Essays on Economic Evaluation in Health Care

Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations

by

Niklas Zethraeus

A Dissertation for the
Degree of Doctor of Philosophy, Ph.D.

Department of Economics
Stockholm School of Economics
1998

STOCKHOLM SCHOOL
OF ECONOMICS
THE ECONOMIC RESEARCH INSTITUTE

ISBN 91-7258-466-1

Stockholm 1998

Acknowledgements

This thesis has benefited from the contributions of many people. First of all I would like to express my gratitude to my main advisor Bengt Jönsson who has spent a lot of time reading and discussing the different essays and for providing adequate comments that helped me focus on the important issues. I am also grateful to Per-Olov Johansson for contributing with constructive comments and suggestions on several papers. I would also like to thank Magnus Johannesson for always taking his time discussing and commenting upon different versions of my papers. I also want to thank Magnus Tambour, who is the co-author of two of the papers, for valuable comments on different versions of my papers and for being a good discussion partner and friend. I am also grateful to Ulf-G Gerdtham (co-author of one of the papers), Ingemar Eckerlund, Freddie Henriksson, Göran Karlsson, Clas Rehnberg and other members of the Health Economic Seminars at the Stockholm School of Economics (SSE) for commenting upon different versions of my papers. I also want to thank other colleagues at the Department of Economics and Department of Economic Statistics for making the stay at the SSE both academically stimulating and enjoyable. Further, I would like to thank Roland Strand and Peter Henriksson (co-authors of one of the papers) for a fruitful collaboration and for helping and assisting me with the data collection at Södertälje Hospital. Carin Blanksvärd has always provided friendly assistance on various practical matters. I am also grateful to Britt-Marie Eisler, Pirjo Furtenbach, and Kerstin Niklasson for providing administrative assistance whenever needed. Financial support from the National Corporation of Swedish Pharmacies is gratefully acknowledged.

Finally, I would like to express my warmest thanks to Susanne for always supporting and encouraging me at every stage of the writing of the thesis.

Stockholm, February 1998

Niklas Zethraeus

List of papers

1. **Willingness to Pay for Hormone Replacement Therapy**
   (forthcoming in *Health Economics*)

2. **The Impact of Hormone Replacement Therapy on Quality of Life and Willingness to Pay**
   Co-authors: Magnus Johannesson, Peter Henriksson and Roland T. Strand
   (Published in *British Journal of Obstetrics and Gynaecology* 1997; 104: 1191-1195)

3. **Estimating Hip Fracture Costs and Potential Savings**
   Co-author: Ulf-G Gerdtham (forthcoming in *International Journal of Technology Assessment in Health Care*)

4. **A Computer Model to Analyse the Cost-Effectiveness of Hormone Replacement Therapy**
   Co-authors: Magnus Johannesson and Bengt Jönsson (submitted for publication)

5. **Bootstrap Confidence Intervals for Cost-Effectiveness Ratios: Some Simulation Results**
   Co-author: Magnus Tambour (forthcoming in *Health Economics*)

6. **Non-Parametric Willingness to Pay Measures and Confidence Statements**
   Co-author: Magnus Tambour (forthcoming in *Medical Decision Making*)

Contents

Chapter 1

Page

Introduction

Chapter 2

Willingness to Pay for Hormone Replacement Therapy

Chapter 3

The Impact of Hormone Replacement Therapy on Quality of Life and Willingness to Pay

Chapter 4

Page

Estimating Hip Fracture Costs and Potential Savings

Chapter 5

A Computer Model to Analyse the Cost-Effectiveness of Hormone Replacement Therapy

Chapter 6

Page

Bootstrap Confidence Intervals for Cost-Effectiveness Ratios: Some Simulation Results

Chapter 7

Non-Parametric Willingness to Pay Measures and Confidence Statements

Chapter 1

Introduction

This thesis contains six papers closely related to current research topics in the field of economic evaluation in health care. The thesis discusses methodological features of cost-effectiveness analysis (CEA) and cost-benefit analysis (CBA). It further relates to the issue of modelling and how to account for uncertainty in economic evaluations. The thesis contributes both with an analysis of the costs and benefits of hormone replacement therapy (HRT) and new approaches for analysing uncertainty in economic evaluations.

The first part of the thesis (*Papers 1-4*) analyses the costs and benefits of HRT in the prevention and treatment of postmenopausal women's health problems. Besides increasing the womens' quality of life by mitigating or eliminating menopausal symptoms, HRT may offer protection against coronary heart disease and osteoporosis and related fractures. Evidence of the effect HRT has on breast cancer is inconclusive; the risk is assumed to increase after a long period of treatment. For non-hysterectomised women taking oestrogens only, an increased risk of endometrial cancer is evident. The increased risk of endometrial cancer is decreased or eliminated by the addition of a progestogen. Combining oestrogen with a progestogen may induce uterine bleedings. Such bleedings may reduce or vanish if a combined HRT is continuously applied, although break-through bleeding often occurs in the first few months (SBU 1996).

At present in Sweden about 180,000 women are treated for menopausal symptoms with HRT, implying a yearly drug cost of about SEK 200 million (SBU 1996). A universal treatment of all women potentially eligible for HRT would then cost around SEK 1.7 billion annually only in terms of drug costs. From a societal perspective the question can be raised whether it is good value for money to treat all, both symptomatic and asymptomatic women, with HRT given that the health care resources are scarce and can be used in different ways in or outside the health care system. Further, if not everyone should be given HRT which patients, if any, should be treated? These and related issues may be considered by using economic evaluations

which means that alternative courses of action are compared in terms of both their costs and their benefits.

The second part of the thesis (*Paper 5* and *Paper 6*) analyses issues of uncertainty in economic evaluations. *Paper 5* is closely related to the current discussion of how to compute confidence intervals for cost-effectiveness ratios. This discussion may be explained by a move toward economic evaluations being carried out alongside clinical trials where patient specific data on costs and benefits are available. The basic question behind this discussion is what inferences can be made for the true unobserved cost-effectiveness ratio based on the observed sample ratio. Different methods have been proposed for constructing confidence intervals for cost-effectiveness ratios (O'Brien 1994, Wakker and Klaassen 1995, Willan and O'Brien 1996). However, until today there is no consensus on which (if any) of these methods to use. One recently suggested approach is to use non-parametric bootstrap methods (Drummond and O'Brien 1993, O'Brien et al. 1994). *Paper 5* uses the proposed bootstrap procedure for estimating confidence intervals for CE ratios and evaluates the performance of this method compared to the approach suggested by Wakker and Klaassen (1995). In contingent valuation studies the question of uncertainty is also an important issue when benefit measures are estimated. One methodological issue is how to obtain confidence statements for the mean WTP based on closed-ended (binary) contingent valuation questions. For the parametric approach different methods have been proposed to account for uncertainty in the mean WTP measure due to sample variation (Cooper 1994, Park et al. 1991). However, it is less clear how to obtain confidence statements for mean WTP estimates using the non-parametric approach developed by Kriström (1990). *Paper 6* addresses this problem and proposes a procedure that allows statistical testing and confidence interval estimation for the mean WTP by employing bootstrap techniques.

1. Economic Evaluation of Health Care Programmes

Economic evaluations can be divided into CBA and CEA[1]. For these two types of studies the treatment of costs is the same and the difference lies in the handling of benefits.

Cost-benefit analysis

In a CBA benefits and costs are measured in monetary units. Benefits are defined as the amount of money gainers of the programme are willing to pay to make sure that the programme is undertaken (willingness to pay (WTP))[2] and costs are defined as the compensation the losers of the programme require to accept that the project is carried out (willingness to accept (WTA)). The benefits and costs can also be expressed as compensating variation (*CV*) measures, where *CV* is defined as a benefit (WTP), if $CV > 0$ and as a cost (WTA), if $CV < 0$. A positive *CV* measure gives the maximum amount of money that can be taken away from an individual while leaving him/her just as well off as before the improvement in health, whereas a negative *CV* measure gives the minimum amount of money that must be given to the individual to compensate him/her for the loss of health (Johansson 1995)[3].

To assess whether a programme is socially profitable a social welfare function (SWF) can be used. The SWF is a function of all the individual utility levels in society and reflects the preferences of society for programmes where a higher value of the function is preferred to a lower one[4]. Using the concept of a SWF in general means that different weights are applied to the individual gains or losses. That is, when summing *CV*:s across individuals it is in general important to consider the marginal social utility of income for each individual in society. However, there are conditions under which a positive sum of *CV*:s is a sufficient condition for an increase in social welfare. Assuming a utilitarian SWF and that government has redistributed income across individuals so as to maximise social welfare this will imply, if we evaluate a small programme, that it is possible to tell whether social welfare has risen just by

[1] Economic evaluations are sometimes divided into cost-minimisation analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis (Gold et al. 1996). Cost-minimisation analysis can, however, be seen as a special case of CEA where the consequences turn out to be equal. CUA can also be seen as a special case of CEA, where the effectiveness measure is an index, which comprises changes in quantity and quality of life.
[2] WTP reflects the amount of consumption of goods and services that individuals are willing to forego.
[3] If the utility level after the change in health is held constant the WTP and WTA measures are referred to as equivalent variation (EV).
[4] For a detailed discussion of social welfare functions (SWF:s) see Boadway and Bruce (1984) and Johansson (1995).

adding up the individual *CV* measures. The sum of *CV*:s across individuals is then proportional to the change in social welfare, and a positive sum of *CV*:s is a sufficient condition for an increase in social welfare (Boadway 1974, Boadway and Bruce 1984, Johansson 1995)[5].

The measurement of WTP (and WTA) for health changes can be based either on revealed preferences[6] or on expressed preferences. In the expressed preference approach, or contingent valuation method (CVM), survey methods are used to investigate the hypothetical WTP for a good or a service. The CVM was originally developed in the environmental field to measure the value of changes in the environment, but recently a number of health care applications have appeared (Eckerlund et al. 1995, Johansson 1995).

Contingent valuation questions can be classified either as open-ended or closed-ended. In open-ended questions the maximum WTP is elicited from each respondent. Due to problems of open-ended questions, e.g., non-response and starting point bias, most studies in the environmental (and health care) field now use binary contingent valuation questions[7]. In binary questions each respondent only accepts or rejects one price (bid). By varying the price in different subsamples the proportion accepting to pay can be calculated as a function of the price. The resulting curve can be interpreted as a market demand curve for the good (treatment). The mean WTP can then be calculated as the area under this curve and the median WTP is the price where 50 percent of the respondents accept or reject to pay.

The market demand curve can be estimated based on either parametric or non-parametric methods (Johansson 1995). A parametric approach necessitates assumptions regarding functional form that usually have been assumed to follow a logistic one (Cameroon 1988, Hanemann 1984). The drawback of a parametric estimator of the mean WTP is that it is sensitive to the assumed functional specification. An alternative is instead to use the non-parametric estimator of the mean WTP developed by Kriström (1990). This estimator is robust against functional misspecification and is simple to compute.

[5] For a recent discussion of the aggregation problem see e.g. Yew-Kwang Ng (1997).
[6] The WTP estimate is then obtained from observations of actual behaviour, e.g. using wage-risk studies (Johannesson 1996).
[7] An expert panel appointed by the National Oceanic and Atmospheric Administration (NOAA 1993) in the US to assess the validity and reliability of the CVM, recommended the use of binary questions.

The CVM based on closed-ended questions is, however, not problem-free, and important issues remain to be resolved in order to establish the validity of the method. Two important problems are that experimental results indicate that hypothetical WTP overestimates real WTP, and that hypothetical WTP is insensitive to the size of the programme (often referred to as insensitivity of scope). To test the validity of the WTP estimates in a contingent valuation study, it can be investigated whether the estimated WTP increases with income and whether the WTP increases with the size (scope) of the programme. In the health field, testing for scope would mean testing whether the WTP increases with the size of the health change.

In the thesis *Papers 1* and *2* address the question of WTP for HRT. The analyses are based on the CVM and on closed-ended contingent valuation questions. The WTP measures are based on a parametric approach assuming a logistic distribution and a non-parametric approach developed by Kriström (1990).

Cost-effectiveness analysis

CEA is based on maximising health effects (benefits) subject to a cost constraint (Weinstein and Zeckhauser 1973). Costs are measured in monetary units and effects in non-monetary units such as gained life-years or quality adjusted life-years (QALYs). The most frequently used health outcome measures including quantity and quality of life are QALYs; although healthy years equivalents (HYEs) have also been proposed (Drummond et al. 1997).

CEA has often been based on a so-called decision-maker approach where the aim of the economic evaluation is to maximise whatever the decision-maker wants to maximise. According to this approach the decision-maker decides what costs and benefits to include. The decision-maker approach can be criticised on many grounds. Firstly, it lacks theoretical foundation in economic welfare theory and the included costs and benefits are arbitrarily decided by the decision-maker. Secondly, the decision-maker is usually not identified and it is not investigated what the decision-maker wants to maximise. Further, following strictly a real world budget perspective, e.g. a health care budget perspective, would likely lead to problems with sub-optimisation compared to a societal perspective (Johannesson 1995).

It has been argued that CEA should instead be based on a societal perspective, which means that all costs and benefits no matter to whom they accrue should be included in the analysis

(Gold et al. 1996, Johannesson 1996). If the purpose of CEA is to maximise social welfare given the limited resources in society, CEA should be based on a societal perspective meaning that all costs and benefits are incorporated in the analysis no matter who pays the costs or receives the benefits (Gold et al. 1996). Using welfare economics as a theoretical basis of CEA also provides guidance in controversial issues in CEA, e.g. whether to include costs resulting from an extended life, i.e. whether to include costs in added life years resulting from a health care programme and deciding what costs to include (Weinstein 1990). The most common current practice in CEA is to include future costs only for "related" illnesses and excluding future costs for "unrelated" illnesses and future non-medical expenditures. In a recent paper by Meltzer (1997) that bases the CEA on the theory of welfare economics, the conclusion is that all future costs (medical and non-medical) should be included in a CEA.

To determine which health programme (if any) to implement using CEA the price per unit of health effects must be determined, e.g. the WTP for a QALY or a life year gained. Without information about the price per unit of health effects, a CEA gives no information about whether a program should be implemented or not, unless it is sorted out as a dominated alternative[8]. Furthermore, if a fixed price is used as a decision rule, the CEA approach can be seen as a special case of CBA where the price per health effect is constant (at all levels of health change) and the same for everyone (Johannesson 1996). CEA can then be interpreted as an estimation of the cost function to produce health effects.

