated quality weights for hip fracture in the year the fracture occurred to be 0.95 for an uncomplicated hip fracture and 0.76 for a disabling fracture, based on opinion from an expert panel (38). Fordham elicited quality of life from 132 hip fracture patients at one and two years using the EuroQol, with a mean score of 0.61 and 0.60 respectively, suggesting that quality of life did not decrease after one year post-fracture (39).

Only Tosteson et al., (36,40) have elicited preference measures using an automated time trade-off (U-titer). U-titer refers to an interactive computer program that is designed to provide a

standardized instrument to assess patients' preferences with the values calculated automatically (41). It has been used to assess a subject's perception of a hypothetical fracture, a subject's perception of overall health and specific adverse events associated with their therapy. In this study subjects were interviewed at one to five years post fracture (42). The mean current health score for vertebral fracture subjects was 0.81 (n=48), and 0.60 for hip fracture subjects (n=40). This means that vertebral fracture and hip fracture subjects, respectively, would give up 19% and 40% of their remaining life years to attain perfect health as compared to non-fracture subjects who would only give up 9%. Tosteson's work is the most comprehensive work to date and her work confirms that both vertebral and hip fractures are associated with large decrements in quality of life. Tosteson et al. have tested the time trade-off (U-titer) in international phase III clinical trials for Raloxifene hydrochloride, a new therapy for osteoporosis. In this trial the elicitation of preferences for side-effects has incorporated. One criticism of Tosteson's work is that when the preferences are elicited, sleep instead of death is used as the metaphor for trading time. The values from the time-trade-off have not been compared to the standard gamble or feeling thermometer values. Also, the elicitation of preferences has not been conducted in a cohort of women over a period of time to assess changes in quality of life that occur after a fracture.

Perhaps the reason more work has not been conducted in this field is that elicitation of preferences is a relatively new discipline and that it is challenging to elicit preferences from patients with hip fractures since these women are often cognitively impaired. Likewise, vertebral fractures may go undetected for years so it is challenging to assess the impact of vertebral fractures on quality of life.

Preferences are an essential component of economic evaluations and decision analyses.

There is a need to improve upon the elicitation of preferences in osteoporosis, given that more cost-effectiveness analyses will be conducted in the future with the increased number of possible therapies for the prevention and treatment of osteoporotic-related fractures.

1.5 Elicitation of Patient Preferences:

Patient preferences have been used to determine the desirability of different health states in various areas of medicine: e.g., rheumatology- ankylosing spondylitis, fibromyalgia (43,44) dermatology - psoriasis (45), renal failure (46), in orthopedics - knee replacement surgery (47) and obesity (48).

Three common techniques have been described for eliciting cardinal preferences directly:

(i) feeling thermometer, (ii) time trade-off and (iii) standard gamble. The standard gamble is a technique which elicits a utility, in that it is measured under uncertainty and incorporates risk assessment, whereas the time-trade-off and feeling thermometer measure values or the strength of preference, under certainty (49). For the purpose of the thesis, the term preferences will be used to describe both values and utilities.

The standard gamble requires patients to choose between living in a less than optimal chronic health state or taking a gamble with a treatment with a certain probability (p) of perfect health, but also risking a corresponding probability ($1-p$) of immediate death. The probability of perfect health is systematically varied until a person is indifferent between the certainty and the gamble. For example, a patient with a hip fracture in a nursing home could be asked whether they would prefer to continue living in their present state or take a gamble with a treatment that would have two outcomes: immediate perfect health or death. The more severe the health state you are

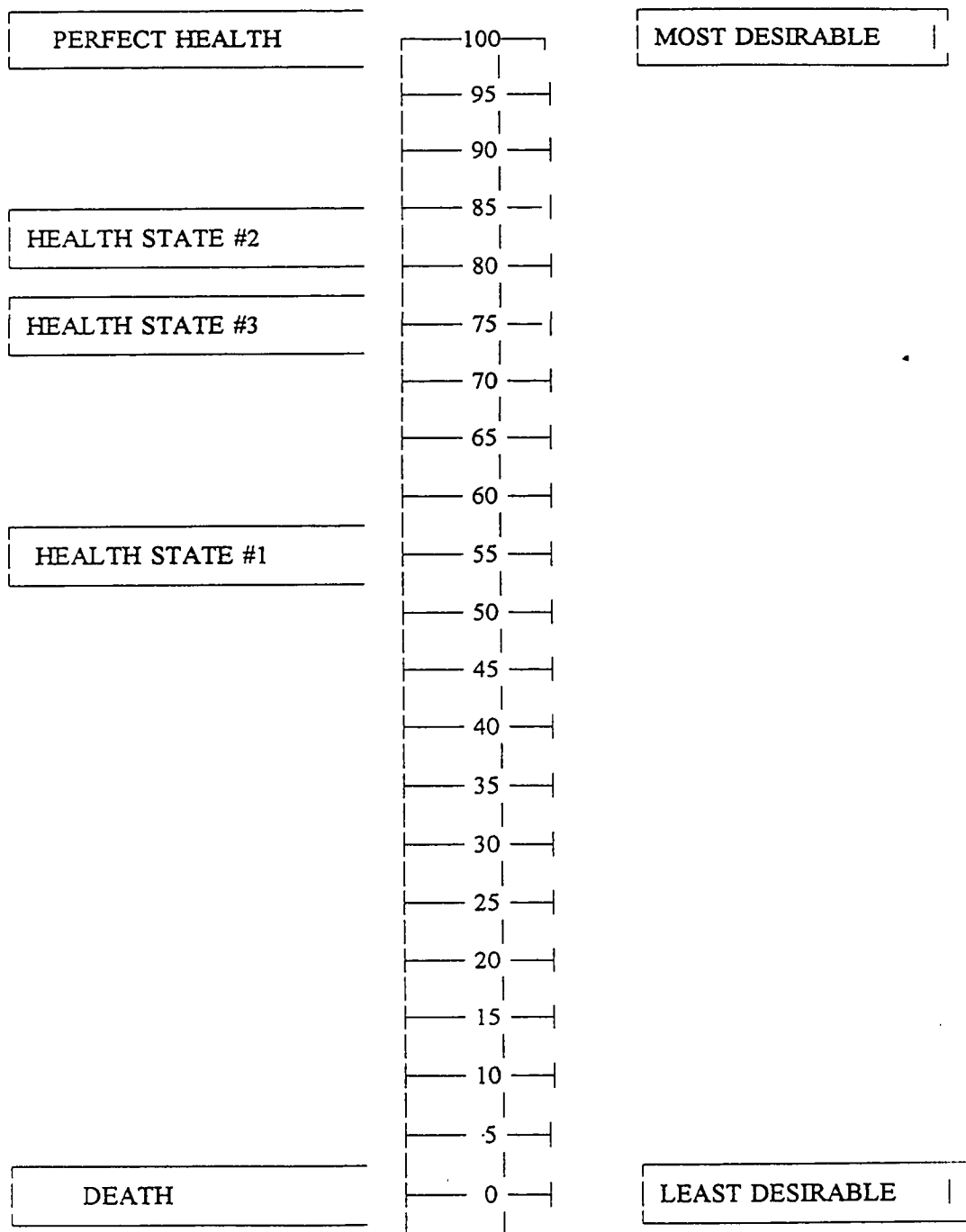
evaluating, the higher the risk of death the individual will be willing to take to avoid the current state.

Many researchers argue that the standard gamble is the only correct method since it is derived from von Neuman-Morgenstern (vNM) utility theory and has a strong axiomatic base (15,50). These axioms are intended as a normative model of how individuals ought to make decisions if they wish to act rationally. Individual behaviour does not always follow the normative model. The conceptual framework of the standard gamble allows for the substitution of outcomes other than perfect health or death (51). The standard gamble has been criticised because it is often confusing to patients and some patients resist the idea of gambling with their own health (49).

In the time trade-off technique, patients are asked to choose between living a fixed number of years in a less than perfect health state or living a lesser number in perfect health (52). The number of years of life that patients are willing to give up to gain perfect health is a measure of their value for the chronic health state. The time- trade off technique may be easier for people to comprehend but it is not based on the axioms of utility theory. Gafni et al. have pointed out that discrepancies exist between the time trade-off and other preferences (17). Time trade-off questions are easier to administer than the standard gamble and in general give lower scores than the SG.

The feeling thermometer (also known as rating scale) is derived from psychological scaling (53) and consists of a horizontal or vertical scale anchored by defined endpoints: death (zero worst imaginable state) at one end and perfect health (100 - best imaginable state) at the other (54) (Figure 1). Patients are requested to rate the desirability of different health (marker) states and his/her own health state along an interval scale between the anchors.

Figure 1: The Feeling thermometer for ranking of Health States



Patients tend to distribute the health states over the entire scale, even if true values are at the end of the scale, a phenomenon which is termed response spreading (i.e. they take it as a ranking than a metric scale). (55). Kaplan has demonstrated that response spreading does not occur if the endpoints of the feeling thermometer are well defined (56). Preferences reflect not only the values of various health states relative to perfect health or death (or the most severe health state) but also the magnitude of the difference between different health states. Preferences from the feeling thermometer are easier and efficient to obtain, but fail to present the subject with a decision.

The standard gamble is the product of the (1) strength of preference and (2) attitude toward risk. The feeling thermometer only elicits the strength of preferences. A number of researchers including Torrance believe that there is a stable relationship between the standard gamble and the rating scale (RS) valuations at the group level. Torrance has used the following formula: $1-SG=(1-RS)^{\beta}$ where RS is the rating scale (feeling thermometer) valuation and beta the coefficient for the power curve which reflects relative risk attitude (60). The exponent can be estimated from a subset of states that have been measured on both the feeling thermometer and the standard gamble. The value of beta is estimated for the original data set by transforming $[1-SG = (1-RS)^{\beta}$ to $\ln(1-SG) = \beta \ln(1-RS)]$ using a linear regression. The conversion approach can be applied at the individual or at the group mean level (54).

Preferences may be assessed indirectly using health status measures such as the HUI, and EuroQol. These are derived from multi-attribute utility theory and have been used in the general population for calculation of population health indices. Preferences are derived in two stages. First a person's health status is elicited along several dimensions. Next, a preference for that particular health state is calculated, based on preferences obtained from previous populations. In

this way, data collection is simplified in that simple health status questions can be asked, rather than the more complicated preference elicitation techniques. However, the indirect measures assume that the preferences for health elicited from previous populations correspond to those of the current study population. Multi-attribute theory has been particularly useful when preferences are required for a large number of health states.

The HUI is an example of the generic approach and is available in a self-administered, telephone or interviewer-administered formats. Torrance and colleagues at McMaster have developed three multi-attribute health status classification systems: HUI :1, HUI: 2, HUI :3 (53). Scores were derived from the preferences of the general public. The HUI: Mark 1 has four attributes, and was initially developed for the economic evaluation of neonatal intensive care and then modified for use in the general population. The HUI : 2 system consists of seven attributes: sensation, mobility, emotion, cognition, self-care, pain and fertility, with 5 levels of function per attribute. Each combination of levels across attributes defines a unique health state. The HUI: 2 was developed based on research to determine the attributes of health status that were considered important by the general public. Preference scoring for the HUI 2 was determined using the visual analogue scale and standard gamble. The HUI has been validated in clinical trials and has shown to be responsive to clinical change (57). The HUI systems are used in clinical trials, cost-effectiveness analyses and in population health surveys.

The EuroQol is another generic preference-based instrument that consists of five dimensions, three levels and classifies patients into one of 245 health states ($3^5 = 243 + \text{dead} + \text{unconscious}$) (58,59). The preferences for these health states have been derived from other population samples using a feeling thermometer. Therefore, potential sources of difference between the HUI and

EuroQol include: attributes and levels which are measured; the source of preferences and the structure of scoring.

1.6 Methodological Issues in Elicitation of Preferences:

In actual practice, these different elicitation methods yield different results which represent a problem when preferences are used for weighting. Torrance evaluated the SG, TTO and feeling thermometer in 43 alumni who rated six scenarios (60). The correlation between SG and TTO for 258 ratings was 0.65, in contrast to lower SG-feeling thermometer correlation of 0.36. The standard gamble measures patient preferences under conditions of uncertainty whereas the feeling thermometer and the TTO do not capture the subject's attitude toward risk. For this reason, the standard gamble gives the highest score, in situations where people are risk averse. In risk averse subjects their indifference probability increases, which increases their preference for the health state. Also the TTO would only yield von Neumann-Morgenstern utilities if the individual's preference function was linear in time (55). Bass et al. state that the rating scale and the SG do not measure the same thing unless the patient is risk neutral (61). Kahneman and Tversky have demonstrated that subjects are risk averse when choices are framed in terms of potential gains and as risk-seeking when choices were framed in potential losses. Some people argue that a patient's reluctance to give up life may distort the values given to health states (62). The standard gamble values may be higher since people tend to overweight sure outcomes compared to highly probable outcomes (63). Head- to-head studies which have compared the three preference measurements have found that standard gamble scores exceed time trade-off scores which in turn exceed feeling thermometer scores (14).

Patients may demonstrate what is termed *regret aversion* in that they fear they will lose the

gamble and end up with the worst outcome (60). Overweighting the chance of death from a treatment may reinforce the attractiveness of the subject's current health.

Differences in scores between the standard gamble and feeling thermometer scores may also be attributed to emotional factors. For example, more intense emotional reactions may occur to the standard gamble if a family member has recently died.