2. Economic Evaluation and the Need for Modelling

The reasons why there usually is a need for modelling in economic evaluation are many and can be explained by lack of data on costs and benefits of a programme as reflected in medical practice, i.e. under real world conditions prevailing once the programme is in routine use (Buxton et al. 1997). For example, modelling is used in economic evaluations in order to be able to generalise from one setting to another or to extrapolate beyond the study period in a clinical trial. The first reason for modelling is actualised for example when data on costs and benefits are collected based on a randomised clinical trial where an intervention is compared with placebo conducted under ideal circumstances. In the economic evaluation, data on costs and benefits for an intervention should reflect clinical practice and the comparator of interest is usually not placebo. Therefore some modelling is needed to make costs and benefits reflect

[8] I.e. the programme has a higher cost for a given level of health outcome compared to the alternative.

actual clinical practice. The second reason comes up e.g. in a clinical trial with a follow-up time of 5 years studying the impact of HRT on the risk of hip fractures. To estimate the cost-effectiveness of HRT risk data for hip fractures is also needed beyond the study period in the clinical trial.

Different approaches for modelling are available (Buxton et al. 1997, Drummond et al. 1997, Gold et al. 1996). One way of modelling frequently used is decision tree analysis (Berggren et al. 1996, Weinstein and Fineberger 1980). Decision tree analysis is a way to structure problems of decision-making where relevant events are identified in a tree structure that specifies the sequence of events. By calculating the costs and benefits of each chain of events, and by weighting the costs and benefits by probabilities, expected costs and benefits can be calculated for different programmes (strategies). A limitation of decision tree analysis is that it is not well suited for programmes involving risks that are ongoing over time. In those cases the tree structure may become very complex, and Markov modelling is an alternative.

Markov models[9] are useful when a decision problem involves risks that is ongoing over time, when the timing of events is important and when important events may happen more than once (Sonnenberg and Peck 1993). This kind of model is defined using a finite number of states (health states) in which an individual may be found at any time. The model assumes that all individuals in a specific state are identical and that each individual obtains the same cost or benefit irrespectively of which transitions led to that specific state, i.e. the model has no memory of prior states (Sonnenberg and Peck 1993)[10]. Markov models occur in a discrete time frame and time progresses in units of arbitrary but fixed length (e.g. one year). A transition occurs when an individual moves from one state to the next and the transitions between states are determined by transition probabilities, which determine the allocation of individuals in each cycle. A feature of the Markov model is that e.g. transition probabilities, benefits and costs may be viewed as a function of not only the state but also population characteristics such as age. In *Paper 4* a Markov model is used for analysing the cost-effectiveness of HRT. Uncertainty surrounding the long-term effects of HRT motivates the use of a Markov model, which enables simulations of the benefits and costs of HRT compared to an alternative intervention strategy (or compared to no intervention).

[9] A Markov model is a type of state transition model in which the transition probabilities depend only on the current state.

[10] When QALYs are calculated it is assumed that the quality weight is the same for a specific health state irrespectively of prior health states. Markov models can then be used to estimate QALYs.

3. Economic Evaluation and Uncertainty

Economic evaluations can be divided into deterministic or stochastic economic evaluations (Drummond et al. 1997). Deterministic evaluations tend to be based on decision analytic models and employ data from many different sources. The deterministic evaluation refers to the case when the available data on costs and benefits are presented as point estimates. This may be due to the source of the data, e.g. secondary data or that the variable may not have been sampled (e.g. discount rate). In order to account for uncertainty in deterministic evaluations sensitivity analysis can be carried out using plausible ranges for included variables. Sensitivity analysis has been criticised for the potential for selection bias for arbitrary interpretation of the results and because possible interaction between parameters is not captured (Drummond et al. 1997).

Stochastic economic evaluations can be carried out alongside clinical trials where patient-specific data on costs and benefits are available. The observations on patient-specific costs and benefits can be viewed as a random sample from a patient population. Issues to be discussed is then e.g. what statements can be made, based on the sample information, about unobservable population parameters of costs, benefits and their ratio (e.g. incremental cost-effectiveness ratios).

Recently there has been a debate referring to statistical issues in economic evaluations, e.g. how to compute confidence intervals for cost-effectiveness ratios when patient specific data on costs and benefits are available. The discussion usually refers to the situation where e.g. a new therapy is both more costly and effective than a customary therapy and the question is whether there is good value for money to replace the customary therapy. One approach to compute confidence intervals for cost-effectiveness ratios is to use parametric methods such as the *Box method*, *Delta method* (*Taylor's method*) or *Fieller's method* (O'Brien 1994, Wakker and Klaassen 1995, Willan and O'Brien 1996). The other approach suggested is to use *non-parametric bootstrap methods* which do not require the analyst to specify the sampling distribution of the ratio, which in general is unknown (Drummond and O'Brien 1993, O'Brien et al. 1994).

The bootstrap technique is a computationally intensive method for obtaining measures of accuracy in statistical estimates. The basic idea of the approach is that the observed random sample of size n is interpreted as an empirical distribution function that estimates the probability distribution of the population (Efron and Tibshirani 1993). Repeated random samples of size n with replacement are then drawn from the observed random sample (each observation in the sample has the probability of $1/n$ to be drawn) to give bootstrap re-samples, and in each of these re-samples the bootstrap replicate of the statistic of interest is calculated. The bootstrap replicates make up the empirical estimate of the statistic's sampling distribution. The simplest method of calculating bootstrap confidence intervals is then to arrange the bootstrap replicates in increasing order and then to cut off, say, 2.5 percent of the observations in each tail which gives a 95 percent, double-sided, *percentile* confidence interval. *Paper 5* in the thesis is closely related to the above discussion and raises the question of how to estimate confidence intervals for cost-effectiveness ratios based on the bootstrap technique.

In a recent paper by Polsky et al. (1997) the above four methods are compared and evaluated (see also Briggs et al. 1997, Chaudhary and Stearns 1996). The conclusion is that the Fieller and Bootstrap methods are more accurate in terms of error percentages close to the stated significance level and in terms of degree of symmetry of miscoverage compared to the other methods. However, until today there is no consensus on which (if any) of these methods to use. An alternative solution to the ratio problem has just recently been proposed (Tambour et al. 1998). By multiplying the effectiveness unit by the price per effectiveness unit both costs and benefits can be expressed in monetary units, and standard statistical techniques can be used to estimate a confidence interval for net benefits. This avoids the ratio estimation problem and explicitly recognises that the price per effectiveness unit has to be known to provide CEA with a useful decision rule.

Also in contingent valuation studies the question of uncertainty is an important issue. One methodological issue is how to obtain confidence statements for the mean WTP based on closed-ended contingent valuation questions. For the parametric approach different methods have been proposed to account for uncertainty in the mean WTP measure due to sample variation (Cooper 1994, Park et al. 1991). However, it is less clear how to obtain confidence statements for mean WTP estimates using the non-parametric approach developed by Kriström (1990). *Paper 6* in the thesis addresses this problem and proposes a procedure that

allows statistical testing and confidence interval estimation for the mean WTP by employing bootstrap techniques.

4. Summary of the Papers

Paper 1 "Willingness to Pay for Hormone Replacement Therapy" addresses the question of WTP for HRT in order to alleviate menopausal symptoms. Further, the WTP for a gained QALY is derived enabling comparisons with the costs for producing a gained QALY. It is assumed that the woman obtains utility from consumption of goods and health and that the purchase of a treatment causes a shift in the health production function during the treatment period. The empirical part is based on a prospective study where data on quality of life and WTP were collected by interviews with patients treated at the Department of Gynaecology, Södertälje Hospital. Participants in the study were one hundred and four women aged 45 to 65 years treated for menopausal symptoms for at least one month. The mean WTP for the HRT is estimated using a parametric and a non-parametric method. The mean WTP based on these two methods is similar and amounts to about SEK 40,000 per year. Further, it is shown that the mean WTP is above the mean treatment cost of HRT. Finally, the WTP per gained QALY is estimated at about SEK 120,000 and SEK 160,000 based on the rating scale (RS) and time trade-off (TTO) methods, respectively.

Paper 2 "The Impact of Hormone Replacement Therapy on Quality of Life and Willingness to Pay" addresses the question of the gain in quality of life due to HRT for women with mild and severe menopausal symptoms. The TTO, RS and the contingent valuation methods are used to measure the gain in quality of life. The empirical part is the same as in *Paper 1*. The increase in the QALY weight due to HRT for women with mild symptoms was 0.26 according to the RS method and 0.18 according to the TTO method. The mean WTP for HRT per month was SEK 2,300 for women with mild symptoms. For women with severe symptoms the QALY weight increased by 0.50 according to the RS method and by 0.42 according to the TTO method. The mean WTP for HRT per month was SEK 4,800 for women with severe symptoms. HRT leads to major improvements in quality of life for women with mild and severe symptoms both in terms of TTO, RS and WTP. For women with mild and severe menopausal symptoms the WTP for the treatment also greatly exceeds the costs, indicating that HRT is good value for money if used for women with severe or mild menopausal symptoms.

Paper 3 "Estimating Hip Fracture Costs and Potential Savings" investigates the determinants of hip fracture costs and estimates potential savings in costs if a fracture is avoided. The costs of hip fracture are comprised of direct costs in the health care system and the municipality. Data are collected for 1,080 post-menopausal women admitted from an independent residence for primary hip fracture surgery during the year of 1992 in the city of Stockholm, Sweden. It is found that hip fracture costs are significantly related to age, mortality the year after a fracture, type of fracture, costs one year before a fracture, and hospital admission. The savings in direct costs for an average woman surviving the year after a fracture amount to SEK 210,000.

Paper 4 "A Computer Model to Analyse the Cost-Effectiveness of Hormone Replacement Therapy" presents a computer model for analysing the cost-effectiveness of HRT. The paper describes the model's design, structure and data requirements; it also discusses the opportunities and data requirements for extending the model to other countries. In an empirical application the cost-effectiveness of HRT in Sweden is analysed from a societal perspective. The economical question raised is whether it is good value for money to treat asymptomatic women with HRT for preventive purposes. Generally, the cost-effectiveness ratios improves with the age at treatment onset, the size of the risk reductions and if only oestrogen is given instead of a combined therapy. If side effects during treatment are prevalent, then HRT is dominated by the no intervention alternative. On the other hand, if the quality of life during HRT is increased the CE ratios improve substantially.

Paper 5 "Bootstrap Confidence Intervals for Cost-Effectiveness Ratios - Some Simulation Results" raises the question how to construct confidence intervals for CE ratios in economic evaluation studies. A bootstrap procedure for estimating bias-corrected confidence intervals for CE ratios is presented and tested in a simulation study based on the assumptions made in a paper by Wakker and Klaassen (1995). Two variants of CE ratio bootstrap confidence intervals are tested. The first is a bootstrap analogue of the parametric method proposed by Wakker and Klaassen, which gives similar results as the parametric method. However, computing bootstrap confidence intervals directly for the CE ratio produces error percentages closer to the stated significance level compared to the Wakker and Klaassen method.

Paper 6 "Non-Parametric Willingness to Pay Measures and Confidence Statements" addresses the question of how to account for uncertainty in contingent valuation studies using a non-parametric estimator of mean WTP. The question raised is whether it is possible to make statistical testing and confidence interval estimation for the mean WTP where the estimator is based on the non-parametric approach developed by Kriström (1990). The paper presents a procedure that allows statistical testing and confidence interval estimation by employing bootstrap techniques. The method is applied on data from the contingent valuation study described in *Paper 2* where the mean WTP for HRT was investigated for women with mild and severe menopausal symptoms. A confidence interval for the mean WTP was estimated for the full sample and separately for women with mild and severe menopausal symptoms. The hypothesis was also tested whether there was a statistically significant difference in the mean WTP between the patient groups with severe and mild menopausal symptoms. One conclusion that can be drawn using the proposed method is that the mean WTP is significantly higher in the group with severe symptoms.

5. Contributions

The papers included in the thesis contribute both empirically and methodologically to the field of economic evaluation in health care. The empirical contribution is primarily in the field of economic evaluation of HRT, whereas the methodological contributions relate to the issues of how to construct confidence intervals for CE ratios and how to account for uncertainty in the mean WTP measure computed by the non-parametric approach developed by Kriström (1990).

Papers 1 and *2* focus on the value of the gain in quality of life that women suffering from menopausal symptoms receive through HRT. The RS, TTO and contingent valuation methods measure the gain in quality of life. The first paper also estimates the WTP for a gained QALY. The second paper focuses on women with mild and severe menopausal symptoms and shows substantial gains in quality of life irrespective of having mild or severe symptoms and that the gain for women with severe symptoms is greater than for women with mild symptoms. The first two papers also show the importance of carrying out empirical studies on quality of life rather than making arbitrary assumptions, which has often been the case in previous studies analysing the CE of HRT.

Paper 3 is an empirical paper focusing on the potential savings when a hip fracture is avoided and on the determinants of hip fracture costs. Potential cost savings were estimated so that they directly fitted into the model described in *Paper 4*. Most previous studies looking into the cost of hip fractures are either prevalence-based studies or restrict themselves to costs for the first complication. In this study all post-menopausal women above the age of 49 years, admitted from a private residence, who sustained a hip fracture during the year 1992 in the city of Stockholm were analysed. The paper shows that savings in costs when preventing fractures would be exaggerated if they were estimated as the costs during one year after fracture, which is explained by considerable consumption of resources without fracture.

Paper 4 contributes a computer model that allows for analysing the cost-effectiveness of HRT from a societal perspective. The model allows for including indirect costs and costs in added life years, which have never been included in previous studies. The model can be used to analyse symptomatic or asymptomatic women, women with an intact uterus or with a hysterectomy. The model is flexible and also allows for different intervention modelling and analysis of women at different ages and with different risk profiles e.g. osteoporotic women. The model needs data on costs, mortality, quality of life and risks and the first three papers of the thesis can be seen as producing input to this model. For example, *Papers 1* and *2* estimate the quality of life gain of HRT for average symptomatic women and also for women with mild and severe symptoms, which enables analyses of the cost-effectiveness of HRT for women with menopausal symptoms. The costs per gained QALY produced by the model can be compared with the WTP per gained QALY estimated in *Paper 1*. *Paper 3* estimates the potential savings if hip fractures are avoided in different ages.

Finally, *Papers 5* and *6* propose new approaches for analysing uncertainty in the field of economic evaluation. *Paper 5* shows that the bootstrap method is a promising approach for computing confidence intervals for cost-effectiveness ratios and concludes that the bootstrap method is more accurate in terms of producing error percentages close to the stated significance level compared to the Wakker and Klaassen approach (1995). *Paper 6* presents an alternative approach for analysing uncertainty in the mean WTP estimate based on the non-parametric method developed by Kriström (1990). By using bootstrap methods, confidence intervals can be estimated for the mean WTP. It is also possible to test hypotheses of differences in mean WTP between different subgroups.