There have been few direct head to head comparisons of different preference measures. Moreover, many different methods have been developed to present the particular health states to patients, consisting of interactive videos, audio-booklets or velcro decision boards (64). It is not clear which method is most effective at presenting the necessary information.

1.7 Source of Preferences:

The QALY concept does not specify how to determine the preferences or whether they should be determined by patients or by the general public (16). Patient groups include those with the condition, those who have previously experienced the condition of interest and patients who have no direct experience of the condition. There is evidence that patients experiencing a health state tend to rate that health state higher than would the general public (65). The appropriate source of preferences depends on the viewpoint and ultimate use of the analysis. For example if one is evaluating a screening program, the public's preferences should be important, since a screening program is designed to prevent the public from entering the undesirable health state. Preferences elicited from patients with the condition would be more relevant when designing a decision-aid for treatment of a particular condition (66).

There is a need to establish the feasibility and validity of using preferences in the area of osteoporosis, so that we can confidently use these measures in cost-utility and decision analyses.

Currently, the reliability and validity of preference measurement techniques are uncertain and specific scenarios to elicit preferences in osteoporosis have not been developed and evaluated. Disease-specific instruments for osteoporosis are being developed, but are not ready for use and they will be limited to comparing within osteoporotic treatments, and not across interventions. Therefore instruments measuring preferences for osteoporosis should be compared to existing generic measures such as the SF-36, until disease-specific instruments are fully developed.

1.8 Psychometric Properties of Preferences in Osteoporosis:

Few studies of osteoporosis have examined the psychometric properties of their preference instrument, in particular: reliability, responsiveness to change and construct validity (67). Prior to recommending that specific methods of preference measurement be used for cost-effectiveness analyses in osteoporosis, it is important that their validity be established. Validity refers to the five main types of validity: (31,32).

(1) **Face validity:** Are the scenarios clinically credible and the results easily interpretable by not only clinicians, but the patients and policymakers? In this study, evaluating scenarios of women with different types of fractures that are related to osteoporosis increases the face validity of the study.

(2) **Content:** Do the preferences cover multiple domains of improvement in osteoporosis such as pain, disability and activities of daily living? The reason for looking at different groups of subjects was to assess patient preferences in women with varying degrees of osteoporosis and different levels of disability.

(3) **Criterion validity:** refers to the extent to which a measuring instrument produces the same results as a gold standard or criterion measure. There is no gold standard to measure health state

preferences.

(4) Construct Validity: A preference should produce results which are consistent with theoretically derived hypotheses concerning the concepts or constructs that are being measured. For example, certain apriori hypotheses can be defined for this study: (I) women with a hip or vertebral fracture should have lower preference scores for current health than women without a fracture and (II) the presence of medication side-effects (which are incorporated in the marker states) should result in a lower preference than the absence of side-effects. The preference scores will have construct validity if it follows these hypotheses. Convergent validity refers to the extent to which the selected measurement technique agrees with other accepted measurement techniques. In this study one would expect a positive correlation between the feeling thermometer and standard gamble scores with both the HUI scores and certain domains of the SF-36, such as the physical function and role - physical. Likewise, one would expect to find a stronger correlation between the feeling thermometer and the HUI than with the standard gamble since the feeling thermometer scenarios are based on the HUI. Divergent validity refers to the fact that the instrument should not correlate with unrelated variables or constructs. In this study I would not expect a correlation between the current health score of the feeling thermometer and the sensation or cognition attributes of the HUI, since osteoporotic-related fractures should not impact on these aspects of quality of life.

(5) Sensitivity to change: Does the estimation technique detect the smallest clinically important improvement, for example does the standard gamble detect change in preferences for a woman who is recovering from a hip fracture?

2. OBJECTIVES OF THESIS:

- (1) To elicit preference measures for specific health states associated with osteoporosis in different subgroups of women using the feeling thermometer. The three major osteoporotic-related fractures (wrist, vertebral, hip) were considered to be separate marker states when developing scenarios for measurement of preferences in osteoporosis. The different subgroups included women commencing hormone replacement therapy, and women who have recently experienced a wrist, vertebral or hip fracture.
- (2) To assess the validity, reliability and sensitivity to change of the standard gamble and feeling thermometer in eliciting preferences for marker states associated with osteoporosis.
- (3) To evaluate the stability of the marker states (functional status of women with different osteoporotic fractures) over time.

3. METHODS:

3.1 Study Design: A cohort followed over a two-month period

Study Population

Inclusion Criteria:

Two major groups of patients were interviewed:

(1) Patients who had not yet experienced a fracture, but was osteopenic and on hormone replacement (HRT) for either osteoporosis prevention or other reasons in order to capture the preference of women on HRT. Patients recently commenced on hormone replacement were identified with the assistance of the nurses and clinicians at the menopause clinic, orthopedic and geriatric wards and from rheumatology clinics.

(2) Females over age 50 with osteoporotic-related fractures of hip, vertebrae, or wrist.

These women were identified after their fracture and interviewed within the first couple of months and then on a second occasion 2-3 months later. Subjects were identified through the fracture clinics, the orthopedic and geriatric wards. Using the ICD-9 codes (733, 805, 820) for vertebral and hip fractures, the charts of female patients admitted with vertebral and hip (inter-trochanteric and femoral neck) fractures were also reviewed for their eligibility via the above criteria (60 women in total). All fractures were confirmed by X-ray.

Exclusion Criteria:

Patients were excluded if they were:

- (1) cognitively impaired
- (2) unable/unwilling to give consent
- (3) not fluent in English

(4) severely depressed

(5) deaf or visually impaired

(6) or had a pathological fracture (related to a known malignancy), since in this case the subject may have a limited life span related to their malignancy which could influence their preference scores.

Re-interviews were conducted in ten patients from each of the four subgroups at two weeks to evaluate test-retest reliability .

An initial feasibility study was conducted in ten patients from the different subgroups with different osteoporotic-related fractures.

3.2 Consent:

The patients were identified as potential candidates as described above and then they were contacted by the author. The objectives of the study were then discussed and they were asked if they would like to participate. At the initial interview, written consent for participation was requested (Appendix A). Confidentiality was maintained by assigning numbers to all patients and patients' names were recorded for the purpose of recontacting for a second interview. The study protocol was approved by the Research Ethics committees of the Ottawa General and Ottawa Civic Hospitals.

3.3 Measures:

Study interviews were performed by AC and JT, with the same interviewer performing both the baseline and two month follow-up interviews with each patient. Prior to elicitation of preferences, an introduction to the interview was given.

(i) For each patient at baseline, demographic information was recorded including: name, level of

education, type of fracture, use of hormone replacement therapy and presence of co-morbidity.

Subjects were not excluded if they had a co-morbid condition (apart from those mentioned above), since this was a pilot project. Co-morbidity was treated as a categorical variable.

(ii) Three health status measures were administered: (a) the SF-36 generic measure, (b) the HAQ pain scale, and for women on HRT, (c) a menopausal quality of a life instrument (MENQol)

(34). (iii) Next preferences were elicited and this included direct measures using: (1) the feeling thermometer (five different marker states, and current health), (2) standard gamble, (current health) and (3) indirect measures using the HUI - Mark II. Velcro boards were used to present the scenarios for the feeling thermometer- and standard gamble. For the HUI, the attribute of fertility was not included in the calculation of the total score, since it was not felt to be relevant given that the subjects were postmenopausal (Dr. G. Torrance, personal communication).

3.4 Assessment Schedule:

Baseline Visit: Demographics, Feeling thermometer (Current Health, Marker States), Standard Gamble, HUI, SF-36, HAQ pain scale.

Follow-up Visit: (2-3 months after baseline) Feeling thermometer (Current Health, Marker States), Standard Gamble, HUI, SF-36, HAQ pain scale.

3.5 Description of Health Status Measures:

(1) Short-Form -36 (SF-36). The SF-36 is a 36-item questionnaire including multi-item scales which measure eight domains detailing the health-related quality of life experienced by a person over the past month (12,68). 1. Physical Function (PF), 2. Role function as limited by physical problems (RF), 3. Bodily pain (BP), 4. General health perceptions (GH), 5. Vitality (VT), 6. Social function (SF), 7. Role function as limited by emotional problems (RE), and 8. Mental

health (MH). Approximately, 80-85% of the reliable variance in the eight scales was explained by two domains which lead to the construction of two summary measures - the Physical Component Summary and the Mental Health Component summary (68). The importance of the eight scales is that they allow us to assess particular domains of health that we think would correlate with the preferences of subjects who have osteoporosis.

(2) The HAQ pain scale was measured on a 15-centimetre horizontal visual analogue scale with terminal markers anchored to no pain (0) and very severe pain (100). A time frame of one week was used (69).

(3) Health Utilities Index - as described earlier on page 23.

3.6 Development of the Scenarios for Elicitation of Preferences for Osteoporosis

There are no evidence-based guidelines for the development of scenarios to elicit preferences. A script for conducting the preference assessments was developed, in order to standardize the interview and make the process clearer (Appendix C). The scenarios were presented using a health state classification system format and were developed by the author based on clinical knowledge and published guidelines that have been derived by George Torrance. The health state classification approach requires the investigator to identify the attributes that define the disease problem, resulting in a comprehensive description of the levels of functioning. After the scenarios were developed, two experts in the area of health economics/preference assessment and four experts (physicians, nurses) involved in the treatment of osteoporosis/menopause were requested to examine the script for content validity. Specifically, they assisted in the development of the health state descriptions associated with osteoporosis. The script was also shown to and evaluated by ten patients who were on hormone

replacement or had experienced an osteoporotic-related fracture in the previous year.

Five marker states for use with the feeling thermometer were constructed to closely resemble osteoporotic states of varying severity. The description of each marker state covered six dimensions: 1) activities of daily living, 2) self-care functions, 3) anxiety and depression, 4) leisure activities, 5) pain and 6) side-effects from treatment. In order to assess the relative importance of side-effects on the overall preference score, identical scenarios were developed for patients with and without side-effects from their medication. The side-effects described with hormonal replacement were: bleeding, breast tenderness, mood swings, weight gain, headache, nausea and record any side-effects and the degree of severity. For each of the marker states, duration was specified as for the *rest of your life*. The marker states describe varying degrees of severity of the disease state in this case: women with wrist, vertebral and hip fractures. The clinical marker states were designed to assist in the clinical interpretation of the meaning of the preference score results and also to assist the individual in determining the position of their own health state (current health) on the overall spectrum (16).

After rating their preferences for the five marker states in a manner that the difference between states represents differences in their preferences, the patients were requested to evaluate their own health. Each patient was requested to describe her own health state by answering questions based on their functional levels (1 corresponding to the best and 5 being the worst functional level). They were also asked if they are on hormonal replacement and if so asked to record the side effects and rate these as moderate or severe.

For the feeling thermometer portion of the interview a sample scenario on how to use the board was given initially. Perfect health (100) and deaths (0) were used as anchors for the

feeling thermometer. After the feeling thermometer portion of the interview was completed, then preferences using the standard gamble were assessed. A probability wheel was used as a visual aid to facilitate the patient's understanding (60). Each patient was offered a choice between remaining in their current health and a gamble with chance p to obtain perfect health and a chance $1-p$ of death. Chance p was systematically varied until the patient was indifferent between current health and taking the gamble, in steps of 10% ($p/1-p$: 100/0, 90/10, 10/90, 20/80 etc.).

4 Statistical Analysis:

4.1 Sample Size Considerations:

For the purposes of a validation study, 10-15 patients per subgroup were interviewed with a total sample size of 42. It was difficult to calculate an accurate sample size due to lack of data on preferences in osteoporosis patients, so the sample size was based on previous validation studies.

4.2 Statistics and Data Analysis:

Descriptive statistics measured to describe the sample: Means, medians and standard deviations were calculated for relevant patient characteristics. Means, medians and the range of the patient preferences at baseline and follow-up for each of the four groups were calculated according to each instrument. Means for the different sub-scale scores of the SF-36 were calculated.

Reliability:

Reliability coefficients were calculated to evaluate short-term test-retest reliability for the marker states, and current health values. Test-retest reliability was assessed by the patient's

valuation of the different marker states; their evaluation of their current health and the standard gamble at two weeks compared to the baseline assessment using an intra-class correlation (ICC) (70,71). There are six types of intra-class correlations described by Shrout and Fleiss (72,73) which are based on the statistical technique of Analysis of Variance (ANOVA) and express the reliability as a ratio of true to total variance. Conventional treatments of reliability omit the effect associated with observer variability (74).

The reliability coefficient was calculated based on the difference between two assessments using a one way ANOVA, with the subject as an independent factor and the difference between scores as the dependent factor. The accepted standard for reliability coefficient is a value greater than 0.70 (74).

Construct validity:

Construct validity was assessed by verification of certain hypotheses: that preferences derived from the hip and vertebral fracture subjects (using the feeling thermometer and standard gamble) would be lower than scores for the wrist fracture and HRT subjects. Likewise a stronger correlation between the feeling thermometer and the HUI would be expected than with the standard gamble, given that the feeling thermometer scenarios are based on the HUI. Next given that the feeling thermometer and standard gamble are assessing quality of life, preference scores from these instruments should correlate with certain domains of the SF-36, (such as the physical function domain) and not with domains such as mental health. These correlations were tested using Spearman's correlation coefficient.