6. References

Berggren U, Zethraeus N, Arvidsson D, Haglund U, Jönsson B. A cost-minimization analysis of laparoscopic versus open cholecystectomy. *American Journal of Surgery* 1996; **172**: 305-310.

Boadway RW, Bruce N. *Welfare Economics*. Oxford: Basil Blackwell, 1984.

Boadway RW. The welfare foundations of cost-benefit analysis. *The Economic Journal* 1974; **84**: 926-939.

Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics* 1997; **6**: 327-340.

Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray M. Modelling in economic evaluation: An unavoidable fact of life. *Health Economics* 1997; **6**: 217-227.

Cameron TA. A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics* 1988; **15**: 355-379.

Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Statistics in Medicine* 1996; **14**: 1447-1458.

Cooper JC. A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. *Land Economics* 1994; **70**: 111-122.

Drummond MF, O'Brien B. Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Economics* 1993; **2**: 205-212.

Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. New York: Oxford University Press, 1997.

Eckerlund I, Johannesson J, Johansson P-O, Tambour M, Zethraeus N. Value for money? A contingent valuation study of the optimal size of the Swedish health care budget. *Health Policy* 1995; **34**: 135-143.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, No. 57. New York: Chapman and Hall, 1993.

Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.

Hanemann WM. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 1984; **66**: 332-341.

Johannesson M. *Theory and methods of economic evaluation in health care*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

Johansson P-O. *Evaluating health risks. An economic approach*. Cambridge University Press, 1995.

Kriström B. A non-parametric approach to the estimation of welfare measures on discrete response valuation studies. *Land Economics* 1990; **66**: 135-139.

Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997;**16**:33-64.

Ng Yew-Kwang. A case for happiness, cardinalism, and interpersonal comparability. *The Economic Journal* 1997; **107**: 1848-1858.

NOAA (National Oceanic and Atmospheric Administration). Report of the NOAA panel on contingent valuation. *Federal Register* 1993; **58**: 4602-4614.

O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; **32**: 150-163.

Park T, Loomis JB, Creel M. Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Economics* 1991; **67**: 64-73.

Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics* 1997; **6**: 243-252.

SBU (The Swedish Council on Technology Assessment in Health Care). SBU report no. 131, Stockholm, Sweden, 1996.

Sonnenberg FA, Beck JR. Markov models in medical decision making. *Medical Decision Making* 1993; **13**: 322-338.

Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. Forthcoming in International Journal of Technology Assessment in Health Care.

Wakker P, Klaassen MP. Confidence Intervals for Cost/Effectiveness Ratios. *Health Economics* 1995; **4**: 373-381.

Weinstein MC, Fineberger HV. *Clinical decision analysis*. Philadelphia: W.B Saunders, 1980.

Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care* 1990; **6**: 93-103.

Weinstein MC, Zeckhauser R. Critical Ratios and Efficient Allocation. *Journal of Public Economics* 1973; **2**: 147-157.

Willan AR, O'Brien BJ. Confidence intervals for cost-effectiveness ratios: An application of Fiellers theorem. *Health Economics* 1996; **5**: 297-305.

Chapter 2

Willingness to Pay for Hormone Replacement Therapy*

Niklas Zethraeus

Abstract: This study addresses the question of willingness to pay (WTP) for hormone replacement therapy (HRT) in order to alleviate menopausal symptoms. The woman obtains utility from consumption of goods and health. The purchase of a treatment is represented as a shift in the health production function during the treatment period. The mean WTP for the HRT is estimated using a parametric and a non-parametric method. The mean WTP based on these two methods is similar in both cases and amounts to about SEK 40,000 per year. Further, it is shown that the mean WTP is above the mean treatment cost of HRT. Finally, the implied WTP per gained quality adjusted life year (QALY) is estimated at about SEK 120,000 and SEK 160,000 based on the rating scale (RS) and time trade-off (TTO) methods, respectively.

Keywords: Hormone Replacement Therapy, Menopausal Symptoms, Willingness to Pay.

* Comments from Magnus Johannesson, Per-Olov Johansson and two anonymous referees for *Health Economics* are gratefully acknowledged. Comments from participants at the Health Economic Seminars at the Stockholm School of Economics are highly appreciated.

1. Introduction

At menopause, around the age of 50, about 80 percent of women experience menopausal symptoms (Daly et al. 1993). Symptoms, for example, are hot flushes, night sweats and atrophy-related symptoms in the urogenital tract. The presence of menopausal symptoms decreases the quality of life of women. The loss in quality of life may be substantial, which is indicated in Daly et al. (1993). Hormone Replacement Therapy (HRT) may alleviate these symptoms and thus increase the quality of life of symptomatic women (Karlberg et al. 1995, Wiklund et al. 1993). HRT may also have a cardioprotective effect and offers protection for osteoporosis and related fractures (Wiklund et al. 1993). The evidence on the effect of HRT on breast cancer is inconclusive (Colditz et al. 1995, Stanford et al. 1995).

In recent years several studies have focused on the benefits and costs of HRT (Cheung and Wren 1992, Daly et al. 1992, Tosteson et al. 1990, Tosteson and Weinstein 1991, Tosteson 1993, Weinstein 1980, Weinstein and Schiff 1983, Weinstein and Tosteson 1990). These are all examples of cost-utility analyses where costs are measured in *monetary units* and benefits are measured in *gained quality adjusted life years* (QALYs). An alternative way to assess the value of the change in quality of life due to HRT is to measure the willingness to pay (WTP) for HRT. Then it becomes possible to compare, directly, the benefits of HRT, measured in monetary units, with the costs of HRT. Further, it also becomes possible to calculate the WTP for a gained QALY, which may be compared to the costs for producing a gained QALY.

The purpose of this study is to use the contingent valuation method (CVM) to analyse how much symptomatic women are willing to pay for HRT. It is assumed that the woman maximises her expected present value utility and that she consumes goods and values her health, which is produced through a health production function. HRT is modelled as a (parametric) shift in the health production function. A WTP measure for the shift in the health production function is then defined. Section 2 presents the methods, while the results are found in Section 3. Section 4 concludes the paper.

2. Methods

The decision for a woman to use HRT is modelled using a life cycle model. It is assumed that the woman maximises her expected present value utility. She consumes goods and values her health, which is produced through a health production function. The treatment under investigation, HRT, is modelled as a shift in the health production function, i.e. it improves the individual's ability to 'produce' health. Since the survival probability is assumed to be a positive function of an individual's health capital, the considered treatment will, in addition to its direct impact on health and utility, also affect her survival probability. The WTP measure corresponds to the value of the change in the health production function caused by the treatment[1].

A questionnaire was consecutively administered to 104 women recruited from the Department of Gynaecology at Södertälje Hospital during the period February 6, 1995 to March 18, 1996. In Sweden, at the time of the study, a woman had to pay SEK 160 for the most costly pharmaceutical product in a prescription and SEK 60 per each further product in the prescription. The patient charge per surgery visit amounted to about SEK 120. However, the largest amount to pay for pharmaceutical products and outpatient care each year was SEK 1,800. The National Social Insurance Board (NSIB) paid for the overshooting cost for pharmaceuticals and outpatient care. The mean age of the entire patient group was 52.2 years (range 45-65 years). 82 percent of the women were treated with oestrogen in combination with a progestogen, while 18 percent were treated with oestrogen alone. Women treated with oestrogen alone were all hysterectomised. 51 percent of the women were given a transdermal HRT preparation while 49 percent were given pills. The mean treatment duration for the women was 3 years. After their consultation with the clinic doctor, a nurse at the clinic interviewed all women. The criteria for eligibility were that the women were between 45 and 65 years of age and that they had been treated with HRT for at least a period of 1 month.

The interview consisted of three parts. In the first part the woman was asked to indicate her health status before initiating HRT and her present health status with HRT treatment on a rating scale (RS) between 0 (dead) and 100 (full health). Based on the answers to the two RS questions we estimated the QALY weight with and without HRT.

[1] The underlying model is more thoroughly described in a paper in the Working Paper Series in Economics and Finance at the Stockholm School of Economics [Zethraeus and Johansson 1997].

Another method used to estimate QALY weights is the time trade-off (TTO) method (Drummond et al. 1987). In the study the woman was asked two TTO questions to estimate the QALY weight with and without HRT. The first TTO question was phrased as: "Suppose that you would experience the symptoms you had before the HRT was initiated for a period of 30 years. Indicate on the scale below how many years in full health followed by death you consider to be equivalent to 30 years with the experienced symptoms followed by death." This question was then repeated for the current health status of the woman with HRT.

In the third part of the interview the WTP for HRT was investigated using the CVM. In the CVM, survey methods are used to investigate the hypothetical WTP for a good. The CVM was originally developed in the environmental field to measure the value of changes in the environment, but recently a number of health care applications have also appeared (Johannesson 1996, Johansson 1995). In Sweden the CVM has been used in several applications, e.g., to investigate the WTP for antihypertensive therapy (Johannesson et al. 1991, Johannesson et al. 1993, Johannesson and Johansson 1996, Kartman et al. 1996).

Contingent valuation questions can be classified either as open-ended or binary contingent valuation questions (Mitchell and Carson 1989). In open-ended questions the maximum willingness to pay is elicited from each respondent, and in binary questions each respondent only accepts or rejects one price (bid). By varying the price in different subsamples the mean and median WTP can be estimated based on binary questions (Hanemann 1978, Hanemann 1984). Due to problems of non-response a bidding game is often used to elicit the WTP in open-ended questions (Randall et al. 1974). A bidding game resembles an auction. A first bid is made to the respondent, who accepts or rejects, and then the bid is raised or lowered depending on the answer. The process goes on until the respondent's maximum WTP is reached. An important problem when using a bidding game is that the WTP is often affected by the first bid made (Boyle et al. 1985, Kartman et al. 1996, Rowe et al. 1980, Stålhammar 1996). This is referred to as starting point bias in the literature (Boyle et al. 1985, Kartman et al. 1996, Mitchell and Carson 1989, Rowe et al. 1980, Stålhammar 1996). Due to the problems of open-ended questions most studies in the environmental field now use binary contingent valuation questions. The use of binary questions was also recommended by an

expert panel, appointed by the National Oceanic and Atmospheric Administration (NOAA) in the US, to assess the validity and reliability of the CVM (NOAA 1993).

The CVM based on binary questions is, however, not without problems, and important issues remain to be resolved in order to establish the validity of the method. The two perhaps most important problems are that some experimental results indicate that hypothetical WTP overestimates real WTP, and that some studies indicate that hypothetical WTP is insensitive to the size of the programme (often referred to as insensitivity of scope). To test the validity of the WTP estimates in a contingent valuation study, it can be tested whether the hypothesised theoretical relationships are supported by the data (Mitchell and Carson 1989). It may for instance, be tested whether the WTP increases with income and whether the WTP increases with the size (scope) of the programme. In the health field testing for scope would mean testing whether the WTP increases with the size of the health change.

In this study we used a binary contingent valuation question, which means that each individual is asked if they would pay a specific price (P) or not. By varying the price in different subsamples it is possible to trace the relationship between the price and the proportion of individuals who are willing to pay.

In the WTP question the woman was asked if she would continue her current HRT if she had to pay SEK P per month out of her own money. The price (P) was randomly varied between SEK 100 and SEK 10,000 in eight different sub samples, and each individual only received one of these prices[2]. The eight different prices were SEK 100, SEK 500, SEK 1,000, SEK 1,500, SEK 2,000, SEK 3,000, SEK 5,000 and SEK 10,000 (*Appendix*)[3]. Data was also collected about the following socio-economic variables: pre-tax household income, education level, age, and household size.

We attempted to estimate a per capita demand function for the considered (take it or leave it) commodity, i.e. we wanted to estimate the probability that the commodity would be purchased. In estimating the probability of agreeing to pay a specified amount of money P in exchange for

[2] The bid vector was established after a small pilot study of 12 patients.
[3] SEK = Swedish crowns. The exchange rates May 21, 1996 were: GB£1=SEK 10.3; US$1=SEK 6.8.

the considered treatment, i.e. HRT, we assumed a logistic model. The acceptance probability Π is written as follows:

$$\Pi = \mathrm{F}(P) = 1/\left[1 + e^{-\Delta v}\right] \qquad (1)$$

where F*(P)* is the 'survivor' function yielding the probability of accepting to pay at least SEK P in exchange for the treatment, and Δv is the change in utility caused by the considered improvement in health if the person pays SEK P for the improvement. In what follows, we assume a linear approximation of the change in utility: $\Delta v = \eta_0 + \eta_1 P + \eta_2 h + \eta_3 \boldsymbol{S}$, where h is a measure of the change in health status (or quality of life) due to the treatment, $\boldsymbol{S}$ is a vector of socio-economic variables, η_i for $i = 0,1,2$ are parameters to be estimated, and η_3 is a vector of parameters to be estimated. The change in health status h is measured using either the RS or the TTO method. Then it is possible to test the hypothesis that the probability of accepting to pay increases if the size of the change in health status increases (sensitivity of scope).

In order to estimate the mean WTP for HRT the following equation was estimated[4]:

$$\ln[\Pi/1\text{-}\Pi] = \eta_0 + \eta_1 P + \eta_2 h + \eta_3 \boldsymbol{S} \qquad (2)$$

As can be seen from equation (1), the regression equations predict that a certain proportion of respondents have a negative WTP since Π will approach one as P approaches minus infinity. However, medical treatment is a private commodity, which you freely may or may not elect to buy. For this reason, we rule out a negative WTP in the estimation of the mean WTP for the treatment. WTP is set equal to zero for the proportion of respondents, which are predicted to have a negative WTP.

In the case where the WTP is non-negative, and the probability of a zero WTP is strictly positive, the mean WTP is equal to (Johansson 1995, O'Conor 1995):

---

[4] The following equation was also estimated: $\ln[\Pi/(1\text{-}\Pi)] = \eta_0' + \eta_1' \ln(P) + \eta_2' h + \eta_3' \boldsymbol{S}$. However, since $[1/\eta_1' < -1]$, the integral did not converge, resulting in an infinite willingness to pay for HRT.

$$dI = \int_0^{\infty} \frac{1}{1+e^{-(\eta_4+\eta_1 P)}} dP = -\frac{1}{\eta_1} \ln(1+e^{\eta_4}) \qquad (3)$$

where *dI* denotes the mean WTP for the treatment and η_4 denotes the magnitude of the constant term in equation (1) when the elements of $\boldsymbol{S}$ and h are assigned particular values. The mean WTP was estimated with the explanatory variables set at their sample means. Thus, we are estimating WTP for an average respondent.