Sensitivity to change:

Sensitivity to change was calculated using two approaches: First, the correlation of the change (follow-up minus baseline) in preferences with changes in the eight domains of the SF-36 were examined. Second, an efficiency score as described by Anderson was used (75). An **efficiency** score was calculated from the mean changes of the preference measure divided by the standard deviation of the change measure $E=d/SD$ where d is the mean of the change scores for the group and SD is the standard deviation of the change measure. Each individual was asked if they had improved subjectively, which was used as a criterion of clinical change and the efficiency measure was calculated for individuals who felt that they had improved.

Transformation of feeling thermometer into standard gamble values:

Torrance has suggested that the standard gamble preferences and the feeling thermometer are related by the formula described earlier on page 23, where Beta is the coefficient reflecting relative risk attitude [$\ln(1-u) = \beta \ln(1-v)$] (60). We tested this relationship in order: (a) to assess the suitability of using the feeling thermometer as a measure of preference in place of the standard gamble and (b) to assess the risk attitudes of the subjects. Beta can be estimated from the subset of states that have been measured on both the feeling thermometer and standard gamble. The feeling thermometer scores were converted to standard gamble scores for each of the groups in order to derive beta, which is the coefficient from the power curve. A beta coefficient greater than one is indicative of risk-averse behaviour.

5. RESULTS:

5.1 Feasibility:

All interviews were completed (baseline and follow-up) except in one individual who died prior to the follow-up visit and in one woman who was lost to follow-up. The time range to complete the full interview was 45-60 minutes.

Thirty of 60 women with hip or vertebral fractures were ineligible due to cognitive impairment.

5.2 Demographics:

Demographic data are displayed in Table 2. The degree of co-morbidity was much higher in the vertebral (45%) and hip fracture (63%) groups, compared to (9%) in both the wrist and HRT groups. The educational levels were similar in all four groups which is relevant, since earlier studies have demonstrated that educational level may have an effect on an individual's preference values (67).

5.3 Reliability of the Feeling Thermometer and Standard Gamble:

Table 3 shows the reliability coefficients (ratio of the variance between patients to the error variance using a one-way ANOVA). The reliability coefficients ranged from 0.65 - 0.87 for the different markers states for the feeling thermometer, and for the standard gamble the coefficient was 0.83. Except for the vertebral fracture state, all other reliability coefficients exceeded the a priori value of 0.70. The reliability coefficient for the current health on the feeling thermometer (0.83) was identical to the reliability coefficient for the standard gamble (0.83).

These values were comparable to those found in previous studies (50).

5.4 Comparison of Feeling Thermometer and Standard Gamble Preferences :

As shown in Table 4 and Figure 2, the baseline preferences for the patient's current health using the feeling thermometer demonstrated a gradient according to the group assessed, with highest preferences for HRT group and lowest for the hip fracture subjects. The HUI preferences for current health were consistent with the preferences from the feeling thermometer, except that for the HRT group, preferences were lower when elicited by the HUI.

The mean (and median) standard gamble preferences were higher than both the HUI and feeling thermometer preferences throughout the wrist, vertebral and hip fracture groups. The standard gamble preferences were particularly high in the hip fracture group (0.91-mean, 0.97 median) and were similar to the preference measurements for the HRT group. The pattern of the follow-up preference scores 2-3 months later were similar to the baseline preferences (Table 5 and Figure 3).

5.5 Valuation of Different Marker States:

In Table 6 and Figure 4, the five different marker states for osteoporosis using the feeling thermometer are presented. The preferences for the marker states decreased in value according to the severity of the marker condition.

The mean values for the different marker states at baseline were similar between the different subgroups; mild (wrist fracture) ranged from 0.86-0.91; moderate (vertebral fracture) from 0.50-0.61 and 0.30 - 0.37 for the severe marker state (hip fracture), which demonstrates consistency between the four groups (Table 7). There was also little difference between the marker states at baseline and follow-up confirming reliability (Table 7). The ranking of the marker states was also consistent with respect to severity. When comparing the values for current health in Table

4 to those obtained for the marker states, the values were comparable for wrist fracture (0.84 = current health); (0.86-0.92 = marker state for wrist fracture). In contrast the preferences for the vertebral (0.49-0.61) and hip fracture (0.30-0.40) scenarios were lower than preferences elicited for current health for the vertebral and hip fracture subgroups (Table 4).

5.6 Difference in the Preferences for the Marker States as Derived from the Feeling Thermometer:

The differences among the five marker states were all statistically significant $p < 0.001$. (Table 8). When analysing the values ($n = 42$) for all the subjects the difference between marker states was largest for the wrist fracture compared to hip fracture state. The difference for the vertebral fracture compared to hip fracture state was smaller. When comparing scenarios that had side-effects secondary to hormone replacement compared to the scenarios that did not describe side-effects the mean difference was 0.11 for the wrist fracture scenario, and 0.07 for the hip fracture scenario.

5.7 SF-36 Scores for Different Groups:

Baseline SF-36 scores were calculated for the four different groups and are shown in Table 9. Overall, there was a trend for the scores to be lower in the hip fracture group, except on the domains of general health, emotional role, and mental health where the hip fracture group rated themselves higher than the HRT group (Figure 5).

5.8 Construct Validity of Preference Assessment.

The results of preferences for current health agree with the a priori axioms that hip and vertebral fracture preferences should be lower than that of wrist or non-fracture patients, at least for the feeling thermometer method.

Table 10 outlines the Spearman correlation values between the different preference measures and the SF-36 domains. Both the feeling thermometer and the HUI correlated with the same domains of the SF-36: Physical functioning, Social functioning, Role physical, Body pain and the Physical health summary. The feeling thermometer correlated significantly with the HUI values, $r=0.38$, $p < 0.05$. The standard gamble, however correlated more strongly with the General health dimension of the SF-36, $r=0.59$ ($p=0.0001$) as well as with Role-emotional, Role-social and Mental health summary. The feeling thermometer correlated with the HUI but not with the standard gamble. The feeling thermometer and standard gamble did not correlate with the sensation, emotional or cognition attributes of the HUI (Table 11).

5.9 Sensitivity to Change:

The results of the efficiency outcomes are displayed in Table 12. The larger the value the greater the sensitivity to change of the instrument. The efficiency score for the feeling thermometer was ($E=1.11$), for the HUI ($E=0.82$) and ($E=0.80$) for the standard gamble. The subjective impression that each individual had improved (according to the subject) was used as the criterion or standard of change. (Table 12).

Sensitivity to change was also tested by the Spearman rank correlation coefficients between the changes in preferences (follow-up minus baseline) and the changes in other outcome measures. Table 13 outlines the correlation in changes in the preferences with the changes in SF-

36 domains and other health-related quality of life measures.

Changes in the current health - feeling thermometer correlated significantly with changes in the HUI, and one domain of the SF-36.

5.10 Transformation of Feeling Thermometer Scores into Standard Gamble Scores:

When comparing the transformed results (feeling thermometer scores to standard gamble scores) to the actual standard gamble scores, only 14 out of 42 scores were within 0.05 of each other (group value for Beta coefficient = 1.21). Our coefficient was lower than that obtained by Torrance but higher than other studies.

The individual values for the beta coefficient were also calculated and when analysing these scores by group; 54.5% of the HRT group, 72.7% of the wrist fracture, 70% of the vertebral fracture and 90% of the hip fracture group were risk-averse.

6. DISCUSSION:

This study was designed to elicit preferences for marker states associated with osteoporosis and to assess the psychometric properties of patient preferences for current health in subjects initiating hormone replacement and patients who recently experienced an osteoporotic-related fracture. The marker health states provide clinical reference points that assist in interpretation of the preference scores.

It is important to include preferences in the effectiveness measure of cost-effectiveness evaluations of therapeutic interventions for osteoporosis. Although there has been some preliminary work on preferences in the area of osteoporosis, the validity and responsiveness of these instruments has not been confirmed. The methods we used to develop the marker states paid special attention to content and clinical validity of the descriptions of the marker states. A comprehensive system of levels of functioning was used in the description of the marker states, along with a description of side-effects. Most of the previous studies located, except for a recent study on knee replacement (47) have neglected to involve physicians, nurses and patients in the development of health state descriptions. I evaluated reliability, construct validity, and sensitivity to change of the preferences elicited via the standard gamble, and the feeling thermometer.

6.1 Reliability: The test-retest reliability of the marker states, and current health using the feeling thermometer and standard gamble in patients with osteoporotic-related fractures was acceptable and comparable to those of other studies. Moreover the preferences of the marker states were stable over the 2-month period, further confirming the reliability. The current

recommendations for acceptable ICC values vary. Donner et al. suggest that $R > 0.60$ is acceptable (76). McHorney and Tarlov recommend that higher requirements are necessary if the instrument is designed to assess change in individual patient (77). In this study, five out of seven values exceeded 0.75 which is similar to values in previous studies (50). Higher coefficients perhaps could have been obtained by decreasing the time between testing, but then this would increase the risk of recall bias. The lower reliability scores may reflect either instability in the instruments used or instability in the health of the subject, given that we sampled subjects who had just experienced a fracture.

6.2 Comparison of the Preferences Elicited by the Feeling Thermometer and Standard Gamble:

The values elicited using the feeling thermometer demonstrated a gradient according to group with the highest values in the HRT group, which demonstrates content validity. The standard gamble values appeared to be higher than the feeling thermometer values in the fracture subgroups. This has been noted in previous studies and the difference between preferences elicited by the standard gamble versus those elicited from the feeling thermometer may be quite large (60). In this study, the women may have regarded their health state as temporary and therefore tended to rate their current health high. This finding may be explained by the risk-averse concept outlined by Kahneman and Tversky (62). Bakker et al. (44) have demonstrated that the standard gamble is internally inconsistent since the results are susceptible to the characteristics of the worst outcome of the gamble. This finding violates the axiom of expected utility theory that the outcomes of the gamble should not influence a patient's preference for a particular health state.

The values obtained for current health using the standard gamble for the vertebral and hip fracture groups were higher than the corresponding marker states. This may have been related to the fact that individuals have a tendency to overestimate their health when they enter that state. It is also conceivable that the description of the marker states may have depicted a poorer quality of life than experienced by the individuals we interviewed.

The values which we obtained for the vertebral and hip fracture sub-groups using the feeling thermometer and the HUI were comparable to those obtained by Tosteson using the automated time-trade-off (42). Tosteson obtained a mean value of 0.81 for vertebral fracture group compared to our values of 0.76 and 0.79 for the rating scale and HUI respectively. For the hip fracture group, Tosteson obtained a mean value of 0.60 compared to our values of 0.71 and 0.67 for the feeling thermometer and HUI respectively. Our HRT group values of current health were comparable to Tosteson's group on HRT : 0.92 (feeling thermometer- current health) compared to her value of 0.90 using the time trade-off technique.

I was unable to find any preference values for women with wrist fractures.

6.3 Valuation of Different Marker States:

The responses were distributed over the feeling thermometer, a phenomenon which could be attributed to *response-spreading* (55). It is difficult to determine whether this is response-spreading or a range of values reflecting the true severity of the different conditions. Kaplan and Ernst have demonstrated that the feeling thermometer provides valid results if the response continuum is made clear and when the endpoints of the scale are well-defined (56).

6.4 Differences in Preferences for the Marker States (side-effects) Elicited from the Feeling

Thermometer:

The estimates derived from the differences in the preference measurements for the marker states may give an estimate of the side effects/toxicity of the medication. For example for the wrist fracture scenario with medication side effects compared to the scenario of the wrist fracture without side effects the difference was 0.11. This concept would need to be validated further in a larger study. The differences between the different marker states was statistically significant supporting the fact that they are distinct health states.

6.5 SF-36 Scores for the Different Subgroups:

The SF-36 scores for the domains: physical function, physical role, social function, and vitality and the physical component summary scores were lower in the vertebral and hip fracture patients with the most dramatic reductions noted in physical functioning. These domains may best reflect the aspects of quality of life that are affected in the hip and vertebral fracture patients. These results are in agreement with results presented in abstract form by Prince et al. (78). The domains of general health and mental health were not lower in the fracture patients as compared to the HRT patients, and in this respect our results differed from those of Prince.

6.6 Construct Validity:

The hypotheses outlined earlier were supported as the results indicate a higher preference score for the HRT and wrist fracture subjects than for the vertebral and hip fracture subjects. Also the scenarios for the marker states which contained side-effects resulted in lower values than for the scenarios without side-effects. Convergent validity of the preferences obtained by the feeling thermometer was supported by correlations with four domains and the physical health

summary of the SF-36. I would expect to find a positive correlation with the physical function domains of the SF-36 in this population, since the primary disability is physical in nature. A correlation between the feeling thermometer and social functioning domain of the SF-36 could be explained by the fact that fracture subjects may limit their social activities due to limitations in physical function. The feeling thermometer also correlated with the HUI, which would be expected given that the feeling thermometer script and the six attributes used are derived from the HUI. In support of divergent validity, the feeling thermometer did not correlate with the mental health summary of the SF-36, the sensation or cognitive attributes of the HUI. A further assessment of construct validity would include comparison to an unrelated construct such as a neuroticism scale, which would not be expected to correlate with preferences of subjects with osteoporosis. The standard gamble values did not correlate with the HUI and correlated with only two completely different domains [general health, and social functioning] of the SF-36.

Overall, the correlations were low with some correlations greater than 0.6, but the majority falling in between 0.4-0.6. The maximum correlation between two measurements is the square root of the product of their reliability coefficients (79). For example when comparing the SF-36 and the feeling thermometer, one would expect a correlation of 0.8.

Lalonde et al. valued the psychometric properties of preferences in subjects with coronary heart disease and also found that the feeling thermometer and standard gamble seemed to assess different aspects of quality of life (80). The results of this pilot study would suggest that with respect to construct validity the feeling thermometer is superior to the standard gamble. Groome et al. reported that the time trade-off and the standard gamble correlated reasonably well, but neither correlated with the feeling thermometer (81). Bakker et al. found that the feeling

thermometer correlated more strongly with disease outcome measures than the standard gamble. Hornberger also demonstrated different results in a cost-utility analysis depending on which preference method was used (82).

If one was to select an instrument based on a theoretical basis, then the standard gamble would be the obvious choice. However, expected utility theory assumes that an individual acts rationally in order to maximize the expected utility, but fails to describe how an individual makes a decision under uncertainty. The standard gamble has the element of risk attitude which is not part of the feeling thermometer. This may explain why the results of the standard gamble did not correlate with the feeling thermometer values. The standard gamble values reflects different aspects of health status than the feeling thermometer which is supported by the fact that the two instruments are derived from different scientific paradigms. The correlation of the standard gamble with the general health dimension of the SF-36 suggests that the subjects who rated their general health poorly were more willing to risk a chance of immediate death for an opportunity of improving their current health (Table 14).

6.7 Sensitivity to Change

The feeling thermometer proved more sensitive to change than the standard gamble based on the efficiency. It is possible that the sensitivity to change was compromised by floor effects in which subjects with the worst score may deteriorate further (83). To really determine the sensitivity to change, a longer follow-up period post-fracture would be necessary, since in some patients two months may be inadequate time period to detect change. Also, the sample size may be too small to reliably assess sensitivity to change in the preferences.

Changes in the standard gamble did not correlate with the other preferences or with any of

the domains of the SF-36 suggesting that the standard gamble is not sensitive to change based on this parameter.

6.8 Transformation of Feeling Thermometer Scores into Standard Gamble Scores:

It would be ideal if a stable relationship existed between the feeling thermometer and the standard gamble, so that preferences derived from the feeling thermometer could be converted by a power curve calculation into standard gamble preferences. The results from our study were different from Torrance and did not support a stable relationship between the feeling thermometer and the standard gamble (53). The coefficient obtained for beta was higher than that quoted in previous studies, although a number of researchers have been unable to confirm the relationship developed by Torrance (54). The feeling thermometer only captures the strength of preferences whereas the standard gamble captures both the strength of preference and the risk attitude of the patient. The high percentage of risk averse subjects in each group may explain why the feeling thermometer does not give a good approximation of the standard gamble in this study. The high percentage of risk averse behaviour explains the higher standard gamble scores that were noted in the vertebral and hip fracture subjects.

The results of this study have important implications for the use of different methods to elicit preferences in subjects with osteoporosis (Table 14). Both methods are reliable based on test-retest results. Some of the older women had more difficulty comprehending the standard gamble and this may limit the feasibility of these measurements. When the major concern is to select an instrument on the basis of construct validity, then the feeling thermometer may be the preferred instrument. The standard gamble however, would be useful in decision analyses with treatments that involved a risk of death, where an individual's attitudes toward risk would be

important to incorporate into the preference measurement.

6.8 Limitations:

There are a number of limitations with this study. The interviews are time-consuming and demand a certain level of training which may not be feasible for use in clinical trials. The SF-36 and HUI can both be self-administered and it may be feasible to develop computer-based interactive (42).

The scenarios and script were developed by myself with the assistance of other clinicians, patients and experts in health economics. It would be useful to further evaluate the content validity by asking different clinicians to design health state descriptions and then arrive at a consensus on the marker states, as a group.

I did not evaluate the effect of using anchors other than perfect health or death. I also did not assess the effect of sequence presentation or framing (84).

When the overall group was separated into individual groups the sample sizes were small, so it would be also useful to interview additional women who have experienced either a hip or vertebral fracture within the past year to assess values for current health. It would also be important to elicit preferences in women one and two years after their fracture occurs, in order to document how fractures affect patient preferences over a period of time.

Another limitation was in the recruitment of patients for the hip fracture group. The majority of the women screened for this group were cognitively impaired or deaf. Three of the older women had some initial difficulty comprehending the standard gamble (SG) and this might explain why the mean value for the SG was higher in the hip fracture subgroup. A more

complete assessment of cognitive function would include administration of the mini-Mental Folstein in those subjects who were suspected of being cognitively impaired. This group also had a greater degree of co-morbidity, which could have acted as an effect modifier. Other studies have included an assessment of co-morbidity using an established co-morbidity index (42,85).

I examined test-retest reliability but did not assess inter-rater reliability, which should also be examined. Also, I would now recommend repeating the test-retest values in a cohort of women that were one year post-fracture in order to minimize the effect of inherent variations in health, that occur in the immediate post-fracture period.

In this study I did not allow for inclusion of a negative score for states worse than death which could result in an overestimation of mean scores in the state is considered to be worse than death by one or more individuals.

6.9 Potential Biases

The recruitment was hospital-based which should not pose a problem for hip and wrist fracture patients since they come to hospital for treatment, but for the vertebral fracture and HRT patients this could represent a bias. Women with vertebral fractures do not always seek medical attention, since up to 60% of vertebral fractures are asymptomatic and I could perhaps have interviewed a sicker group which may limit the external validity.

I may have also introduced bias by selecting a well, ambulatory group of elderly women with hip fractures, who were not representative of the overall population of women recovering from hip fractures and this may limit the external validity of this type of assessment in elderly women.

The method of presentation of the scenarios could have resulted in bias, in particular the

description of side-effects.

It would also be interesting to see if education about osteoporosis prior to the administration of the instruments would impact on the results.

7. CONCLUSIONS:

This study is one of the few that has developed scenarios for the feeling thermometer, based on different osteoporotic-related fractures, in order to derive preferences specific for osteoporosis. The presence of side-effects was incorporated into the scenarios. The incorporation of side-effects within the marker states may be an accurate method to assess the side-effects of a medication and the impact on quality of life.

The validity of the current health scores of the feeling thermometer and the standard gamble were evaluated in comparison to scores derived from the SF-36 and HUI.

The use of the feeling thermometer and the standard gamble to determine preference values in the area of osteoporosis was feasible and reliable. The recruitment and follow-up of patients, in particular the hip fracture patients were the two main limiting factors.

The methodology and results of this study will be of assistance in deriving sample size estimates for future studies of patient preferences.

The differences between the preferences for current health obtained with the feeling thermometer and the standard gamble and their validity testing with the SF-36 and HUI are important. The feeling thermometer, in comparison to the standard gamble appeared to have better evidence of construct validity. These results would suggest that the feeling thermometer and the standard gamble are measuring different aspects of health-related quality of life which has potential implications for their use. Neither instrument proved to be sensitive to change over

the follow-up period of 2-3 months which may limit the potential use of these instruments in clinical trials.

More extensive validity testing is necessary, particularly in the area of construct validity and sensitivity to change, before preference values can be used as a measure of quality of life in health economic evaluations or decision aids for osteoporosis. It would be important to verify the degree of measurement error, or instability of an individual's preference measurements. Preferences may be prove to be less useful for individual decision-making than for group decision-making.

Table 2 Demographics of Patients

Group	Hormone replacement	Wrist fracture	Vertebral fracture	Hip fracture	
Number approached	15	20	20	40	
Number eligible	13	19	15	15	
Sample size (number consenting)	11	11	10	10	
Age: median (range)	56.0 (45-69)	68.0 (51-80)	75.5 (65-88)	79.5 (63-91)	
Educational level (n)	> Grade 13 < Grade 13	45% 54%	45% 54%	40% 60%	50% 50%
Co-morbidity (percent)	9%	9%	45%	63%	

Table 3

Reliability Coefficients

Test-retest Reliability: Between test and retest scores two weeks apart.
n=10

Marker state-Feeling thermometer	Reliability Coefficient
Wrist fracture/side effects	0.79
Wrist fracture/no side effects	0.72
Vertebral fracture	0.65
Hip fracture	0.85
Hip fracture /no side effects	0.87
Current Health	Reliability Coefficient
Feeling Thermometer	0.83
Standard Gamble	0.83

Table 4**Baseline Preferences for Current Health according to Group:**Mean, (standard deviation), median and [Range]

Group	Feeling Thermometer	Standard Gamble	Health Utilities Index
HRT n=11	0.92 (0.08) <u>0.95</u> [0.78 -1.0]	0.90 (0.11) <u>0.95</u> [0.70 -1.0]	0.80 (0.10) <u>0.80</u> [0.62 - 0.92]
Wrist n=11	0.84 (0.11) <u>0.90</u> [0.63 -0.95]	0.87 (0.19) <u>1.00</u> [0.50 - 1.0]	0.86 (0.06) <u>0.87</u> [0.75 - 0.92]
Vertebral fracture n=10	0.76 (0.13) <u>0.75</u> [0.50 -0.95]	0.84 (0.20) <u>0.92</u> [0.50 - 1.0]	0.79 (0.22) <u>0.73</u> [0.25 - 0.92]
Hip n=10	0.71 (0.11) <u>0.72</u> [0.50 -0.85]	0.91 (0.12) <u>0.97</u> [0.75 - 1.0]	0.67 (0.12) <u>0.64</u> [0.53 -0.89]

Table 5**Follow-up Preference Scores for Current Health**Mean,(standard deviation), median and [Range]

Group	Feeling Thermometer	Standard Gamble	Health Utilities Index
HRT	0.88 (0.12) <u>0.93</u> [0.6- 1.0]	0.93 (0.07) <u>0.95</u> [0.8 - 1.0]	0.82 (0.07) <u>0.80</u> [0.69-0.92]
Wrist Fracture	0.88 (0.07) <u>0.90</u> [0.75 - 0.95]	0.91 (0.15) <u>0.98</u> [0.5 -1.0]	0.87 (0.07) <u>0.92</u> [0.70-0.95]
Vertebral Fracture	0.83 (0.08) <u>0.80</u> [0.73 - 0.97]	0.91(0.10) <u>0.95</u> [0.70 -1.0]	0.76 (0.14) <u>0.77</u> [0.43-0.92]
Hip Fracture	0.76 (0.18) <u>0.80</u> [0.40- 0.95]	0.84 (0.18) <u>0.85</u> [0.50 -1.0]	0.71 (0.09) <u>0.73</u> [0.58-0.82]

Table 6
Baseline Preference Scores for Marker States Elicited with the Feeling Thermometer

MARKER STATES - Mean preference score and range

Group	Wrist fracture without side effects	Wrist fracture with side effects	Vertebral fracture	Hip fracture without side effects	Hip fracture with side effects
HRT	0.88 (0.77-1.00)	0.76 (0.6-0.95)	0.61 (0.35-0.8)	0.47 (0.25-0.80)	0.37(0.20-0.65)
Wrist	0.86 (0.70- 0.95)	0.75 (0.55-0.90)	0.50 (0.25-0.65)	0.39 (0.15-0.55)	0.30(0.10-0.5)
Vertebrae	0.91 (0.75- 0.99)	0.80 (0.70-0.90)	0.56 (0.40-0.85)	0.42 (0.25-0.65)	0.37 (0.20-0.55)
Hip	0.86 (0.65-0.95)	0.78 (0.55-0.90)	0.5 (0.25- 0.70)	0.38 (0.20-0.55)	0.33 (0.10-0.50)

* Scale Anchors

1- perfect health
0- death

Table 7 Preference Scores for the Marker Health States at Baseline and Follow-up**Mean (SD)**

Group	Marker State		
	Wrist fracture scenario	Vertebral fracture scenario	Hip fracture scenario
HRT- Baseline	0.88 (0.07)	0.61 (0.12)	0.47 (0.15)
Follow-up	0.88 (0.09)	0.57 (0.17)	0.45 (0.16)
Wrist Baseline	0.86 (0.10)	0.49 (0.13)	0.39 (0.11)
Follow-up	0.90 (0.05)	0.54 (0.12)	0.42 (0.12)
Spine Baseline	0.91 (0.07)	0.56 (0.12)	0.42 (0.13)
Follow-up	0.92 (0.07)	0.61 (0.07)	0.46 (0.08)
Hip Baseline	0.86 (0.11)	0.50 (0.14)	0.38 (0.13)
Follow-up	0.86 (0.11)	0.54 (0.11)	0.41 (0.07)

Table 8

**Difference in Preference Scores for the Marker States
(Elicited by the Feeling Thermometer)**

MARKER STATES	Mean difference 95% confidence interval	
Wrist fracture without side-effects compared to Wrist fracture with side-effects	0.11	(0.09 - 0.13)
Wrist fracture compared to Vertebral fracture	0.25	(0.21 - 0.29)
Wrist fracture compared to Hip fracture	0.40	(0.36 - 0.45)
Vertebral fracture compared to Hip fracture scenario -	0.18	(0.13 - 0.23)
Hip fracture without side-effects compared to scenario with side-effects	0.07	(0.05 - 0.09)