To estimate the mean and median WTP for HRT based on the answers to the contingent valuation question, a non-parametric method was also used (Kriström 1990). With this method, the proportion of yes answers at the different price levels was used to construct a curve that shows the relationship between the price and the proportion of yes answers. This curve can be interpreted as a demand curve and the mean WTP is measured as the area below the curve. In the estimation of the mean WTP we assumed that the highest WTP is equal to the highest price of SEK 10,000 used in the study. We also assumed that every woman would continue her HRT if the bid was SEK 0. The median WTP is the price where 50 percent would accept to pay and 50 percent would reject to pay.

3. Results

Table 1 shows mean values and standard deviations in parenthesis for the included explanatory variables in the logistic regression equations according to equation (2). It can be noted that the patient group is relatively homogeneous with respect to age, which reports a standard deviation of 3.87. This is explained by the inclusion criterion, which restricts the patient sample to women at the age of 45 to 65 years (*Table 1* in here).

Table 2 shows the results of the logistic regressions of the intention to pay for the HRT. We report two goodness-of-fit measures: the percentage of correctly predicted responses and the likelihood ratio index (LRI) (Statistisk årsbok 1995). The estimated parameter of the bid variable is significant at the 1 percent level in the two regressions with the expected sign. In the first regression the perceived change in health status is represented by the change in the TTO score with and without HRT. The estimated parameter of ΔTTO has the expected sign and is

significant at the 10 percent level using a one-sided t-test (Mitchell and Carson 1989)[5]. In the second regression the perceived change in health is represented by the change in the RS score with and without treatment. The estimated parameter of ΔRS has the expected sign and is significant at the 10 percent level using a one-sided t-test (Greene 1993)[6]. The income variable is not significant but has the expected sign. The education variable is significant at the 10 percent level using a one-sided t-test (since the hypothesis is that a higher education level increases the probability of agreeing to pay). The estimated mean WTP in the two regression equations is SEK 3,772 and SEK 3,651 per month, respectively[7] (*Table 2* in here).

Figure 1 shows the relationship between the price on the WTP question and the proportion of patients accepting to pay the price (*Figure 1* in here).

Based on the curve in *Figure 1* the mean and median WTP were estimated. The mean and median WTP was SEK 3,508 and SEK 2,000 respectively. Thus, 50 percent of the women accept to pay SEK 2,000 for continuing their treatment of menopausal symptoms.

Finally, the implied WTP for a gained QALY was calculated by dividing the mean WTP (*Table 2*) by the mean gain in quality of life (*Table 1*). Based on the TTO method the WTP per gained QALY was estimated at SEK 156,100. Based on the RS method the WTP per gained QALY was estimated at SEK 118,400.

4. Concluding Remarks

The purpose of this study was to use the CVM to analyse how much symptomatic women are willing to pay for HRT. The mean WTP per month for HRT is estimated at about SEK 3,700 using logistic regression analysis. The mean WTP per month using a non-parametric method is estimated at about SEK 3,500. The results, based on the parametric and non-parametric method,

[5] This test is relevant for the hypothesis that the perceived change in health status represented by TTO or RS increases the probability of agreeing to pay. For a further discussion of one-sided hypothesis tests, see Mitchell and Carson (1989).

[6] A reason why the health status variables are significant only at the 10 percent level using a one-sided test, may be measurement errors in these variables (according to TTO and RS). Measurement errors in one variable imply that the estimated coefficient of the variable is biased towards zero, which is denoted attenuation (Greene 1993).

[7] We also tested for if there was a significant difference in the WTP of patients given pills versus patients given transdermal HRT preparations. No significant difference was found.

are thus similar. The yearly WTP for HRT amounts to about SEK 42,000. This amount can be compared to the yearly WTP for reducing high blood pressure which was estimated at about SEK 9,600 (Johannesson et al. 1993). The WTP for HRT is thus quite high and can be compared with the average yearly pre-tax household income in the sample of SEK 334,000. An average woman is then prepared to pay 12 percent of her yearly pre-tax household income for continuing the HRT. In Johannesson et al. (1993) an average individual was prepared to pay 4 percent of his/her yearly pre-tax household income for continuing the treatment for high blood pressure.

The estimated equations are in accordance with the predictions of the theory. An increased price reduces the demand for HRT and an increased perceived change in health status between HRT and no HRT implies that the WTP for HRT increases. Also the education variable was significant and showed that a higher education level implied an increased WTP for HRT. This result conforms to the predictions of Grossman investment model, where education is positively related to health because education increases the ability of producing health (Grossman 1972). Thus, the marginal cost of investing in health is reduced. A given increase in health status is more easily produced if an individual has a higher education level.

The mean WTP can also be compared with the treatment costs of HRT. Based on the total patient sample we estimated the yearly mean treatment costs of HRT[8]. The annual treatment costs per patient amounted to about SEK 1,600 for patients treated with oestrogen alone and to about SEK 2,200 for patients treated with oestrogen in combination with a progestogen. Thus, the yearly mean WTP for HRT is well above the mean treatment costs associated with HRT.

The implied WTP for a gained QALY was estimated at SEK 156,100 and SEK 118,400 using the TTO and RS method respectively. These figures may be compared to the costs for producing QALYs. If the WTP for a gained QALY exceeds the costs for producing a QALY, it is indicated that the treatment is motivated from an economical perspective. In Sweden the price per QALY

[8] The treatment costs included the costs for the drugs, the physician's visits, and the patients' time and travelling costs. Costs for oestrogen only: Drugs SEK 860, 1 physician outpatient visit SEK 601, costs for time and travelling SEK 180; Costs for oestrogen combined with a progestin: Drugs SEK 1,055, 1.5 physician outpatient visit SEK 901, time and travelling SEK 270.

gained is implicitly stated in cost-benefit analysis of road investments where the price per discounted QALY gained was estimated at about SEK 700,000 (The discount rate was 3%) (Johannesson et al. 1997).

The results in this study demonstrate that the increase in quality of life from using HRT greatly exceeds the assumed increases made in earlier studies. Weinstein, for instance, assumed an increase in the quality of life weight of 0.01 due to HRT for women with menopausal symptoms (Weinstein 1980, Weinstein and Schiff 1983). This shows the importance of carrying out empirical studies on quality of life rather than making arbitrary assumptions. Thus, the high WTP for HRT for symptomatic women can be explained by a considerable increase in the quality of life in terms of changes in the TTO and RS[9].

[9] The improvement in the quality of life based on the RS method is similar to the results in the study by Daly et al. (1993).

5. References

Boyle KJ, Bishop RC, Welsh MP. Starting point bias in contingent valuation bidding games. *Land Economics* 1985; **61**: 188-194.

Cheung AP, Wren BG. A cost-effectiveness analysis of hormone replacement therapy in the menopause. *The Medical Journal of Australia* 1992; **156**: 312-316.

Colditz GA, Hankinson SE, Hunter DJ, Willett WC, Manson JE, Stampfer MJ, Hennekens C, Rosner B, Spezer FE. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *New England Journal of Medicine* 1995; **332**: 1589-1593.

Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the Impact of Menopausal Symptoms on Quality of Life. *British Medical Journal* 1993; **307**: 836-840.

Daly E, Roche M, Barlow D, Gray A., McPherson, K. and Vessey, M. HRT: An analysis of benefits, risks and costs. *British Medical Bulletin* 1992; **48**: 368-400.

Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford Medical Publications, 1987.

Greene WH. *Econometric Analysis, Second Edition*. New York: Macmillan, 1993.

Grossman M. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 1972; **80**: 223-255.

Hanemann MW. *A methodological and empirical study of the recreation benefits from water quality improvement*. Ph.D. Dissertation, Harvard, 1978.

Hanemann MW. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 1984; **66**: 332-341.

Johannesson M, Johansson P-O, Kriström B, Gerdtham U-G. Willingness to pay for antihypertensive therapy: further results. *Journal of Health Economics* 1993; **12**: 95-108.

Johannesson M, Johansson P-O. To be, or not to be, that is the question: An empirical study of the WTP for an increased life expectancy at an advanced age. *Journal of Risk and Uncertainty* 1996; **13**: 163-174.

Johannesson M, Jönsson B, Borgquist L. Willingness to pay for antihypertensive therapy: results of a Swedish pilot study. *Journal of Health Economics* 1991; **10**: 461-474.

Johannesson M, Meltzer D, O'Conor RM. Incorporating future costs in medical cost-effectiveness analysis: Implications for the cost-effectiveness of the treatment of hypertension. *Medical Decision Making* 1997; **17**: 382-389.

Johannesson M. *Theory and Methods of Economic Evaluation of Health Care*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

Johansson P-O. *Evaluating Health Risks. An Economic Approach.* Cambridge: Cambridge University Press, 1995.

Karlberg J, Mattsson L-Å, Wiklund, I. A quality of life perspective on who benefits from estradiol replacement therapy. *Acta Obstet Gynecol Scand* 1995; **74**: 367-372.

Kartman B, Andersson F, Johannesson M. Willingness to pay for reductions in angina pectoris attacks. *Medical Decision Making* 1996; **16**: 248-253.

Kriström B. A non-parametric approach to the estimation of welfare measures on discrete response valuation studies. *Land Economics* 1990; **66**: 135-139.

Mitchell RC, Carson RT. *Using Surveys to Value Public Goods: the Contingent Valuation Method, Resources for the Future*. Washington DC: The John Hopkins University Press, 1989.

NOAA (National Oceanic and Atmospheric Administration). Report of the NOAA panel on contingent valuation. *Federal Register* 1993; **58**: 4602-4614.

O'Conor RM. *Consumer-Patient Valuation of Drug Safety and Efficacy*. Ph.D. Dissertation, Department of Economics, University of Kentucky, 1995.

Randall A, Ives BC, Eastman C. Bidding games for valuation of aesthetic environmental improvements. *Journal of Environmental Economics and Management* 1974; **1**: 132-149.

Rowe RD, d'Arge RC, Brookshire DS. An experiment on the economic value of visibility. *Journal of Environmental Economics and Management* 1980; **7**: 1-19.

Stanford JL, Weiss NS, Voigt LF, Daling JR, Habel LA, Rossing MA. Combined estrogen and progestin hormone replacement therapy in relation to risk of breast cancer in middle-aged women. *Journal of the Medical Association* 1995; **274**: 137-142.

Statistisk Årsbok 96 [Statistical Yearbook of Sweden]. Stockholm, Sweden: Norstedts Tryckeri AB, 1995.

Stålhammar N-O. An empirical note on willingness to pay and starting point bias. *Medical Decision Making* 1996; **16**: 242-247.

Tosteson A. *A review and update of cost-effectiveness of hormone replacement therapy in the menopause*. In Cosséry J-M, ed. Medical-Economic aspects of Hormone Replacement therapy, The Parthenon Publishing Group, 1993.

Tosteson A, Rosenthal DI, Melton LJ, Weinstein MC. Cost-effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Annals of Internal Medicine* 1990; **113**: 594-603.

Tosteson A, Weinstein MC. Cost-effectiveness of hormone replacement therapy after the menopause. *Baillière's Clinical Obstetrics and Gynaecology* 1991; **5(4)**: 943-959.

Weinstein MC. Estrogen use in postmenopausal women: costs, risks, and benefits. *New England Journal of Medicine* 1980; **303**: 308-316.

Weinstein MC, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstetrical and Gynaecological Survey* 1983; **38**: 445-455.

Weinstein MC, Tosteson A. Cost-effectiveness of hormone replacement. *Annals New York Academy of Sciences* 1990; **592**: 162-172.

Wiklund I., Karlberg J, Mattsson, L-Å. Quality of life of postmenopausal women on a regimen of transdermal estradiol therapy: A double-blind placebo-controlled study. *American Journal of Obstetrics and Gynecology* 1993; **168**: 824-830.

Zethraeus N, Johansson P-O. *Willingness to pay for hormone replacement therapy*. Published in the Working Paper Series in Economics and Finance at the Stockholm School of Economics, 1997, Working paper No. 214.

Appendix

The formulation of the willingness-to-pay question:

This question focuses on how much you value the continuation of your hormone replacement therapy. Presume that you have to pay the majority of the treatment costs for drugs and physician visits by yourself. Would you choose to continue your current treatment for menopausal symptoms if you had to pay SEK 1,000 each month[10] as patient charges for the treatment? Be aware that the money is taken from your own disposable income and hence decreases your private consumption.

Alternatives:

Yes ☐

No ☐

Statements of your motives:

Follow-up question:

Are you sure or uncertain that you want to pay SEK 1,000 for continuing the hormone replacement therapy?

Certain ☐

Uncertain ☐

[10] The bid varies from SEK 100 to SEK 10,000 in eight sub-samples.

Tables

Table 1: Mean values of the explanatory variables included in the logistic regressions[a].

Variable	Bid	ΔTTO[b]	ΔRS[b]	Age	Income[c]	Household Size	Education[d]
TOTAL	2,640	0.29	0.37	52.2	27,840	2.15	0.55
Standard Deviation	(2,936)	(0.28)	(0.26)	(3.87)	(12,092)	(0.98)	(0.5)

[a] To be noted in interpreting the results of the present study is that the patient sample in the study may not be representative of the overall patient population receiving HRT in Sweden. Care should thus be taken in extrapolating the results to other populations and settings. To compare the patient population with women from the Swedish population, the mean values of the socio-economic variables household income, household size and education level were compared with the mean values for the Swedish population. The mean household size in the Swedish population is 2.1 and the monthly mean pre-tax household-income is SEK 18,500 (Statistisk årsbok 1995). The mean education level for women, 45-54 years of age, is 0.7 (Statistisk årsbok 1995)[11]. Thus, the mean education level in the patient sample is lower compared to the mean education level for women in the Swedish population. However, the mean household income in the patient sample is above the mean household income in the Swedish population. Only the household size is similar in the patient sample and in the Swedish population.

[b] Difference in the quality of life score with and without HRT.

[c] Pre-tax household income per month.

[d] Coded 0 for primary education and 1 for secondary and university or higher education.

[11] Coded 0 for primary education and 1 for secondary and university or higher education. 1 January, 1995.