Differences are all significant at $p < 0.0001$, Using T-test statistic

Table 9

Mean SF-36 Scores for each Group at Baseline

SF-36 Domain	Sub-Group				
	HRT	Wrist	Vertebrae	Hip	
Physical Functioning	85	76	51	42	
Role - Physical	70	30	25	22	
Body Pain	63	58	44	50	
General Health	72	79	72	81	
Vitality	60	60	58	55	
Social functioning	82	73	67	69	
Role Emotional	79	51	86	83	
Mental Health	76	82	79	85	
Physical Health summary	47	43	23	32	
Mental Health summary	50	51	43	58	

Table 10**Correlation Coefficients between Baseline Preferences and Health Status Outcomes**

n=42

	Feeling Thermometer - Current Health	Standard Gamble Current Health	HUI	HAQ -Pain
Feeling Thermometer	-----	0.09	0.38*	-0.36*
Standard Gamble	0.09	-----	0.15	-0.20
HUI	0.38*	0.15	-----	-0.33*
SF-36 dimensions: Physical functioning	0.56 ***	0.15	0.55***	-0.39
Role-Physical	0.44***	0.22	0.39**	-0.14
Body Pain	0.32 **	0.25	0.43***	-0.34*
General Health	0.14	0.59***	0.05	-0.40**
Vitality	0.14	0.29	0.36*	-0.18
Social functioning	0.52 ***	0.31*	0.35*	-0.29
Role - Emotional	0.02	0.29 (p=0.05)	0.08	-0.11
Mental Health	0.03	0.26	0.03	-0.01
Physical Health summary	0.60***	0.22	0.62***	-0.36*
Mental Health summary	-0.07	-0.36*	-0.09	0.004

* p<0.05

** p<0.01

*** p<0.005

Table 11
Relationship of the HUI and the SF-36 with the Feeling Thermometer and Standard Gamble

SF-36 domain	HUI - SENS	HUI - MOB	HUI - COGN	HUI - EMOT	HUI - PAIN	HUI - SELF
Physical Functioning	0.19	-0.59**	0.05	0.26	-0.3*	-0.29*
Role Physical	0.2	-0.37*	-0.05	0.05	-0.17	-0.17
Body Pain	0.40**	-0.26	0.09	0.00	-0.44**	0.04
General Health	-0.02	0.24	-0.16	-0.21	-0.16	0.18
Vitality	0.05	-0.001	-0.15	-0.20	-0.06	-0.10
Social functioning	0.37*	-0.008	-0.15	-0.05	-0.10	-0.06
Role Emotional	-0.0045	0.25	-0.14	-0.10	-0.038	0.33*
HAQ-PAIN	-0.03	0.16	-0.08	0.05	0.48**	0.24
Feeling thermometer	0.08	-0.39*	0.11	0.10	0.10	-0.34*
SG	-0.02	0.10	0.08	-0.10	-0.10	0.02

* p<0.05

** p<0.001

Table 12
Sensitivity to Change for Preferences
(Efficiency score)

Preference measurement in those subjects who changed subjectively	
Feeling Thermometer	1.11
Health Utilities Index	0.82
Standard Gamble	0.80

Efficiency =
$$\frac{[\text{Change in preference score}]}{[\text{Standard deviation of change}]}$$

Table 13 Correlation Coefficients between Change in Preferences and SF-36 Outcomes.

n=40

	Feeling Thermometer- Current Health	Standard Gamble Current Health	HUI
Feeling Thermometer	-----	0.01	0.47 **
Standard Gamble	0.01	-----	-0.01
HUI	0.47 **	-0.01	-----
SF-36 dimensions			
Physical Functioning	0.06	0.27	0.18
Role-Physical	0.24	0.19	0.24
Body Pain	0.32 *	0.12	0.35 *
General Health	0.05	0.28	0.15
Vitality	-0.01	-0.06	0.06
Social Functioning	0.04	-0.01	0.23
Role-Emotional	0.20	0.13	0.06
Mental Health	0.30 *	0.19	0.13
Physical Health summary	0.12	0.26	0.19
Mental Health summary	0.17	0.04	0.12

* p<0.05

** p<0.01

*** p<0.005

Table 14 Summary of Psychometric Properties of Feeling Thermometer and the Standard Gamble.

Psychometric Property	Instrument	
	Feeling Thermometer	Standard Gamble
Reliability	83.2%	83.0%
Feasibility	Interviewer administered, comprehension not a problem	Difficulty with comprehension of probabilistic information in older subjects
Methodological problems	Response spreading	Risk averse behaviour leads to higher values
Construct Validity	Correlated with the HUI ($r=0.38$) Correlated with four domains of the SF-36: physical functioning, role-physical, body pain, and social functioning and with the physical component summary	Did not correlate with the HUI Correlated with two dimensions of the SF-36: general health, social functioning and the mental health summary
Sensitivity to change	Relative efficiency - 1.11 Current Health	Relative efficiency - 0.80 Current Health