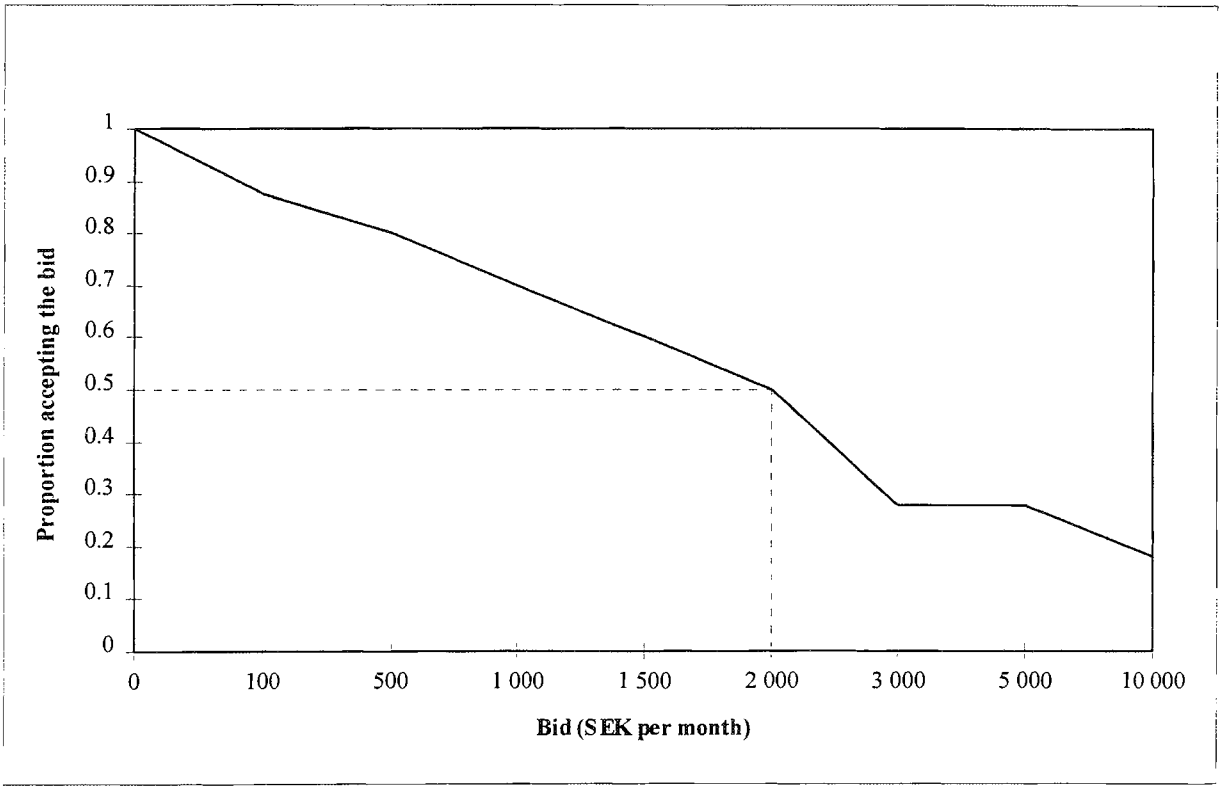
Table 2: Coefficients of the logistic regressions for the full sample (t-ratios are shown within parentheses).

Variable	Equation 1	Equation 2
Intercept	2.6524	1.7952
	(0.72)	(0.53)
Bid	-0.0004***	-0.0004***
	(-3.49)	(-3.67)
ΔTTO	1.2528	/
	(1.44)	/
ΔRS	/	1.2301
	/	(1.31)
Age	-0.0443	-0.0291
	(-0.67)	(-0.48)
Income	9.78×10^{-6}	1.00×10^{-5}
	(0.44)	(0.45)
Education	0.6267	0.8122
	(1.25)	(1.61)
Household size	-0.0734	-0.1097
	(-0.27)	(-0.41)
N	104	104
Goodness of fit:		
Correctly predicted		
Responses (%)	75.00	73.79
LRI	0.18	0.17
Mean WTP per month (SEK):	3,772	3,651

Notes: *,**,*** Significant at 10%, 5% and 1% levels (two-sided t-test).

Figures

Figure 1: The relationship between the bid level and the proportion of patients willing to pay the bid. *N*=104.



Chapter 3

The Impact of Hormone Replacement Therapy on Quality of Life and Willingness to Pay*

Niklas Zethraeus, Magnus Johannesson,

Peter Henriksson and Roland T. Strand

Abstract: The purpose of this paper is to measure the gain in quality of life due to hormone replacement therapy (HRT) for women with mild and severe menopausal symptoms. It is a prospective study where data on quality of life and willingness to pay (WTP) were collected by interview. A questionnaire was administered to 104 consecutive women attending the Department of Gynaecology at Södertälje Hospital near Stockholm. The time trade-off (TTO) and rating scale (RS) methods measured the gain in quality of life. The WTP for HRT was investigated using the contingent valuation method (CVM). The increase in the quality adjusted life year (QALY) weight due to HRT for women with mild symptoms was 0.26 according to the RS method and 0.18 according to the TTO method. For women with severe symptoms the QALY weight increased by 0.50 according to the RS method and by 0.42 according to the TTO method. The mean WTP for HRT per month was SEK 2,300 for women with mild symptoms and SEK 4,800 for women with severe symptoms. HRT leads to a major improvement in quality of life for women with menopausal symptoms. Both for women with mild and severe menopausal symptoms the WTP for the treatment exceeds the costs, indicating that HRT is economically beneficial for women with menopausal symptoms.

Keywords: Economic Evaluation, Hormone Replacement Therapy, Quality of Life, Willingness to Pay.

---

* We are grateful to Ms V. Lundqvist and Ms M. Östh at the Department of Gynaecology, Södertälje Hospital, for assisting with the data collection and the two anonymous referees and the editor for *British Journal of Obstetrics and Gynaecology* for helpful comments.

1. Introduction

Hormone Replacement Therapy (HRT) reduces the menopausal symptoms, which many women experience about the age of 50, and therefore also increases the woman's quality of life during these years (Karlberg et al. 1995, Wiklund et al. 1993). To allocate scarce resources effectively in the health care sector economic evaluation methods should be carried out, for example by cost-effectiveness analysis (Drummond et al. 1987). The cost-effectiveness of HRT has been evaluated in several studies (Cheung and Wren 1992, Daly et al. 1992, Tosteson et al. 1990, Weinstein 1980, Weinstein and Schiff 1983). The results of these studies are very sensitive to the quality of life weights assigned to different health states affected by the HRT. These studies have based their estimations of the quality of life weights on assumptions, which are more or less qualified guesses rather than based on empirical data.

Daly et al. (1993) indicated that the effect of menopausal symptoms on the quality of life might have been underestimated in earlier studies. In their study 63 women were asked to assess mild and severe menopausal symptoms with the rating scale (RS) and time trade-off (TTO) methods. Daly's results showed that menopausal symptoms seem to have a much more severe impact on quality of life than assumed in analyses of cost-effectiveness, and that HRT caused a much greater improvement in quality of life than previously been shown (Cheung and Wren 1992, Daly et al. 1992, Tosteson et al. 1990, Weinstein 1980, Weinstein and Schiff 1983). Due to the small sample size these results need to be interpreted with caution; furthermore it is accepted that the RS method tends to exaggerate the gain in quality of life compared to other methods (Hornberger et al. 1992, Read et al. 1984). It is therefore important to carry out further studies to confirm the Daly's (1993) results and to measure the gain in the quality of life using other methods. Presently there is no consensus on what method to use for measuring the gain in quality of life, justifying the use of more than one method (Weinstein et al. 1996).

The purpose of this study was to estimate the gain in quality of life from HRT for women with mild and severe menopausal symptoms, by the RS and the TTO method. For the first time the willingness to pay (WTP) for HRT is also measured.

2. Methods

A questionnaire was administered to 104 consecutive women attending the Department of Gynaecology at Södertälje Hospital during the period 6 February 1995 to 18 March 1996. Fifty-six women had mild symptoms and 48 severe symptoms. The mean age of the whole group was 52.2 years (range 45-65). The mean age for women with mild and severe menopausal symptoms was 52.0 (range 45-60) and 52.4 years (range 45-65), respectively. Eighty-five women were treated with oestrogen in combination with a progestogen, while 19 women were treated with oestrogen alone. The women treated with oestrogen alone had all had a hysterectomy. The mean duration of treatment for all the women was 3 years. A nurse at the clinic interviewed all the women. To classify each woman as having mild or severe symptoms she was asked to read a description of mild and severe symptoms used in the study by Daly et al. (1993), and to choose which alternative that best corresponded to her own symptoms before taking HRT. About 180,000 or 10% of all women aged 45 years or older are treated with oestrogen in Sweden (SBU 1996). The total cost per year of oestrogen treatment are estimated at about SEK 300-400 million. Oestrogen prescriptions are mainly made by general practitioners (24%) and gynaecologists (69%) (NCSP 1995).

The interview consisted of three parts. In the first part the woman was asked to indicate her health status before starting HRT and her present health status with HRT, on a RS between 0 (dead) and 100 (full health). The RS method is commonly used to estimate the quality of life weights to construct quality adjusted life years (QALYs). QALYs is the measurement typically used in cost-effectiveness analysis (Drummond et al. 1987). To obtain the QALY weight with the RS method the score on the scale is divided by 100: for example, if a woman places her current health state on 70, the QALY weight will be 0.7 (70/100). The QALY weights were estimated with and without HRT.

Another method used to estimate QALY weights is the TTO method (Drummond et al. 1987). This involves a trade-off between quantity and quality of life. An individual is asked to state the number of years in full health followed by death that is deemed as being equivalent to a specific number of years in the health state to be assessed, followed by death. To obtain the QALY weight the number of years in full health is divided by the number of years in the health state to be assessed. For example, if a woman thinks that 20 years in full health

followed by death is of equal value to 30 years with mild menopausal symptoms followed by death, the QALY weight for mild menopausal symptoms is 0.67.

In this study the women were asked two TTO questions to estimate the QALY weights with and without HRT. The first time trade off question was: "Suppose that you experienced the symptoms you had before you started HRT for 30 years. Indicate on the scale below how many years in full health followed by death that are equivalent to 30 years with your symptoms followed by death." This question was then repeated for the current health status of the woman when taking HRT. It should be noted that we assume both for the RS and TTO that the women's recollection of the 'before' state is representative of the 'no treatment' state.

In the third part of the interview the WTP for HRT was investigated using the contingent valuation method (CVM). This method was originally developed to measure the value of changes in the environment, but has recently been applied to a number of health care applications (Carson 1991, Johannesson and Jönsson 1991, Mitchell and Carson 1989). In this study we used a binary contingent valuation question which asked each woman if she would pay a specific price or not. By varying the price in different subsamples it is possible to trace out the relationship between the price and the proportion of individuals that are willing to pay.

In the WTP question the woman was asked if she would continue her present HRT if she had to pay SEK P per month out of her own money. The prices were established after a small pilot study of 12 women. The eight different prices were SEK 100; 500; 1,000; 1,500; 2,000; 3,000; 5,000 and 10,000 (£1 = SEK 10.3). The price (P) was randomly varied between SEK 100 and SEK 10,000 in eight different subsamples, and each individual only received one of these prices

To estimate the mean and median WTP for HRT a non-parametric method was used (Kriström 1990). With this method the proportion of 'yes' answers at the different price levels are used to construct a curve that shows the relation between the price and the proportion of 'yes' answers and the mean WTP is measured as the area below the curve. The median WTP is the price where 50% would accept to pay and 50% would reject to pay. In the estimation of

the mean WTP we assumed that the highest possible WTP is equal to the highest price of SEK 10,000 used in the study.

We present the results separately for women with mild and severe menopausal symptoms. Mean values within samples was compared by means of the non-parametric Wilcoxon matched-pairs signed-rank test (Newbold 1991). Mean values between samples were compared by means of the non-parametric Mann-Whitney test (Newbold 1991).

Since it is not possible to test if the WTP differs between women with mild and severe menopausal symptoms by the non-parametric method, we tested this using logistic regression analysis (Greene 1993). In the logistic regression analysis we tested if the probability of accepting to pay the price (i.e. the probability of answering 'yes' to the WTP question) differed significantly between women with mild and severe symptoms. In this analysis the yes/no answer to the WTP question for the full sample was used as the dependent variable and the logarithm of the price received by the women and a dummy variable for mild/severe symptoms were used as independent variables. We use the logistic regression analysis to test if the probability of answering 'yes' to the WTP-question differs significantly between patients with mild and severe symptoms (i.e. if the mild/severe symptoms dummy variable is statistically significant).

3. Results

Table 1 shows the results of the RS and TTO questions (*Table 1* in here). For women with mild symptoms the quality of life weight is 0.60 without HRT and 0.86 with HRT, according to the RS method. The increase in quality of life amounts to 0.26 ($P < 0.001$). The TTO method also shows a significant improvement in quality of life for women with mild symptoms, from 0.73 to 0.90 ($P < 0.001$).

For women with severe symptoms the quality of life weight, based on the RS method, is 0.32 without HRT and 0.82 with HRT. The increase in quality of life amounts to 0.50 ($P < 0.001$). With the TTO method the quality of life for women with severe symptoms increases from 0.52 to 0.95 with HRT ($P < 0.001$).

The improvements in quality of life with the RS and the TTO methods with HRT is about doubled for women with severe symptoms as compared with those with mild symptoms ($P < 0.001$). This larger improvement is due to the lower quality of life without HRT for the women with severe symptoms ($P < 0.001$ for both RS and TTO). For mild symptoms the RS method shows a greater improvement in quality of life than the TTO method ($P = 0.003$). For severe symptoms the RS shows a greater improvement in quality of life than the TTO method ($P = 0.06$). For the whole study the change of improvement in quality of life between RS and the TTO was significant ($P < 0.001$).

Figure 1 shows the relationship between the price on the WTP question and the proportion of women accepting to pay that price for women with mild and severe symptoms (*Figure 1* in here). The proportion of women that are willing to pay a certain price is the same or higher for women with severe symptoms for every price. Based on the curves in *Figure 1* the mean and median WTP were estimated for women with mild and severe symptoms. These WTP estimates are shown in *Table 2* (*Table 2* in here).

For women suffering mild menopausal symptoms the mean WTP for HRT is about SEK 2,300 and the median WTP is about SEK 1,500. For women with severe menopausal symptoms the mean WTP is about SEK 4,800 and the median WTP is about SEK 2,000. The results of the logistic regression analysis shows that the probability of accepting to pay is higher for women with severe menopausal symptoms than for women with mild symptoms ($P = 0.002$).

4. Discussion and Conclusion

The results of this study using the RS method are similar to those of Daly et al. (1993). In that study the quality of life weight increased from 0.61 to 0.79 with HRT for women with mild symptoms and from 0.29 to 0.85 for women with severe symptoms, compared with increases from 0.60 to 0.86 for women with mild symptoms and 0.32 to 0.82 for women with severe symptoms in our study. The results of both studies show that the effect of menopausal symptoms on the quality of life has been underestimated.

In this study we also measured the gain in the quality of life with HRT by the TTO method. The TTO method gives a significantly lower gain in quality of life than the RS. The increase in the quality of life weight of 0.18 for women with mild symptoms and 0.42 for women with severe symptoms are, however, clinically significant. These increases can be compared with the assumptions made in cost effectiveness analyses of HRT. Weinstein, for instance, assumed an increase in the quality of life weight of 0.01 due to HRT for women with menopausal symptoms (Weinstein 1980, Weinstein and Schiff 1983). Our study shows the importance of carrying out empirical studies on quality of life rather than making arbitrary assumptions. The improvement in quality of life with HRT in our study was similar to that with transplantation in renal failure (Russell et al. 1992). The RS method shows a greater improvement in quality of life than the TTO method, which is in accordance with previous studies (Hornberger et al. 1992, Read et al. 1984).

This is the first study to measure willingness to pay for hormone replacement therapy, and shows that women will pay large sums of money for HRT. The mean WTP per month was about SEK 2,300 for women with mild symptoms and about SEK 4,800 for women with severe symptoms. These are sizeable amounts and can be compared with the average monthly pre-tax household income in the sample of SEK 28,000. Note that the monthly pre-tax income stated in this paper is relatively high compared with the average pre-tax income for the Swedish population, which amounts to about SEK 18,500. The WTP may be positively related to income explaining the relatively high WTP in our sample. In the Swedish population the mean WTP is probably lower because of a lower mean income. In the estimation of the mean WTP we assumed that the highest possible WTP is equal to the highest price of SEK 10,000 used in the study. This assumption had to be used only for women with severe symptoms (*Figure 1*) and implies that the mean WTP for women with severe symptoms is conservative.

The WTP amounts can be compared with the treatment costs of HRT. We estimated the treatment costs of HRT for the whole study, including the costs of drugs, visits to the physician, and the costs of time and travelling for the women. The total treatment costs amounted to about SEK 130 per month for women treated with oestrogen alone and to about

SEK 190 per month for women treated with oestrogen in combination with a progestogen. The WTP for HRT is thus well above the treatment costs associated with HRT both for women with mild and severe menopausal symptoms.

The results of the WTP part of the study should be interpreted with caution, however, for the validity of the contingent valuation method to measure WTP has not been established. The results of the WTP part of the study is, however, consistent with the RS and TTO measurements of quality of life, for women with severe menopausal symptoms have a greater increase in quality of life and are willing to pay more for their HRT than women with mild symptoms.

Another limitation of this study is that the women may not be representative of all women receiving HRT in Sweden and that the sub-groups may not be representative of all women with mild and severe menopausal symptoms in Sweden. The results of our study therefore may not be general. The results with the RS method are, however, very similar to those of Daly et al. (1993) in British women. Finally, it may be the case that the women's recollection of the 'before' state is not representative of the 'no treatment' state. A prospective study may be required where each woman is followed from the 'no treatment' state to reduce the extent of this problem.

5. References

Carson RT. *Constructed markets*. In J.B. Braden and C.D. Kolstad (Eds.), Measuring the Demand for Environmental Quality. Amsterdam: Elsevier/North Holland, 1991.

Cheung AP, Wren BG. A cost-effectiveness analysis of hormone replacement therapy in the menopause. *The Medical Journal of Australia* 1992; **156**: 312-316.

Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the Impact of Menopausal Symptoms on Quality of Life. *British Medical Journal* 1993; **307**: 836-840.

Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: An analysis of benefits, risks and costs. *British Medical Bulletin* 1992; **48**: 368-400.

Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford Medical Publications, 1987.

Greene WH. *Econometric Analysis*, Second Edition. New York: Macmillan, 1993.

Hornberger JC, Redelmeier DA, Peterson J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *Journal of Clinical Epidemiology* 1992; **45**: 505-512.

Johannesson M, Jönsson B. Economic evaluation in health care: is there a role for cost-benefit analysis? *Health Policy* 1991; **17**: 1-23.

Karlberg J, Mattsson L-Å, Wiklund I. A quality of life perspective on who benefits from estradiol replacement therapy. *Acta Obstetricia et Gynecologica Scandinavica* 1995; **74**: 367-372

Kriström, B. A non-parametric approach to the estimation of welfare measures of discrete response valuation studies. *Land Economics* 1990; **66**: 135-139.

Mitchell, RC, Carson RT. *Using Surveys to Value Public Goods: the Contingent Valuation Method, Resources for the Future*. Washington DC, 1989.

NCSP (National Corporation of Swedish Pharmacies). Östrogen/Gestagen. Kvinnor över 50 allt mer hormoniska. IA kontakt 1995; **5**: 1-4.

Newbold P. *Statistics for business and Economics, 3rd Ed.* Englewood Cliffs, New Jersey: Prentice-Hall, 1991.

Read, JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preference for health outcomes: comparison of assessment methods. *Medical Decision Making* 1984; **4**: 315-329.

Russell JD, Beecroft ML, Ludwin D, Churchill DN. The quality of life in renal transplantation-A prospective study. *Transplantation* 1992; **54**: 656-660.

SBU (The Swedish Council on Technology Assessment in Health Care). Oestrogen Treatment. SBU-rapport nr. 131, 1996, Stockholm Sweden.

Tosteson A., Rosenthal DI, Melton LJ, Weinstein MC. Cost-effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement theory. *Annals of Internal Medicine* 1990; **113**: 594-603.

Weinstein MC. Estrogen use in postmenopausal women: costs, risks, and benefits. *New England Journal of Medicine* 1980; **303**: 308-316.

Weinstein MC, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstetrical and Gynaecological Survey* 1983; **38**: 445-455.

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on cost-effectiveness in health and medicine. *Journal of the American Medical Association* 1996; **276(15)**: 1253-1258.

Wiklund I, Karlberg J, Mattsson L-Å. Quality of life of postmenopausal women on a regimen of transdermal estradiol therapy: A double-blind placebo-controlled study. *American Journal of Obstetrics and Gynecology* 1993; **168**: 824-830.

Tables

Table 1: Quality of life weights for women with mild and severe menopausal symptoms with and without hormone replacement therapy (HRT), based on the rating scale (RS) and time trade-off (TTO) methods. Standard deviations within parentheses. *N*=104 (Mild=56, Severe=48).

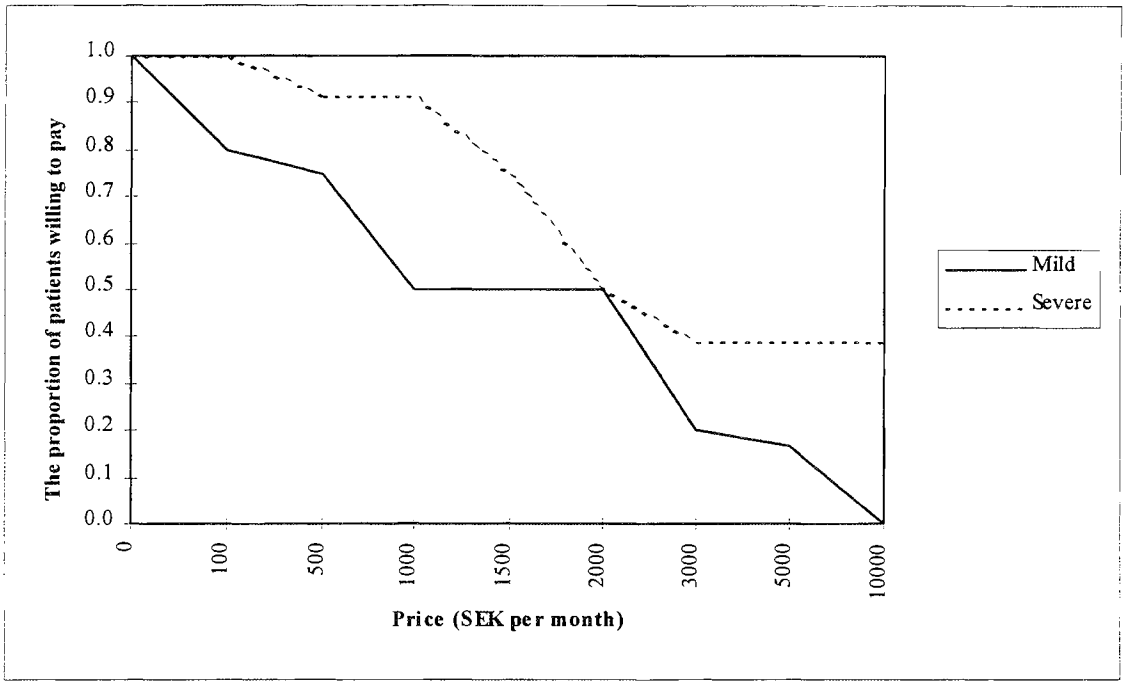
	RS			TTO		
	Without HRT	***With HRT***	***Difference***	***Without HRT***	***With HRT***	***Difference***
Mild	0.60 (0.21)	0.86 (0.13)	0.26 (0.21)	0.73 (0.24)	0.90 (0.21)	0.18 (0.21)
Severe	0.32 (0.18)	0.82 (0.16)	0.50 (0.25)	0.52 (0.27)	0.95 (0.12)	0.42 (0.29)
Total	0.47 (0.24)	0.84 (0.14)	0.37 (0.26)	0.63 (0.28)	0.92 (0.17)	0.29 (0.28)

Table 2: Mean and median willingness to pay (WTP) for hormone replacement therapy per month for women with mild and severe symptoms respectively. *N*=104 (Mild=56, Severe=48). £1= SEK 10.3.

	Mean WTP (SEK)	**Median WTP (SEK)**
Women with mild symptoms	2,346	1,500
Women with severe symptoms	4,838	2,000
All women	3,508	2,000

Figures

Figure 1: The relationship between the price level and the proportion of women willing to pay each price. *N*=104 (Mild=56, Severe=48).



Chapter 4

Estimating Hip Fracture Costs and Potential Savings*

Niklas Zethraeus and Ulf-G Gerdtham

Abstract: This paper examines the determinants of hip fracture costs and further evaluates potential savings in costs when the occurrence of hip fracture is prevented. The costs of hip fracture are comprised of direct costs from the health care and the social welfare system. Data are collected for 1,080 post-menopausal women admitted from an independent residence for primary hip fracture surgery during the year of 1992 in the city of Stockholm, Sweden. It is found that hip fracture costs are significantly related to age, mortality the year after a fracture, type of fracture, costs one year before a fracture, and hospital admission. The savings in direct costs for an average woman surviving the year after a fracture amount to SEK 210,000.

Keywords: Cost Savings, Determinants, Hip Fracture.

* Comments from Anders Odén, participants at the Health Economic Seminars at the Stockholm School of Economics and two anonymous referees for *International Journal of Technology Assessment in Health Care* are highly appreciated.

1. Introduction

Hip fracture is associated with the most severe morbidity and mortality of all osteoporotic or age-related fractures (Cummings 1993, OTA 1994). In Sweden, there are about 17,000 hip fractures each year (SCB 1993, Stockholm County Council 1990). Individuals at greatest risk of developing osteoporosis are post-menopausal women aged 50 years and older, who account for a major part of hip fracture patients (Cooper et al. 1992). The risk of osteoporotic fractures can be reduced, for instance, by altering certain behavioural factors, increasing calcium intake, and replacing deficient sex steroids[1] (Riggs and Melton 1992). Estimates of potential savings from preventing hip fractures are essential in economic evaluation studies of technologies, e.g. hormone replacement therapy (HRT), which decrease the fracture risk. During the last few decades attempts have also been made to lower the hip fracture treatment costs (Bauer 1985, Borgquist 1991). In order to be able to lower the treatment costs it is important to establish which factors determine the hip fracture costs.

Earlier studies focusing on hip fracture costs and potential cost savings have shown that fractures are associated with substantial costs for the health care system in particular and for society in general (Agarwal et al. 1986, Borgquist et al. 1991, Campion et al. 1987, Chrischilles et al. 1994, Hollingworth et al. 1993, Holmberg and Thorngren 1988, Sernbo and Johnell 1993). Previous studies suffer from several shortcomings. First, the time perspective is often limited, i.e. rehabilitation and long-term costs are seldom included. Second, these studies rarely considered the fact that individuals sustaining a hip fracture may have had a substantial consumption of both health care and municipal community care even if they did not sustain the fracture. Third, not one of the studies entertained the notion that hip fracture patients belong to a particular fragile group of individuals, suggesting that these patients in general may have a higher consumption of health care compared with a matched sample of individuals not sustaining a fracture (Sernbo 1988). If this fragility component cannot be accounted for in a matched sample this may imply that potential cost savings are overestimated. Fourth, when considering potential cost savings from the prevention of hip fracture, no distinction has explicitly been made between patients who survive and those who

[1] There are also other alternatives for reducing the risk of fractures such as hip protectors and weight bearing exercise.

die. The purpose of the present paper is to investigate the determinants of hip fracture costs and to provide estimations of potential savings in costs when hip fracture is prevented.

This study is designed to overcome the above listed shortcomings of earlier studies. First, the paper examines direct costs arising in the health care and social welfare systems during one year after hip fracture (costs for orthopaedic care, geriatrics, other acute hospital care, nursing homes, homes for the elderly, group residence and municipal home-help). Second, the savings in costs due to the prevention of hip fracture (extra hip fracture costs) are basically estimated as the difference between costs one year after and one year before the fracture (i.e. the patient is used as her own control). This approach accounts for the fact that hip fracture patients may belong to a fragile group of individuals and that the patients may have consumed resources even without fracture. Third, a problem in comparing the costs one year after and one year before fracture is that some of these patients die in the year after the fracture. Cost savings were therefore estimated separately for patients who survived and those who died within one year after fracture. The direct costs during a year after hip fracture are hereafter denoted as *hip fracture costs*, whereas savings in direct costs due to the prevention of hip fracture are denoted as *extra hip fracture costs*.

The present paper is organised into five sections. Section 2 presents the data and variables, Section 3 describes the methods used to analyse the hip fracture costs, Section 4 reports the results, and Section 5 discusses and concludes the paper.

2. Data

This paper focuses on *all* (1,080) post-menopausal women above the age of 49 years, admitted from a private residence, who sustained a hip fracture during the year 1992 in the city of Stockholm (OTA 1994). The patients were operated in one of five hospitals in the city of Stockholm[2]. The patients were followed one year before and one year after the fracture by combining two data bases: the inpatient data base of the Stockholm County Council (SCC) and the municipal community database. The costs are direct costs (costs for orthopaedic

[2] The five hospitals were: Danderyd Hospital (DS), Huddinge University Hospital (HS), Sabbatsberg Hospital (SABB), S:t Göran Hospital (SG), and Södersjukhuset (SÖS). The patients were mainly operated on using osteosynthesis, sliding screw-plate or screws.

department, geriatric care, other acute hospital care, nursing home, home for the elderly, group residence and municipal home-help) arising in the health care sector and the social welfare system (see *Appendix*)[3].