Figure 2: Baseline Preferences for Current Health by group

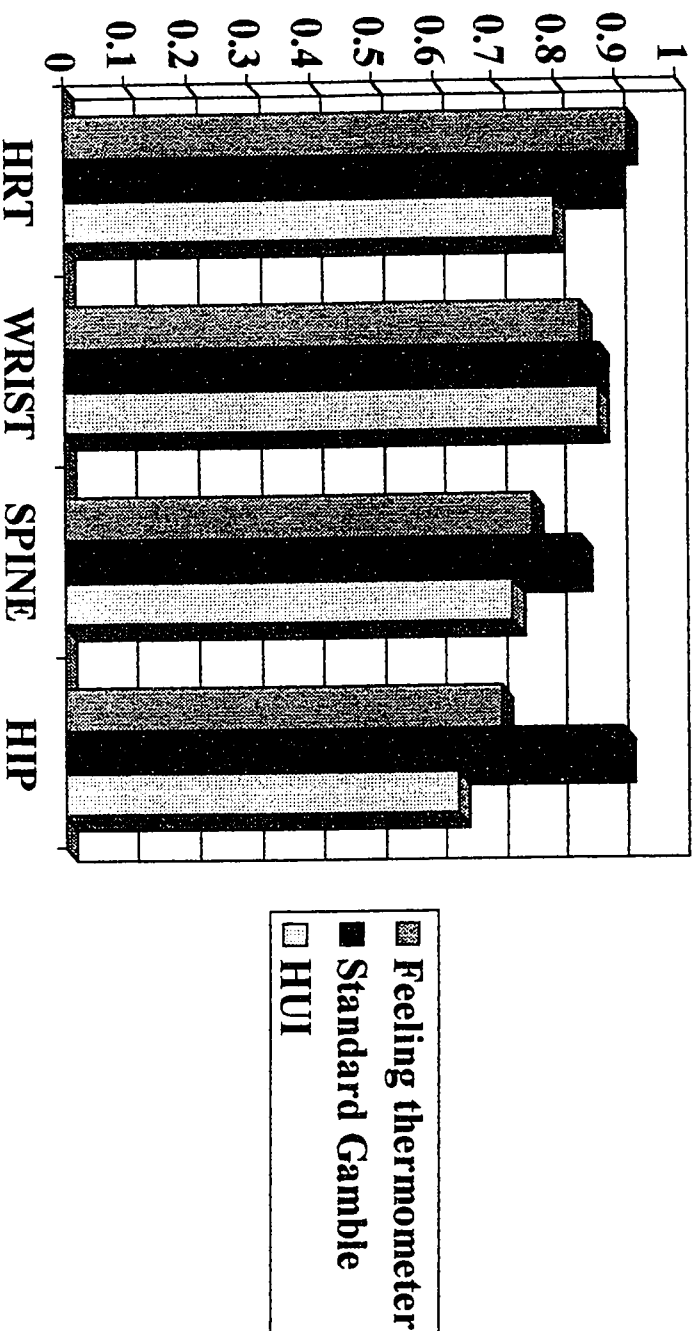


Figure 3: Follow-up preference scores for Current Health according to group

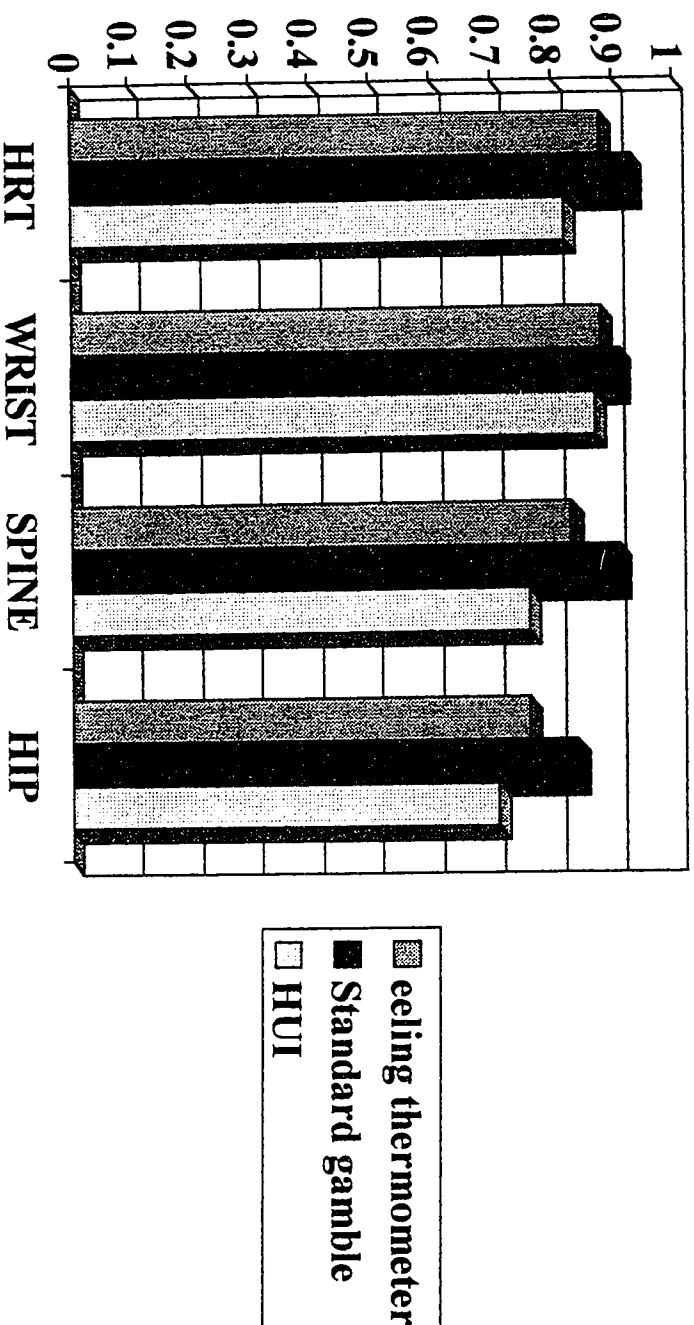


Figure 4: Marker states at
Baseline using Feeling
thermometer

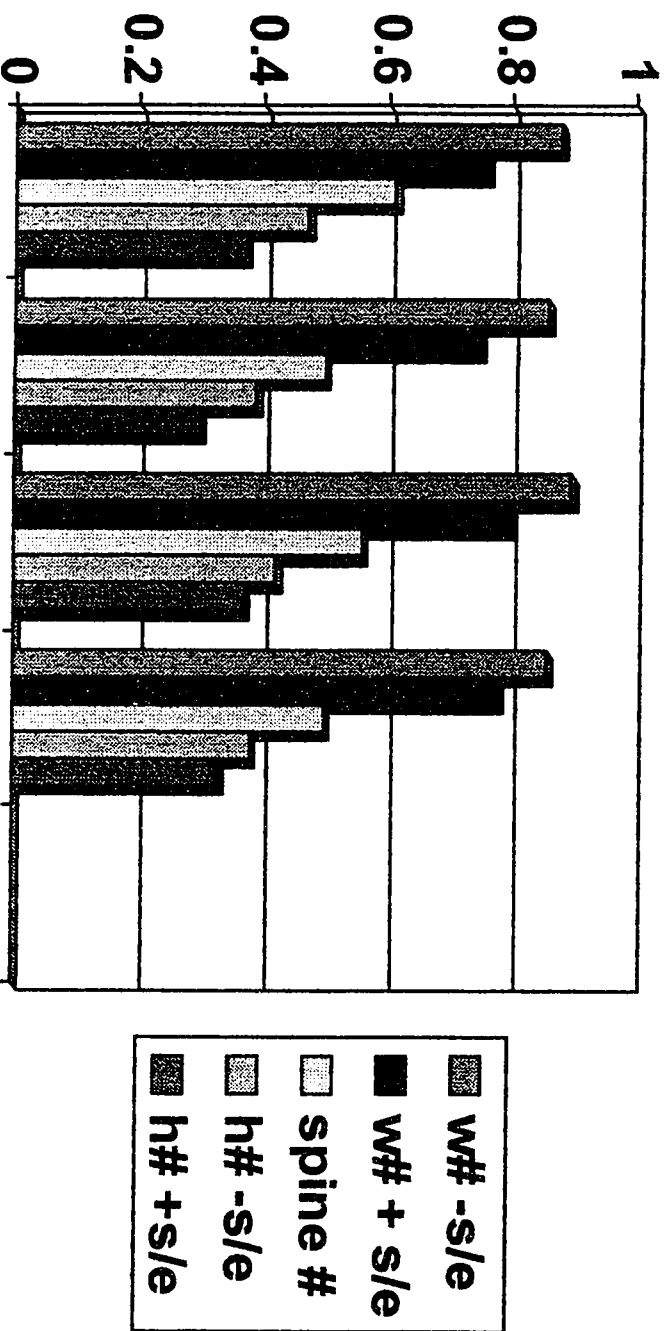
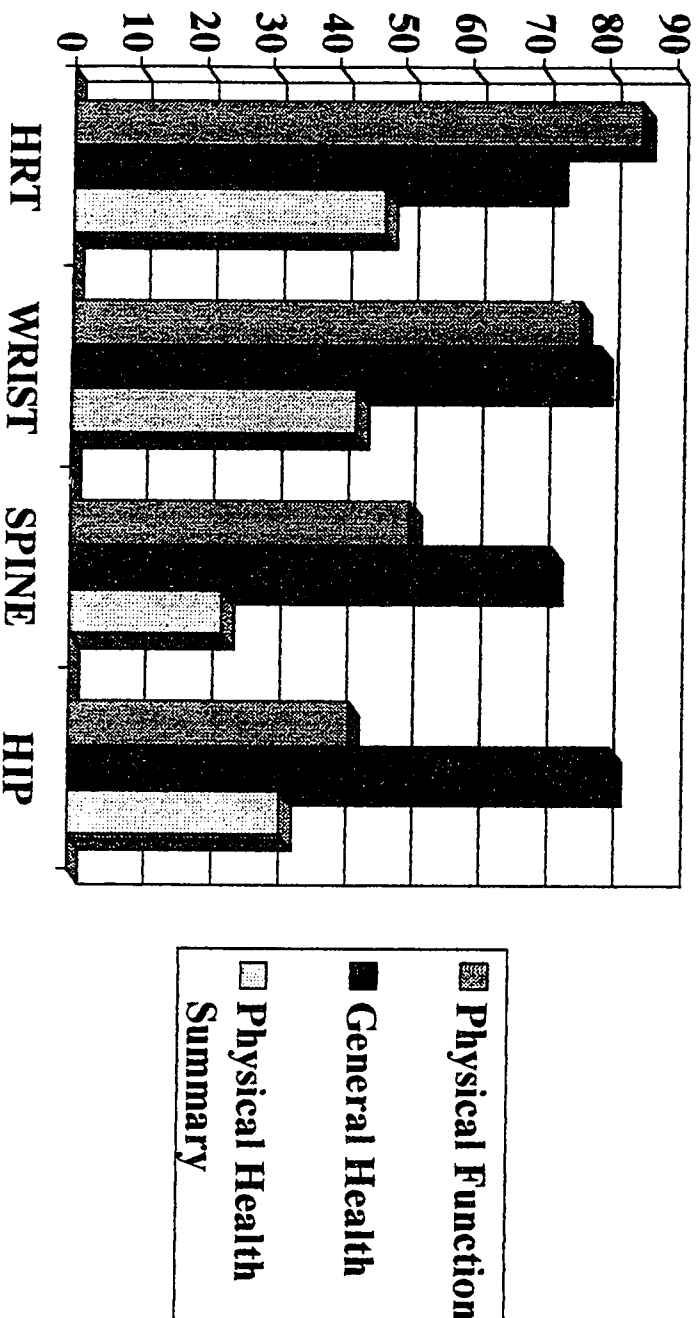


Figure 5: SF -36 scores by group



REFERENCES:

- (1) Goeree R, O'Brien B, Pettit D, Cuddy L, Ferraz M, Adachi J. An assessment of the burden of illness due to osteoporosis in Canada. *SOGC* 1996;18(7):S 15-24.
- (2) Cummings S, Black DM, O'Fallon WM, Wahner HW, Riggs BL. Bone density at various sites for prediction of hip fractures. *Lancet* 1993; 341:72-5.
- (3) Cummings S, Black DM, Rubin SM. Lifetime risks of hip, Colles' or vertebral fracture and coronary heart disease among white postmenopausal women. *Arch Intern Med* 1989;149:2445-8.
- (4) Henderson V. Estrogen, cognition, and a woman's risk of Alzheimer's disease. *Am J Med* 1997;103(3 A):11 S-8 S.
- (5) Canadian Coordinating office for health technology assessment - Guidelines for economic evaluation of pharmaceuticals: Canada. 1994; p.1-85.
- (6) Hillman AL. Economic analysis of health care technology. A report on principles. *Ann Intern Med* 1995;122:61-70.
- (7) Udvarhelyi S, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature. *Ann Intern Med* 1997;116(3):238-44.
- (8).Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. *JAMA* 1996;276:1172-7.
- (9) Siegel JE, Weinstein MC, Russell LB, Gold MR . Recommendations for reporting cost-effectiveness analyses. *JAMA* 1996;276:1339-41.
- (10) Detsky AS. Guidelines for economic analysis of pharmaceutical products: A draft document for Ontario and Canada. *Pharmacoeconomics* 1993;3(5):354-61.
- (11) Woden (ACT) Department of Health, H.A. Guidelines for the pharmaceutical industry in preparation of submissions to the pharmaceutical benefits advisory committee: including submissions involving economic analysis. 1990.
- (12) McHorney CA, Ware JE, Raczek A. The MOS 36-item short-form health survey (SF-36):II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31(3):247-63.

- (13) Silverman SL, Mason J, Greenwald M. The osteoporosis assessment questionnaire (OPAQ): a reliable and valid self-assessment measure of quality of life in osteoporosis. *J Bone Miner Res* 1997; 12(8):S 343. *J Gerontology Med Sci* 1995; 50A: M91-8.
- (14) Gold M, Patrick DL, Torrance GW, et al. Gold M, Siegel J, Russell L, Weinstein M, editors. *Cost-effectiveness in health and medicine*. New York: Oxford University Press; 1997; 4: Identifying and valuing outcomes. p. 82-134.
- (15) von Neumann J, Morgenstern O. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press; 1994.
- (16) Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J of Technology assessment in health care* 1989;55:59-75.
- (17) Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYE versus QALYs. *J Health Econ* 1993;11:325-39.
- (18) Mehrez A, Gafni A. The healthy-years equivalents: How to measure them using the standard gamble approach. *Med Decis Making* 1991;11(2):140-9.
- (19) Wakker P. A criticism of healthy-years equivalents. *Med Decis Making* 1996;16(3):207-14.
- (20) Daly E, Gray A, Barlow D, McPherson K, Roche M, and Vessey M. Measuring the impact of menopausal symptoms on quality of life. *Br Med J* 1993;307:836-40.
- (21) Geelhoed E, Harris A, Prince R. Cost-effectiveness analysis of hormone replacement therapy and lifestyle intervention for hip fracture. *Aust J Public Hlth* 1994;18(2):153-60.
- (22) Weinstein M, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstet Gynecol Survey* 1983;38(8):445-55.
- (23) Office of Technology Assessment. *Effectiveness and costs of osteoporosis screening and hormone replacement therapy*. Volume 1: Cost-effectiveness analysis. Washington, DC. 1995; OTA-BP--H-160. p.1.
- (24) Andersen L, Kristiansen IS, Falch J, Aursnes I. Cost-effectiveness of alendronate for the prevention of osteoporotic fractures in Norwegian women. Working paper #11, 1995.
- (25) Ankjaer-Jensen A and Johnell O. Prevention of osteoporosis: cost-effectiveness of different pharmaceutical treatments. *Osteoporosis Int* 1996;6:265-75.
- (26) Jonsson B, Christiansen C, Johnell O, Hedbrandt J. Cost-effectiveness of fracture prevention in established osteoporosis. *Osteoporosis Int* 1995;5(2):136-42.

- (27) Tosteson ANA, Rosenthal DI, Melton J III, Weinstein MC. Cost-effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Ann Intern Med* 1990;113(8):594-603.
- (28) Cheung AP, Wren BG. A cost-effectiveness analysis of hormone replacement therapy in the menopause. *Med J Austral* 1992;156(3):312-6.
- (29) Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: an analysis of benefits, risks and costs. *Br Med Bull* 1992;48(2):368-400.
- (30) Weinstein MC, Tosteson ANA. Cost-effectiveness of hormone replacement. *Ann NY Acad Sci* 1990;592:162-72.
- (31) Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol* 1982;9:758-62.
- (32) Guyatt G, Kirschner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;45(12):1341-5.
- (33) Bowe TR. Measuring patient preferences: rating scale versus standard gamble. *Med Decis Making* 1995;15 (3):283-5.
- (34) Hilditch JR, Lewis J, Peter A, et al. A menopause-specific quality of life questionnaire: development and psychometric properties. *Maturitas* 1996;24:161-75.
- (35) Ross PD, Wasnich RD, MacLean CJ, Hagino R, Vogel JM. A model for estimating the potential costs and savings of osteoporosis prevention strategies. *Bone* 1988;9:337-47.
- (36) Tosteson ANA, Houpt L, Kneeland TS, et al. Longitudinal analysis of utility data from a randomized controlled health economics trial. *Med Decis Making* 1997;17(4):540.
- (37) Tosteson ANA, Gabriel SE, Kneeland TS, et al. Impact of hormone replacement therapy (HRT) on quality of life. *Med Decis Making* 1997;17(4):532.
- (38) Hillner BE, Hollenberg JP, Pauker SG. Postmenopausal estrogens in prevention of osteoporosis: benefit virtually without risk of cardiovascular effects are considered. *Am J Med* 1986;80:1115-7.
- (39) Fordham RJ. Hip fracture rehabilitation: Economic evaluation & outcome assessment of geriatric orthopedic care (thesis). Department of Public Health: University of Western Australia, 1995.

- (40) Tosteson ANA, Kneeland TS, Nease RD, et al. Feasibility of automated utility assessment in phase III clinical trials: U-titer and a raloxifene osteoporosis prevention trial. *Med Decis Making* 1995;15:413.
- (41) Sumner W, Nease R, Littenberg B, et al. U-titer: A utility assessment tool. *J* 1992; AMIA: 701-5.
- (42) Tosteson ANA, Gabriel SE, Kneeland T, et al. Impact of fractures on quality of life. *J Bone Miner Res* 1997;12:(1)S 366.
- (43) Bakker C, van Rutten M, van Doorslaer EK, Bennett KJ, van der Linden S. Feasibility of utility assessment by rating scale and standard gamble in patients with ankylosing spondylitis or fibromyalgia. *J Rheum* 1994;21(7):269-74.
- (44) van Rutten M, Bakker C, van Doorslaer EK, van der Linden S. Methodological issues of patients' utility measurement. *Med Care* 1997;33:922-37.
- (45) Zug KA, Littenberg B, Baughman RD, et al. Assessing the preferences of patients with psoriasis. *Arch Dermatol* 1995;131:561-8.
- (46) Churchill DN, Torrance GW, Taylor DW. Measurement of quality of life in end-stage renal disease: the time trade-off approach. *Clin Invest Med* 1987;10(1):14-20.
- (47) Bennett KJ, Torrance GW, Moran L, Smith F, Goldsmith CH. Health state utilities in knee replacement surgery: the development and evaluation of Mcknee. *J Rheumatol* 1997;24(9):1796-805.
- (48) Mathis SD, Williamson CL, Colwell HH, et al. Assessing health-related quality-of-life and health state preferences with obesity. *Qual Life Res* 1997;6:311-22.
- (49) Bennett KJ, Torrance GW. Spilker B, editors. *Quality of life and pharmacoeconomics in clinical trials*. second ed. Philadelphia: Lippincott-Raven; 1996; 27, Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. p. 253-65.
- (50) Froberg DG, Kane RL. Methodology for measuring health-state preferences - II: scaling methods. *J Clin Epidemiol* 1989;42(5):459-71.
- (51) Torrance GW. The measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5(1):1-30.
- (52) Health services research group. Studying patients' preferences in health care decision making. *CMAJ* 1992;147(6):859-76.

- (53) Torrance GW, Feeny D, Furlong W, Barr RD, Zhang Y, Wang Q. Multi-attribute utility function for a comprehensive health status classification system. *Med Care* 1996;34(7):702-72.
- (54) Bleichrodt H, Johannesson M. An experimental test of a theoretical foundation for rating-scale valuations. *Med Decis Making* 1997;17(2):208-16.
- (55) Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein M. Preferences for health outcomes. *Med Decis Making* 1984;4(3):314-28.
- (56) Kaplan RM, Ernst JA. Do category rating scales produce biased preference weights for a health index? *Medical Care* 1983;21(2):193-207.
- (57) Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions - Health Utilities Index. *Pharmacoeconomics* 1995;7(6):503-20.
- (58) Anonymous EuroQol: a new facility for the measurement of health-related quality of life. EuroQol Group. *Health Policy* 1990;16(3):199-208.
- (59) Hurst NP, Jobanputra P, Hunter M et al. Validity of the Euroqol - A generic health status instrument - in patients with rheumatoid arthritis: Economic and health outcomes research group. *Br J Rheumatol* 1994;33:655-62.
- (60) Torrance GW. Social preferences for the health states. An empirical evaluation of three measurement techniques. *Socioecon Plann Sci* 1976;10:129-36.
- (61) Bass EB, Wills S, Scott IU, et al. Preference values for visual states in patients planning to undergo cataract surgery. *Med Decis Making* 1997;17(3):324-30.
- (62) Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 1979;47(2):263-91.
- (63) Hellinger FJ. Expected utility theory and risky choices with health outcomes. *Med Care* 1989;27(3):273-9.
- (64) Froberg DG, Kane RL. Methodology for measuring health state preferences - III: population and context effects. *J Clin Epidemiol* 1989;42(6):585-92.
- (65) Llewellyn Thomas H, Sutherland HJ, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? *Med Care* 1993;31(11):1002-12.
- (66) Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. The primer on medical decision analysis: part 3 - estimating probabilities and utilities. *Med Decis Making* 1997;17(2):136-41.