The values in *Appendix* are then multiplied by the average unit costs, which include both variable and fixed costs. It is important also to include fixed costs using a long-term perspective, because in the long run all costs become variable. (For example, by preventing fractures it may be possible in time to decrease the number of hospital beds and to eventually build smaller hospitals.) The average unit costs used in this paper are taken from a paper by Zethraeus et al. (1997). The average unit costs concerning orthopaedic and other acute hospital care are extracted from the Huddinge University Hospital patient-related accounting system, whereas the geriatric department at the Huddinge Hospital calculates the average unit cost for geriatric care. The average unit costs for nursing home, home for the elderly, group residence, and municipal home-help are collected from the social welfare authority. The costs are collected using a societal perspective. Due to limitations in the data it was not possible to include all relevant costs in the analysis. For example, indirect and outpatient costs are not included in the study. However, according to Borgquist et al. (1991), the outpatient cost (in primary health care) is a relatively small part of the total cost of hip fracture, amounting to about one percent of the total short-term cost within 4 months after hip fracture. However, it may still be true that costs in outpatient care are of importance during the year before fracture. Further, the percentage of patients who belong to the age group 50-64 years was only 3.9 percent (N=43), indicating that indirect costs may be of minor importance. Costs for the patient such as time and travelling costs are also excluded. To determine the importance of the excluded costs a separate investigation is needed. All costs are in SEK[4], expressed as arithmetic means, and refer to the prices of 1994.

[3] Data regarding days in orthopaedic departments, geriatric care and other acute hospital care are extracted from the inpatient database of the Stockholm county council, whereas data regarding days in a nursing home, home for the elderly, group residence and hours of municipal home-help are extracted from the municipal community database.

[4] 3 rd June 1997: US$1 = SEK 7.8.

The patient characteristics are depicted in *Table 1*, which also lists the included variables (*Table 1* in here). A total of 869 patients (80 percent) survived the first year after fracture and 211 patients (20 percent) died. In the morbidity and mortality groups, the direct costs within one year after fracture are higher than the costs one year before fracture. Note that costs one year before fracture for the mortality group exceed the costs for the surviving group. The costs one year after fracture are slightly higher for the morbidity group. However, for patients who died within the year after fracture, the mean period of survival is only 125 days. Thus, the cost per patient for each survival day is substantially higher in the mortality group (SEK 1,820) compared with the survival group (SEK 670). *Table 1* (TRADRES1) also shows that 66 percent of the women had returned to an independent residence while 20 percent had died within one year after hip fracture had occurred.

3. Methods

Treatment of a disease may be viewed in terms of a treatment cost function where the costs are related to output and input prices. In this paper we basically estimate two treatment cost functions: one for the treatment and rehabilitation of hip fractures during one year after fracture, and one for the treatment during one year before fracture. The costs for an individual during one year after hip fracture may be expressed as: $C_a^i(.) = \sum_z w_a^z x_a^z$, where z=1,2,3,...,Z and i=1,2,3,...,I. Z is the number of health care providers involved in the treatment and rehabilitation of hip fracture patients, while I is the number of treated patients. x_a^z is an input variable reflecting resource utilisation related to hip fracture for provider z, (e.g. hospital days in an orthopaedic department). w_a^z denotes the input prices at the time of the fracture for provider z. Analogously, the costs for a woman during one year before fracture may be viewed as: $C_b^i(.) = \sum_z w_b^z x_b^z$, where z=1,2,3,...,M, and i=1,2,3,...,I. M is the number of health care providers involved in the treatment of the women during the year before the fracture, while I is the number of treated patients. x_b^z is an input variable reflecting resource utilisation the year before fracture for provider z. w_b^z denotes the input prices at the time before fracture for provider z.

Econometric model specification

The costs one year before fracture (C_b^i) were estimated by the use of Cragg's model for a censored dependent variable, i.e. the costs were estimated in two steps (Cragg 1971). This model is a variant of the Tobit model where the probability of a nonlimit outcome is determined apart from the level of the nonlimit outcome. For our purposes this is an advantage compared to the Tobit model if it is suspected that the decision on whether to utilise the health care differs from the decision on how much to utilise, conditional on already utilising health care. Thus, first the probability of having positive costs one year before fracture was estimated using the following binary probit model:

$$\text{prob}(C_b^i > 0) = \Phi(a_0 + a_1*\text{AGE}_i + a_2*\text{AGE}_i^2 + a_3*\text{MORT}_i + u_i) \quad (1)$$

where Φ is the cumulative of the standard normal distribution, a_j, $j=0,1,2,3$, is a set of parameters to be estimated and u_i is the error term, which is conditionally independent of the regressors. Equation (1) separates users from nonusers and takes into account the fact that several women are nonspenders of medical services during the year before fracture.

Then, conditional on having positive costs the year before a fracture a regression model truncated at zero was estimated using maximum likelihood methods as[5]:

$$\log(C_b^i \mid C_b^i > 0) = b_0 + b_1*\text{AGE}_i + b_2*\text{AGE}_i^2 + b_3*\text{MORT}_i + v_i \quad (2)$$

Equation (2) describes the relation between the logarithm of the conditional costs one year before a fracture and the explanatory variables AGE, AGE^2 and MORT. The logarithmic transformation of the cost variable was employed to eliminate the skewness in the distribution of costs among users of medical services. AGE is assumed to affect the costs one year before fracture because of a greater health care consumption in higher ages. Also, a non-linear relationship was tested in which AGE^2 was included. MORT is a binary variable equal to 1 (0) if the patient dies (is alive) within one year after the fracture. This variable was included to account for the fact that health care costs are higher for patients close to death and may be

[5] A truncated regression model, truncated at zero, is used knowing that every patient has positive costs.

viewed as a proxy for the individual's health status (Zweifel 1995). The expected costs for a woman the year before a fracture ($E(C_b^i)$) was obtained by multiplying (1) with a re-transformation of (2); i.e. after (2) was antilogged, and can be expressed as:

$$E(C_b^i)=\text{prob}(C_b^i>0)\times E(C_b^i \mid C_b^i>0) \quad (3)$$

The logarithm of the costs one year after a fracture was estimated using maximum likelihood methods in a regression model truncated at zero:

$$\log(C_a^i \mid \text{fracture})=c_0+c_1*AGE_i+c_2*AGE_i^2+c_3*MORT_i+c_4*DIACODE_i+c_5*\log(COSTBEF_i)+\sum d_k*HOSP_{ik}+w_i \quad (4)$$

The model for the logarithm of costs after fracture is truncated as in the second step in the first model. We also estimated the logarithm of the conditional costs one year before fracture (the second step in Cragg's model) and the logarithm of the costs one year after fracture by ordinary least squares (OLS). Equation (4) describes the relationship between the logarithm of the costs one year after fracture and the explanatory variables AGE, AGE^2, MORT, DIACODE, log(COSTBEF) and $HOSP_k$. One effect of mortality on costs the year after a fracture is that the costs per unit of time may increase, implying higher costs one year after fracture. An opposite effect of mortality is that the time period is shorter, implying a lower cost one year after fracture. The strongest of these effects determines the effect of mortality on the costs one year after fracture. DIACODE was included as an explanatory variable in that it has been indicated that trochanteric fractures are associated with higher costs compared with cervical fractures (Borgquist et al. 1991). COSTBEF may reflect patient morbidity, and a patient who has high costs the year before a fracture is assumed to have higher costs the year after fracture compared to a patient with low costs the year before. Finally, $HOSP_k$ was included to account for differences in efficiency between hospitals that may affect the costs after fracture. Hospital efficiency may depend on e.g. how treatment and after care at different hospitals are organised.

One problem in comparing the costs between the year after and the year before fracture is that some women die in the year after fracture (i.e. for women who died, different time periods are compared). The extra hip fracture costs were therefore estimated separately for patients who

survived and died in the year after fracture. Conditional on surviving, the extra costs were estimated for different ages and were estimated by subtracting equation (3) from the antilog of equation (4). The extra costs were estimated with the other explanatory variables set at their sample means. Thus, conditional on surviving, the extra hip fracture costs were estimated for average women of different ages. The mean extra costs for women who died in the year after fracture were estimated by comparing the costs after the fracture with the costs before the fracture for the same time period. For instance, for a woman who died 4 months after fracture, the costs for 4 months before the fracture were subtracted from the costs after the fracture.

4. Results

Our results are reported in three parts. The first part estimates the costs one year before fracture. The second part estimates the costs one year after fracture. The third part estimates the extra hip fracture costs (difference between costs one year after and one year before fracture) for patients surviving the year after hip fracture.

4.1 Estimating the costs one year before hip fracture

The probability of having positive costs the year before hip fracture is significantly related to AGE^2 ($p<0.05$) and MORT ($p<0.01$) in the year following fracture (*Table 2*). Conditional on the survival status the year after fracture, the probability of having positive costs the year before fracture is relatively constant up to 60 years of age and thereafter increases: first at an increasing rate up to the age of 85 years, and then at a decreasing rate until the age of 100 years. A 1 year increase in age for an average woman results in a 3% increase in the probability of having positive costs. Conditional on age, the probability of having positive costs the year before fracture shifts upwards if a patient dies within the year after fracture (i.e. patients who die within one year after fracture have a greater probability of utilising resources in the health care and social welfare systems the year before fracture compared with patients surviving the year after fracture). An average woman who dies during the year after fracture has a 28% higher probability of having positive costs the year before fracture compared to a woman who survives (*Table 2* in here). *Table 2* further shows that, conditional on having positive costs the year before fracture, the logarithm of the costs one year before fracture is

significantly related to AGE ($p<0.05$), AGE^2 ($p<0.05$) and MORT ($p<0.01$)[6]. Note that this second step in the Cragg model applies to the number of patients having positive costs (N=704). Conditional on survival status the year after fracture, the conditional costs before fracture decrease up to the age of 76 years and increase thereafter. A 1 year increase in the age for an average woman having positive costs results in a 2% increase in the conditional costs. Conditional on age, the conditional costs before fracture shift upwards if a patient dies within the year after fracture. An average woman who dies during the year after fracture has 66% higher conditional costs compared to a woman who survives.

The estimated costs for an average woman the year before fracture is obtained by multiplying the probability of having positive costs by the conditional costs (i.e. given that they are positive). The expected costs for an average woman the year before fracture decreases up to the age of 70 years and then increases until 100 years of age. Expected costs one year before fracture is positively related to mortality, indicating that women who die the year after fracture are subject to a greater morbidity the year before fracture compared with women who survive. An average woman who dies during the year after fracture has 113% higher expected costs compared to a woman who survives. The goodness of fit measure for models 1 and 2 in *Table 2* shows that the variation in the dependent variable is badly explained by the included explanatory variables. This is also the case when model 2 is estimated by OLS with a resulting R^2 of 0.03.

4.2 Estimating the costs one year after hip fracture

The estimated costs one year after fracture are an increasing function of age, although the partial effect of AGE and AGE^2 on costs one year after fracture was not statistically significant[7]. However, we tested for joint significance the null hypothesis that the effects of AGE and AGE^2 on costs one year after fracture are zero. Using a Wald-test, the $\chi^2(2)$-statistic was 49.66 which is significant at the 1% level and the null hypothesis could be rejected[8]. A 1

[6] Conditional on having positive costs, the costs one year before fracture were also estimated using ordinary least squares (OLS). The estimated parameter values were the same as in the truncated model but the standard errors were higher.

[7] Costs one year after fracture were also estimated using ordinary least squares (OLS). The estimated parameter values were the same as in the truncated model but the standard errors were higher.

[8] In an unreported model we excluded AGE from the model and found that AGE^2 was significant at the 1% level.

year increase in age for an average woman results in a 3% increase in the costs one year after fracture. MORT is negatively related to costs within one year after fracture ($p<0.01$), showing that the costs are lower for patients who die compared with patients who survive. An average woman who dies has 35% lower costs compared to a surviving woman. Further, trochanteric fractures are associated with higher costs ($p<0.10$) compared to cervical fractures during the year after fracture. An average patient having a trochanteric fracture has 10% higher costs compared to a patient with a cervical fracture. The costs the year after fracture is positively related to the amount of costs one year before fracture ($p<0.01$). A 1% increase in the costs one year before results in a 0.08% increase in the costs one year after fracture. Hospital 4 is associated with lower costs than Hospital 1 ($p<0.10$). A patient admitted to hospital 4 has 17% lower costs compared to a patient admitted to hospital 1. The estimated costs one year after fracture increase for an average individual from SEK 142,000 for a 50-year old woman to SEK 406,000 for a 100-year old woman. Although the explained variation in the dependent variable in model 3 is relatively low, the variation is better explained by the included variables in model 3 compared to models 1 and 2. This is also the case when model 3 is estimated by OLS with a resulting R^2 of 0.26.

4.3 Extra hip fracture costs

Figure 1 shows that the expected costs the year before fracture are higher for an average woman who dies during the year after a fracture than for one who survives (*Figure 1* in here). This can be explained by a higher degree of morbidity for the mortality patients the year before fracture. On the other hand, the expected costs the year after fracture are higher for an average woman who survives the year after fracture than for an average woman who dies. Thus, the higher costs that are due to a longer survival overshadow the lower costs due to a lesser cost per day in the morbidity group compared with the mortality group.

Surviving patients

The age dependent extra hip fracture costs are calculated separately for surviving women and for women who die during the year after fracture. As *Figure 1* shows, the costs one year after fracture always exceeded the costs one year before fracture for an average woman surviving the first year after fracture. Thus, avoiding hip fracture implies cost savings for women of all

ages in this sub-group ranging from SEK 107,000 for a 50-year old woman to SEK 346,000 for a 100-year old woman. *Figure 2* shows that the predicted extra costs increase from 50 years up to the age of 100 years. The mean age in the surviving group is 81 years and the predicted cost savings at this age are SEK 212,000 (*Figure 2* in here).

Dying patients

For patients dying within a year after fracture the costs one year after fracture exceed the costs one year before fracture in the age range 57 to 100 years (*Figure 1*). However, the extra costs in the mortality group cannot be calculated as the difference between costs one year after fracture and one year before. This would imply that different time periods are compared for the mortality patients (e.g., a patient surviving one month would be compared with her costs over an entire year). In the mortality group, the mean time of survival is 125 days the year after fracture. This is further depicted in *Figure 3*, which shows that the costs per survival day in the mortality group are higher in the period after fracture compared with the period before (*Figure 3* in here). The mean cost per surviving day in the mortality group during the year after fracture is SEK 1,820, which can be compared with the mean cost per surviving day of SEK 390 in this group during the year before fracture.

5. Discussion and Conclusion

This paper investigated the determinants of hip fracture costs and estimated savings in costs when hip fractures are prevented. The paper focused on 1,080 post-menopausal women above the age of 49 years who were admitted to hospital from a private residence. All patients were residents of the city of Stockholm. The costs consisted of direct costs from the health care and social welfare systems during one year after fracture (i.e. rehabilitation and long-term costs were also included). The study showed that hip fracture costs for post-menopausal women were significantly related to age, mortality the year after fracture, type of fracture, costs one year before fracture, and hospital admission (as compared to a benchmark hospital). The estimated hip fracture costs for an average woman between 50 and 100 years ranged from SEK 140,000 to SEK 410,000. For women who survived, cost savings were present in all ages and ranged from SEK 110,000 to SEK 350,000. These results are similar to results presented in previous studies. For example, Sernbo and Johnell (1993) show that the costs of a hip

fracture in the first year after the fracture were SEK 200,000 and that the extra costs of a hip fracture were SEK 160,000. Further, Borgquist et al. (1991) showed that the hip fracture costs during 4 months after a hip fracture amounted to SEK 81,000.

The improvements in this study relative to previous studies are several. First, it was found that savings in costs when preventing fracture would be exaggerated if they were estimated as the costs during one year after fracture, which is explained by a considerable consumption of resources without fracture. Thus, when estimating cost savings it is important to consider the costs that would arise if the fracture had not occurred. Second, to estimate savings in costs if fracture is avoided the woman was used as her own control. The advantage of this approach over an approach where a matched patient sample is used is that we take into account the fact that hip fracture patients belong to a particularly fragile group (Sernbo 1988). If it is not possible to account for the fact that hip fracture patients belong to a fragile group of patients using a matched patient sample, it would imply that costs without a fracture are underestimated and cost savings overestimated. Third, the best method for estimating extra hip fracture costs for patients dying within a year after fracture largely depends on whether the mortality risk following fracture is affected or not. If we assume that the mortality risk is not affected by hip fracture, then the extra costs can be estimated by subtracting the costs before fracture over the same time period that the patient survives after fracture. For example, for a woman who died 4 months after the fracture, the costs for 4 months before the fracture should be subtracted from the costs after fracture to obtain the extra costs. The extra costs for an average woman dying during the year after a fracture are then estimated to be SEK 180,000. On the other hand, if the mortality risk is increased by the fracture, then avoiding fractures implies additional costs for the health care system due to longer survival. Whether these costs, which are a result of a longer life, should be included or not is controversial (Weinstein 1990). However, previously it has been suggested that costs in added life years should be included in economic evaluation studies of health care programmes (Meltzer 1997). The costs in added life years are equal to changes in unrelated and related medical costs and consumption minus changes in the production due to changes in mortality produced by the programme.

Although this study involves several improvements relative to previous studies, it suffers from some limitations. First, due to limitations in the databases not all costs relevant from a societal

perspective are included. Costs that are excluded from the study are indirect costs, outpatient costs and costs for the patient such as time and travelling costs. Another limitation is that the included costs were collected for just one year after fracture and the study does not reveal how the costs will develop in the long run. In order to extend the follow-up time and to capture the excluded costs so as to determine their importance, a special investigation is needed. This is a matter for future research.

6. References

Agarwal N, Reyes JD, Westerman AD, Cayten CG. Factors influencing DRG 210 (hip fracture) reimbursement. *Journal of Trauma* 1986; **26(5)**: 426-431.

Bauer GCH. Orthopaedic technology for the elderly. *International Journal of Technology Assessment in Health Care* 1985; **1**: 59-74.

Borgquist L, Lindelöw G, Thorngren K-G. Costs of hip fracture - Rehabilitation of 180 patients in primary health care. *Acta Orthopaedica Scandinavia* 1991; **62(1)**: 39-48.

Campion EW, Jette AM, Cleary PD, Harris BA. Hip fracture: A prospective study of hospital course, complications, and costs. *Journal of General Internal Medicine* 1987; **2**: 78-82.

Chrischilles E, Shireman T, Wallace R. Costs and health effects of osteoporotic fractures. *Bone* 1994; **15**: 377-386.

Cooper C, Campion G, Melton LJ. Hip fractures in the elderly: a world-wide projection. *Osteoporosis International* 1992; **2**: 285-289.

Cragg JG. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 1971; **39**: 829-844.

Cummings SR. Bone mass and bone loss in the elderly: A special case? *International Journal of Fertility* 1993; **38(suppl 2)**: 92-97.

Hollingworth W, Todd C, Parker M, Roberts JA, Williams R. Cost analysis of early discharge after hip fracture. *British Medical Journal* 1993; **307**: 903-906.

Holmberg S, Thorngren K-G. Consumption of hospital resources for femoral neck fracture. *Acta Orthopaedica Scandinavia* 1988; **59(4)**: 377-381.

Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997; **16**:33-64.

OTA (Office of Technology Assessment). *Hip fracture outcomes in people age fifty and over - background paper*. Washington DC: US Congress, OTA, 1994, OTA-BP-H-120.

Phillips S, Fox N, Jacobs J, Wright WE. The direct medical costs of osteoporosis for American women aged 45 and older, 1986. *Bone*, 1988; **9**: 271-279.

Riggs BL, Melton LJ. The prevention and treatment of osteoporosis. *New England Journal of Medicine* 1992; **327**: 620-627.

SCB, Population statistics, part 3. Stockholm: Statistics Sweden, 1993.

Sernbo I. *Hip fracture*. Thesis, Lund University, Malmö, Sweden, 1988.

Sernbo I, Johnell O. Consequences of a hip fracture: a prospective study over 1 year. *Osteoporosis International* 1993; **3**: 148-153.

Stockholm County Council. Stockholm inpatient register. 1990.

Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *International Journal of Technology Assessment in Health Care* 1990; **6**: 93-103.

Zethraeus N, Strömberg L, Jönsson B, Svensson O, Öhlén G. The cost of a hip fracture. Estimates for 1,709 patients in Sweden. *Acta Orthopaedica Scandinavica* 1997; **68(1)**: 13-17.

Zweifel P, Felder S, Meier M. Ageing of population and health care expenditure: a red herring? Paper presented at "Third European Conference on Health Economics", The Stockholm School of Economics, Sweden August 20-22, 1995.

Appendix

Mean number of days or hours within one year after and one year before the occurrence of hip fracture. Standard deviations are given in parentheses. *N*=1,080.

	1 year Before	**1 year after**
Days in department of orthopaedics	1.2 (5.6)	14.6 (11.7)
Days in geriatrics	5.2 (21.6)	30.8 (46.9)
Days in other acute hospital care	5.2 (24.0)	4.8 (21.5)
Days in nursing home	0.3 (4.6)	28.2 (75.5)
Days in home for the elderly	0.2 (4.5)	3.2 (23.9)
Days in group residence	- -	2.6 (24.8)
Hours of municipal home-help	218.2 (421.7)	228.7 (376.5)

Tables

Table 1: Characteristics of hip fracture patients and included variables. Standard deviations are given in parentheses.

VARIABLES		Total	Alive	Dead
Abbreviation	*Interpretation*	(*N*=1, 080)	(*N*=869)	(*N*=211)
COSTBEF	Cost one year before fracture	89,600 (161,120)	77,170 (156,120)	140,920 (171,330)
COSTAFT	Cost one year after fracture	243,870 (209,630)	247,640 (212,280)	228,340 (198,060)
COSTDIFF	Difference in costs one year after and one year before fracture	154,250 (189,810)	170,480 (174,740)	87,420 (230,950)
AGE	Age at admission	81.6 (8.1)	80.8 (8.1)	85.0 (7.1)
DIACODE	Dummy variable, 1 for Trochanteric, zero otherwise (Cervical)	0.52 (0.5)	0.51 (0.5)	0.58 (0.49)
MORT	Dummy variable, 1 for dead one year after fracture, zero otherwise (Alive)	0.20 (0.40)	/ /	/ /
TRADUT	At discharge after initial orthopaedics (percent):			
	Institution	50.1	48.2	57.8
	Independent residence	47.2	51.8	28.4
	Dead	2.7	/	13.7
TRADRES1	1 year after fracture (percent):			
	Institution	14.6	18.2	/
	Independent residence	65.8	81.8	/
	Dead	19.5	/	100.0
HOSP1 (DS)	Benchmark hospital	0.10 (0.3)	0.10 (0.3)	0.09 (0.3)
HOSP2 (HS)	One for hospital 2, zero otherwise (benchmark Hospital 1)	0.10 (0.3)	0.10 (0.3)	0.09 (0.3)
HOSP3 (SÖS)	One for hospital 3, zero otherwise (benchmark Hospital 1)	0.36 (0.5)	0.35 (0.5)	0.41 (0.5)
HOSP4 (SABB)	One for hospital 4, zero otherwise (benchmark Hospital 1)	0.24 (0.4)	0.24 (0.4)	0.24 (0.4)
HOSP5 (STG)	One for hospital 5, zero otherwise (benchmark Hospital 1)	0.20 (0.4)	0.20 (0.4)	0.18 (0.4)

Table 2: Results of estimations for women admitted from a private residence; t-ratios are shown in parentheses.

REGRESSOR VARIABLES:	Costs one year before fracture *Binomial probit Model*	Costs one year before fracture *Truncated regression model*	Costs one year after fracture *Truncated regression Model*
DEPENDENT	0/1	log(COSTBEF)	log(COSTAFT)
INTERCEPT	2.3239	20.055***	11.866***
	(0.93)	(4.76)	(7.52)
AGE	-0.096274	-0.24739**	-0.035203
	(-1.49)	(-2.34)	(-0.87)
AGE^2	0.00087204**	0.0016242**	0.00037484
	(2.09)	(2.45)	(1.45)
MORT	0.49561***	0.50981***	-0.43398***
	(4.34)	(3.43)	(-6.42)
DIACODE	-	-	0.095548*
	-	-	(1.83)
log(COSTBEF)	-	-	0.080051***
	-	-	(15.85)
HOSP2	-	-	-0.082752
	-	-	(-0.72)
HOSP3	-	-	0.020490
	-	-	(0.22)
HOSP4	-	-	-0.18983*
	-	-	(-1.95)
HOSP5	-	-	-0.057291
			(-0.57)
$\hat{\sigma}$	-	1.663	0.844
LRI	0.072	0.00845	0.110
L	-647.46	-1356.87	-1349.62
N	1,080	704	1,080

Notes:

***,**,* Significant at the 1%, 5%, and 10% levels, respectively.

$\hat{\sigma}$ *:* Estimated standard deviation in the model.

LRI: Likelihood ratio index. LRI=1-[L(general)/L(restricted)], where L(general) is the maximum likelihood value of the log-likelihood function, and L(restricted) is the maximum likelihood value of this function under the constraint that the effects of the explanatory variables are zero.

L: The value of the log-likelihood function.

Figures

Figure 1: Predictions of expected costs one year before and one year after a hip fracture, conditional on survival status the year after a fracture for an average woman admitted from a private residence, by age.

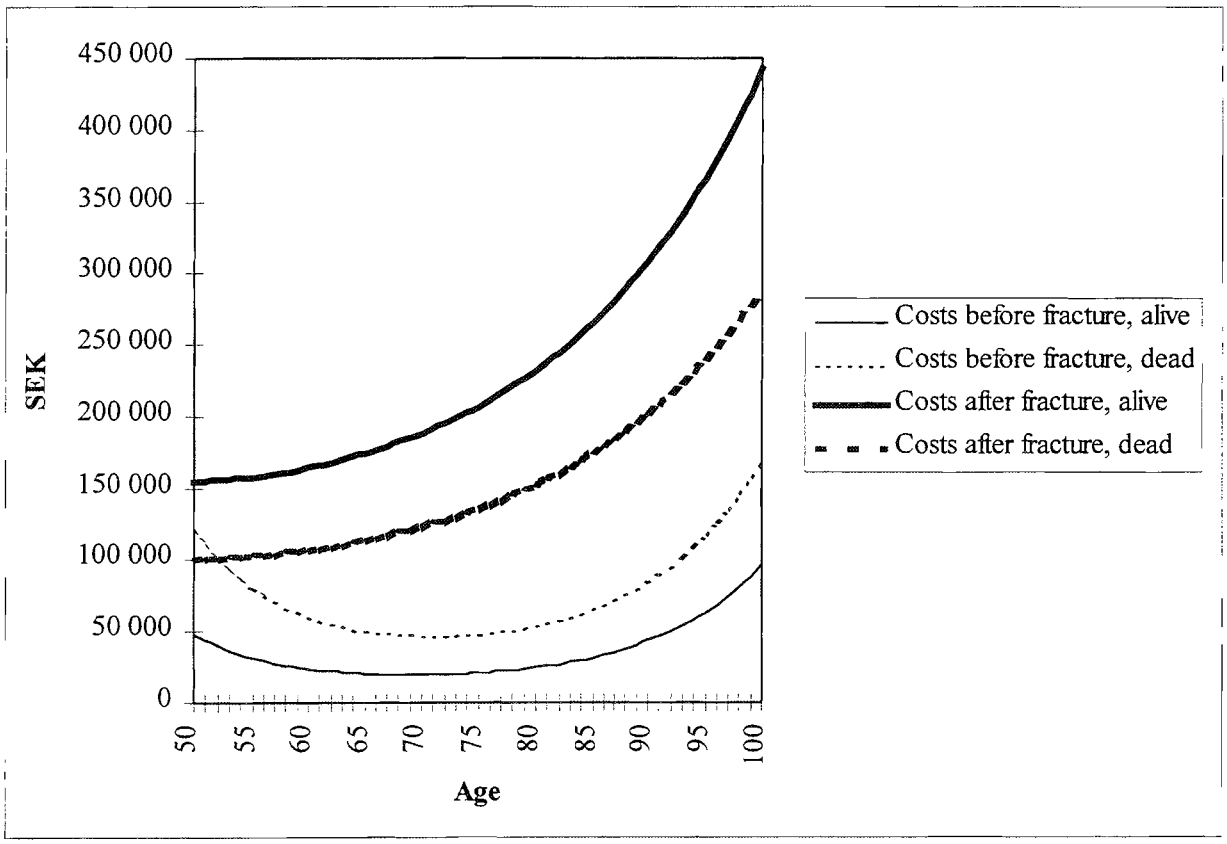


Figure 2: Predicted extra hip fracture costs for an average woman surviving the first year after a fracture, by age. *N*=869.