- (67) Gabriel SE, Campion ME, O'Fallon M. Patient preferences for non-steroidal anti-inflammatory drug related gastrointestinal complications and their prophylaxis. *J Rheumatol* 1993;20(2):358-61.
- (68) Ware J, Spilker B, editors. *Quality of life and pharmacoeconomics in clinical trials*. 2nd ed. Philadelphia: Lippincott-Raven; 1996; 34, The SF-36 health survey. p. 337-45.
- (69) Fries JF, Spitz, Young DY. The dimension of health outcomes: The health assessment questionnaire. *J Rheumatol* 1982;9:789-93.
- (70) Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Statistics and strategies for evaluation. Control Clin Trials* 1991; 121: 42S-58S.
- (71) Bartko JJ. General methodology II measures of agreement: A single procedure. *Stat in Med* 1994;13:737-45.
- (72) Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat in Med* 1994;13:2465-76.
- (73) Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1967;86:420-28.
- (74) Streiner DL, Norman G. *Health Measurement Scales*. 2nd ed. 1995; 8: Reliability. p. 104-27.
- (75) Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 1993;20:535-7.
- (76) Donner AP, Eliasziw M. Sample size requirements for reliability studies. *Stat Med* 1987; 6:441-8.
- (77) McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293-307.
- (78) Prince RL, Criddle T, Comito T, et al. The effects of vertebral fracture on the quality of life and physical function: a case control study. [Abstract] *J Bone Miner Res* 1997;12:(1)S 356.
- (79) McDowell I, Newell C. *Measuring Health*. second ed. New York: Oxford University Press; 1996; 2, The theoretical and technical foundations of health measurement. p. 10-46.
- (80) Lalonde L. Psychometric properties of valuations quality of life measures in coronary heart disease prevention and treatment. *Med Decis Making* 1997;17:(4)517.

(81) Groome PA, Hutchinson TA, Tousignant P. Comparison of utility assessment methods: visual analogue method yields unacceptable results. *Med Decis Making* 1993;13:387.

(82) Hornberger J, Redelmeier D, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *J Clin Epidemiol* 1992;45:505-12.

(83) Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-29.

84. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. *Med Decis Making* 1982;2(4):449-62.

85. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613-9.



HÔPITAL GÉNÉRAL D'OTTAWA OTTAWA GENERAL HOSPITAL

PATIENT INFORMATION AND CONSENT FORM

**Study: ASSESSMENT OF PATIENT PREFERENCES AND QUALITY OF LIFE IN
OSTEOPOROSIS PATIENTS.**

Investigator: A. Cranney

Phone: 737-8983

I would like to ask you to participate in a study which will assess the quality of life/preferences in women with osteoporosis. These questionnaires can be used to assign a number to a patient's overall health.

The results of this research will help us demonstrate if patient preferences can be used to help patients in deciding on treatment options for osteoporosis.

The questionnaires/interview will be conducted on two occasions: during the initial visit and 2 months later. This should take approximately 45-60 minutes on each occasion.

Quality of Life Measures:

(Short Form - 36):

This questionnaire measures health in general and measures three major health attributes: functional status, well being, and overall evaluation of health. It contains 36 items and requires about 5-10 minutes to complete.

MENQoL:

This is a quality of life questionnaire which assesses the severity of menopausal symptoms.

Health Utilities Index:

This questionnaire was developed to provide a comprehensive description of health status. Responses are required to questions related to day to day health. The questionnaire measures a number of attributes: including vision, hearing, speech, mobility, pain and emotion. This questionnaire is self-administered and takes less than 10 minutes to complete.

Patient Preferences:

Both of these will be administered by an interviewer.

Feeling Thermometer:

This is used to rank different states of health from 0 to 100, with 0 being the worst possible health and 100 the best possible health a person could have. We will be asking you to place cards describing different health situations related to osteoporosis on the Feeling Thermometer depending where you think they belong on a scale of 0-100. You will also rate your own health.

Standard Gamble:

With this instrument, patients are asked to compare different health states in another way. You will be offered the opportunity to taking an imaginary treatment which will offer you a chance of perfect Health and Death or stay in your current health state. We will be able to arrive at a point where you cannot choose between current health state and the treatment.

Additional Information

If you have any questions about this study you can contact Ann Cranney (737-8983). If you have questions about your rights as a research subject, you should contact the Research Ethics Committee, Ottawa General (737-8392). This protocol has been approved by the Research Ethics Committee of the Ottawa General Hospital and the Ottawa Civic Hospital. The Research Ethics Committee considers the ethical aspects of all hospital research projects using human subjects.

Benefits:

There are no benefits to participation.

Voluntary Participation/Withdrawal

Participation in this study is voluntary. You may choose not to participate, or withdraw from the study at a later date, without prejudice to your care at this hospital.

Consent

I understand that I have been asked to participate in a research study to evaluate patient preferences and quality of life in osteoporosis patients. Before giving my consent by signing the consent form, I have read the patient information sheet and I feel sufficiently informed.

Patient Name

Patient Signature

Witness Name

Witness Signature

Investigator Signature -A.Cranney

Date

APPENDIX B

Clinical Report Form: Patient Preferences/Quality of Life in Osteoporosis

Visit Number _____

Patient Initials _____

Interview Date _____
D M YR

AGE _____

Patient Number _____

Time to complete interview _____

History of Osteoporotic Fracture _____

Educational level _____

Household Income < \$25,000 _____; \$25-50,000 _____; >\$50,000 _____

On HRT or other osteoporosis medication _____ Co-morbidity _____

QUALITY OF LIFE SCORES:
SF 36

Baseline

Followup

Health Assessment Questionnaire Pain Scale _____

PREFERENCES:

HUI - Current Health

FEELING THERMOMETER:

Score

Perfect Health

Marker States:

Sarah

Joanne

Michelle

Monica

Angela

Current Health

STANDARD GAMBLE:

APPENDIX C

Patient preference assessment interview at Baseline

The respondent should be seated at a table with the interviewer seated to the side of the table. The props should be within easy reach.

Introduction

Good morning/afternoon Mr./ Ms _____ . I am _____ . The purpose of this interview is to understand better how osteoporosis affects the lives of people who have it. Because you have osteoporosis, we feel that you best know how it affects your life. I will be asking you about imaginary situations, situations that are not real. We can stop the interview at any time and there are no right or wrong answers.

The questions we will ask are simply meant to find out HOW YOU FEEL about different situations.

Display the Feeling Thermometer

This is called a feeling thermometer and it is used to rank different states of health from 0 to 100. 0 is the worst possible health a person could have, and 100 is the best possible health a person could have. I will be asking you to place cards describing different health situations on the Feeling thermometer depending on where you think they belong on the scale from 0 to 100.

Do you understand? YES: Go to Section 1

NO: Continue with Example

Section 1

I will now go through an example of a health state with you. If at any time there is something you do not understand, please do not hesitate to ask me.

Example A

Consider the following example to help you understand the information on the scenarios.

Hand Example Card to Respondent

The card is split into 2 sections. The top of the card describes the health state of an individual.

Point to verbal description of the state.

This indicates the situations in which people with osteoporosis might find themselves, with respect to physical functioning and mobility, self-care functions, emotional function, leisure activities, pain and discomfort.

Point to the columns on the card

This section describes the side effects that a person with osteoporosis might have from medications (hormone replacement).

These side effects include:

Point to the 7 symptoms in the first column

Bleeding, breast tenderness, mood swings, headache, weight gain, nausea, depression.

The card indicates whether the side effects have been a problem over the last 6 months and if so, how serious the problem has been.

Point to the first row specifying months 1 to 6.

If the patient found nausea to be a mild problem for them during the first month, there will be an M indicating a MILD problem in the box corresponding to month 1 and nausea. If the symptom, in this case bleeding, is severe there will be an S in the box corresponding to month 1 and bleeding.

In this example, this patient experienced mild bleeding and severe breast tenderness and mood swings in month 1. POINT In the second month the patient suffered mild weight gain and headache. POINT In the third month, the patient had mild breast tenderness and mild nausea. POINT In the fourth month, the patient had mild headaches and breast tenderness. POINT After the fourth month no side effects were a problem. Do you understand the way the health card is supposed to work.

Example - Individual 1

- Able to carry out all responsibilities at home and/or work.
- Able to carry out self-care functions (eating, bathing, dressing) without assistance.
- Rarely worried about osteoporosis and whether it will get worse.
- Unable to participate in some leisure activities.
- Occasional mild to moderate pain and discomfort.

Side Effects	1	2	3	4	5	6
Bleeding	M					
Breast Tenderness	S		M	M		
Mood swings	S					
Weight gain		M				
Headache		M		M		
Nausea			M			
Depression						

Section 2

In this section I would like you to place some health states on the Feeling Thermometer.

This card describes a person without osteoporosis or any other health problems and not on HRT. We call this perfect health.

Perfect Health

- 1- Able to carry on all activities at home/work without difficulty
- 2- Able to carry out self-care (eating, bathing, dressing) without assistance
- 3- Never worries about osteoporosis
- 4- Able to fully participate and enjoy all social activities
- 5- Free from pain

No side effects are suffered from the treatment medications, as indicated.

Months

Side Effects	1	2	3	4	5	6
Bleeding						
Breast tenderness						
Mood swings						
Weight gain						
Headache						
Nausea						
Depression						

In this case we can also see that there are no side effects, all squares in the side effect columns are **blank**.

Point to blank columns.

This is the **best** you could be, so we will put this card at the very top of the 'Feeling thermometer'. Each card has an arrow attached to it. I would like you to place the card with the arrow on the left hand side of the Feeling Thermometer pointing at a number. The top number is 100, so for Perfect Health you would place this arrow pointing at 100.

Wait for Respondent to place arrow and card (if respondent needs help, place arrows and card for them accordingly). Adjust if necessary.

This

Select death card and arrow

card represents **death**, which will be quick and painless. Remember this is imaginary.

We will put this card at the bottom of the Thermometer, at the least desirable position. The arrow should be pointing to zero.

Section 3

I would now like you to rank some different health states on the 'Feeling Thermometer'. These next few cards describe both the state of the disease and side effects that may affect people with osteoporosis. I'd like to remind you that these are imaginary situations in which people with osteoporosis might find themselves. I would also like to remind you that there are no right or wrong answers, only what you think or feel.

The first case describes Sarah's situation.

Select and hand Respondent Sarah's card. Read card slowly, aloud, with Respondent. Point to side effect columns and indicate S and M s (Sarah) suffered from mild breast tenderness, mood swings and severe bleeding in month 1.

Sarah: 52 year old female with a history of a wrist fracture 4 months ago

- 1- Able to carry out all activities at home/work without difficulty
- 2- Able to carry out all self-care functions (eating, dressing, bathing) without assistance
- 3- Rarely thinks about osteoporosis and whether a fracture will occur
- 4- Eliminates some leisure activities, due fear of another fracture
- 5- Occasional mild pain in wrist

Sarah suffers from the following side effects from the medication, as indicated.

Side Effects	Months					
	1	2	3	4	5	6
Bleeding	S	M	M	M	M	M
Breast tenderness	M	M	M			
Mood swings	M	M				
Weight gain				M	M	M
Headache						
Nausea		M	M			
Depression						

Please place Sarah on the thermometer somewhere between zero and 100 according to how good or bad you feel it would be if you had to live the rest of your life like Sarah compared to perfect health, and death.

(b) Joanne who also had a wrist fracture and suffers the identical health state as Sarah but has no side effects from medication.

Side Effects	1	2	3	4	5	6
Bleeding						
Breast tenderness						
Mood swings						
Weight gain						
Headache						
Nausea						
Depression						

Please place Joanne on the feeling thermometer somewhere between zero and 100 according to how good or bad you feel it would be if you had to live the rest of your life like Joanne compared to perfect health, death and Sarah.

MICHELLE a 75 year old woman with a previous history of a vertebral fracture.

- 1- Unable to carry out some activities at work/home.
- 2- Difficulty with some self-care functions (dressing, bathing)
- 3- Worries about falling/ sustaining a fracture intermittently
- 4- Unable to participate in a number of leisure activities
- 5- Intermittent low back pain at rest and with activity

Michelle suffers from the following side effects.

Side Effects	Months					
	1	2	3	4	5	6
Bleeding	M	M	M	M		
Breast tenderness	S	M	M	M	M	M
Mood swings	M	M				
Weight gain			M	M	M	M
Headache						
Nausea						
Depression			M	M		

Please place Michelle on the thermometer somewhere between zero and 100 according to how good or bad you feel it would be if you had to live the rest of your life like Michelle compared to perfect health, death and Sarah and Joanne.

MONICA a 78 year old woman who had a hip fracture six months ago.

- 1- Unable to carry out a number of activities at home
- 2 - Needs assistance with most self-care functions (bathing, dressing).
- 3- Worries about falling, fracture on a daily basis
Worries about inability to remain at home.
- 4- Leisure activities markedly limited by current functional level
- 5- Persistent pain in hip, on a daily basis

Monica suffers the following side-effects from the medication, at the indicated times.

Side Effects	Months					
	1	2	3	4	5	6
Bleeding	M	M		M		
Breast tenderness	S	S	M	M	M	M
Mood swings	S	M	M			
Weight gain			M	M		
Headache				M	M	
Nausea						
Depression	S	S	M	M		

Please place Monica on the thermometer somewhere between zero and 100 according to how good or bad you feel it would be if you had to live the rest of your life like Monica compared to perfect health, Sarah, Joanne, Michelle and death.

ANGELA - Hip fracture within the last six months

- 1- Unable to carry out a number of activities at home
- 2 - Needs assistance with most self-care functions (bathing, dressing).
- 3- Worries about falling, fracture on a daily basis
Worries about inability to remain at home.
- 4- Leisure activities markedly limited by current functional level
- 5- Persistent pain in hip, on a daily basis

Angela suffers the following side-effects from the medication, at the indicated times.

	Months					
Side Effects	1	2	3	4	5	6
Bleeding						
Breast tenderness						
Mood swings						
Weight gain						
Headache						
Nausea						
Depression						

Please place Angela on the thermometer somewhere between zero and 100 according to how good or bad you feel it would be if you had to live the rest of your life like Angela compared to perfect health, Sarah, Joanne, Michelle, Monica and death.

Section 4

Now I would like to ask you some questions about your current health state.

Take out Question Sheet

for each question please check the statement which best describes your situation. If there is anything on the cards that you do not understand, please do not hesitate to ask me.

If no immediate questions go to Question (6) and explain columns before having Respondent complete questions.

1. PHYSICAL FUNCTIONING AND MOBILITY

- 1) able to carry out all responsibilities at home and/or at work without difficulty
- 2) able to carry out all responsibilities at home and /or work, but with some difficulty
- 3) unable to carry out SOME responsibilities at home and/or work
- 4) unable to carry out MANY responsibilities at home and/or at work
- 5) unable to carry out any responsibilities at home and/or at work

2. SELF-CARE FUNCTIONS (EATING, BATHING & DRESSING)

- 1) able to carry out all self-care functions without assistance
- 2) able to carry out all self-care functions without assistance, but has difficulty in the morning
- 3) unable to carry out some self-care functions without assistance
- 4) unable to carry out MOST self-care functions without assistance
- 5) unable to carry out any self-care functions without assistance

3. EMOTIONAL FUNCTION:

- 1) Never worried about osteoporosis and whether it will get worse
- 2) rarely worried about osteoporosis and whether and whether it will get worse
- 3) occasionally worried about osteoporosis and whether it will get worse
- 4) frequently worried about osteoporosis and whether it will get worse
- 5) always worried about osteoporosis and whether it will get worse

4. LEISURE ACTIVITIES

- 1) able to participate in all leisure activities without difficulties
- 2) able to participate in all leisure activities, but with difficulties
- 3) unable to participate in SOME of my leisure activities
- 4) unable to participate in MOST of my leisure activities
- 5) unable to participate in any of my leisure activities

5. PAIN AND DISCOMFORT

- 1) no pain or discomfort
- 2) occasional MILD to MODERATE pain and discomfort
- 3) frequent MILD to MODERATE pain and discomfort
- 4) frequent MODERATE to SEVERE pain and discomfort
- 5) persistent SEVERE pain and discomfort

This question asks you about the side effects you might have had from medications to treat osteoporosis. Did you ever suffer from any of the side effects listed in the past month? If so was the side effect severe?

Respondent should place a check on the appropriate boxes, indicating if ("Yes") the side effect was a problem, or if ("No") the side effect was not a problem. As well if ("Yes") it was severe, or ("No") it was not severe.

SIDE EFFECT QUESTION:

	<i>Did you suffer from any side effects?</i>		<i>Was Side Effect Severe</i>	
<i>Side Effects</i>	<i>YES</i>	<i>NO</i>	<i>YES</i>	<i>NO</i>
Bleeding				
Breast tenderness				
Mood swings				
Weight gain				
Headache				
Nausea				
Depression				

Section 5

This last health situation is not imaginary. This card represents you and your **current** health situation.

Consider the answers you have just given concerning the impact osteoporosis and side effects have on your life, we will use this card to represent your own health situation with respect to osteoporosis.

YOU (Your own current health situation)

Consider the impact the following symptoms currently have on your life.

- Your ability to carry out the activities you have at work and at home.
- Your ability to carry out self-care functions (eating, bathing, dressing).
- Your concern about osteoporosis and whether you will sustain another fracture.
- Your ability to engage in social activities and hobbies.
- The frequency and intensity of pain and discomfort.

Also consider whether you have had the following possible side effects in the past month. If so, also consider whether the side effect was severe.

SIDE EFFECTS

Bleeding
Breast tenderness
Mood Swings
Headache
Nausea
Depression

Place your current health state description and it's arrow on the thermometer. Remember that the distance between the arrows should represent how good or bad you think your own health situation is compared to the others.

Remember there are no right or wrong answers, only what you think.

wait for respondent to place card and arrow

Are there any changes you would like to make to the arrows on the board? You may move the arrows and cards on the thermometer if you have changed your mind. Would you like to make any changes before we write down your scores ?

Allow Respondent to change arrows if desired.

Starting from the bottom of the thermometer and working toward the top, please tell me the name on each card, and then read me the closest number it is pointing to on the thermometer.

Record Respondent's scores on Patients response Sheet.

Thank you, that ends the Feeling Thermometer section of our interview.

GLOSSARY OF TERMS:

Content Validity: Content validity refers to the completeness with which an instrument covers the important issues of the health state which it is attempting to represent.

Construct Validity: Construct validity is assessed by comparing the results of several contrasting tests of validity (including convergent, and divergent validity), with predictions from a theoretical model. An instrument with construct validity is able to discriminate between groups that have known differences.

Convergent Validity: An instrument exhibits convergent validity when it is shown to co-vary with a number of other different measures, each of which is thought to be a direct or indirect correlate for some aspect of the construct.

Cost-effectiveness analysis: In cost-effectiveness analysis the incremental costs are compared to the incremental outcomes as a ratio, measured in physical units, for example life year gained.

Cost-utility analysis: Cost-utility analysis refers to a particular form of cost-effectiveness analysis where the outcomes are measured in terms of quality-adjusted life years (QALY) gained.

Criterion Validity: Validity indicated by comparing the results obtained using a measurement scale with a gold standard or indicator of the true situation.

Discount rate: Rate of discount used to convert future costs and health outcomes into equivalent present values.

Effectiveness: The extent to which therapeutic intervention achieves health improvements in a real practice settings.

Efficacy: The extent to which the therapeutic intervention achieves health improvements under ideal circumstances.

Face Validity: A judgement of the reasonableness of a measurement or model based on its examination by persons with expertise in the health problem and intervention being measured.

Feeling Thermometer: An instrument which is used to elicit preferences that reflect the severity of disease along an interval scale.

Healthy-Years Equivalent (HYE): The number of years of perfect health that has the same utility as the lifetime path of health states under consideration. It can be measured by a two stage

standard gamble questions or by one time trade-off.

Intra-class correlation: The intra-class correlation coefficient is a measure of agreement that records the average similarity of raters' actual scores on the ratings being compared.

Preference: Preference refers to the desirability of a health outcome. Both utility and value are special cases of the concept of preference.

QALY -Quality-adjusted life year: An outcome measure used in cost-utility analysis to assist in comparisons among programs; expressed as a monetary cost per unit of outcome. QALYs provide weights for various health states based on their relative desirability. QALYs combine changes in quantity (period of time) and quality of life (mortality and morbidity) into one composite measure or weight (0 to 1) where 1 corresponds to optimal health.

Reliability: The reliability often refers to the stability of a measurement and this is influenced by the internal consistency of the method i.e. how far the all the questions measure the same theme. The proportion of variance in a measurement that is not error variance is the reliability. Reliability can be assessed in different ways, each of which differs in the definition of error variance.

Response Spreading: This phenomenon occurs with the feeling thermometer, where there is a desire to use all categories of the scale or thermometer even if the true values are concentrated at one end of the thermometer.

Risk Averse behaviour: A person is risk averse if he prefers the certain prospect to any risky prospect with expected value x . In expected utility theory, risk aversion is equivalent to the concavity of the utility function.

Sensitivity to Change: This term refers to the ability of an instrument to detect clinically important change over time.

Sensitivity Analysis: A process through which the robustness of an economic model is assessed by examining the changes in the results of the analysis when key variables are varied over a specified range.

Societal perspective: A viewpoint for conducting a cost-effectiveness analysis that incorporates all costs and all health effects regardless of who incurs the costs and who obtains the effects.

Standard gamble: Subjects are asked to compare life in a particular health state that is a sure thing to a gamble with a probability p that perfect health is the outcome and $1-p$ that immediate death is the outcome. The probability is varied until the preference for the sure thing, the certainty of the particular health state is equal to the preference for the gamble.

Test-retest reliability: The stability or reproducibility of a measurement is evaluated in terms of the agreement between a measurement applied to a sample of people and the same measurement repeated later (typically one to two weeks later). Any changes in the scores can be regarded as error variance, and hence the level of agreement is used as an indicator of reliability.

Time trade-off: A method of measuring health state utilities in which patients are asked to trade-off life years in a state of less than perfect health for a shorter life span in a state of perfect health. The ratio of the number of years of perfect health that is equivalent to a longer life span in a less-than perfect health provides a measure of the preferences for that health state.

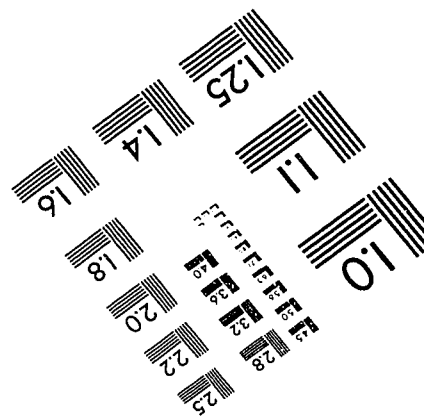
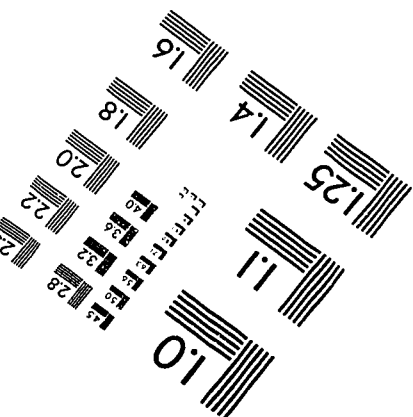
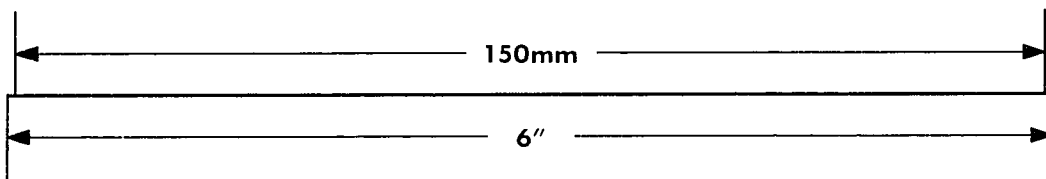
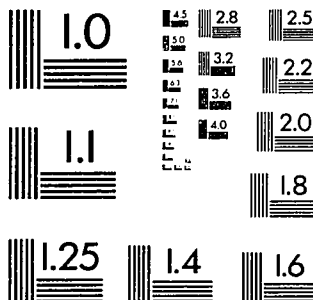
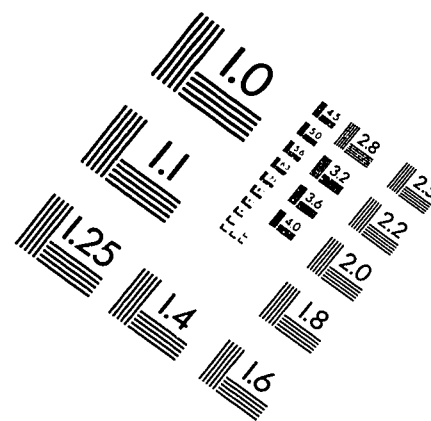
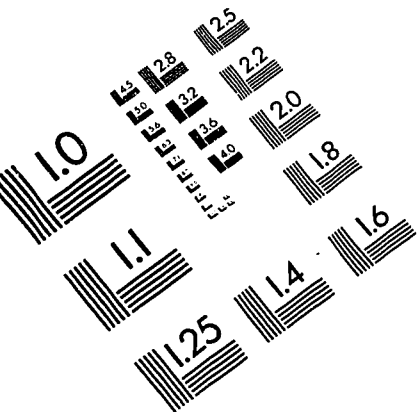
Visual Analogue scale: A format for a measurement scale in which a subject places a mark at a point on a 10 cm line that indicates the intensity of his response.

Utility: Utility is the cardinal measure of the desirability of a specific level of health status or a specific health outcome, measured under uncertainty. (e.g. Standard gamble)

Validity: The extent to which a technique measures what it is intended to measure.

Value: A cardinal measure of the desirability of, a specific level of health status or a specific health outcome, measured under certainty.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
 1653 East Main Street
 Rochester, NY 14609 USA
 Phone: 716/482-0300
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved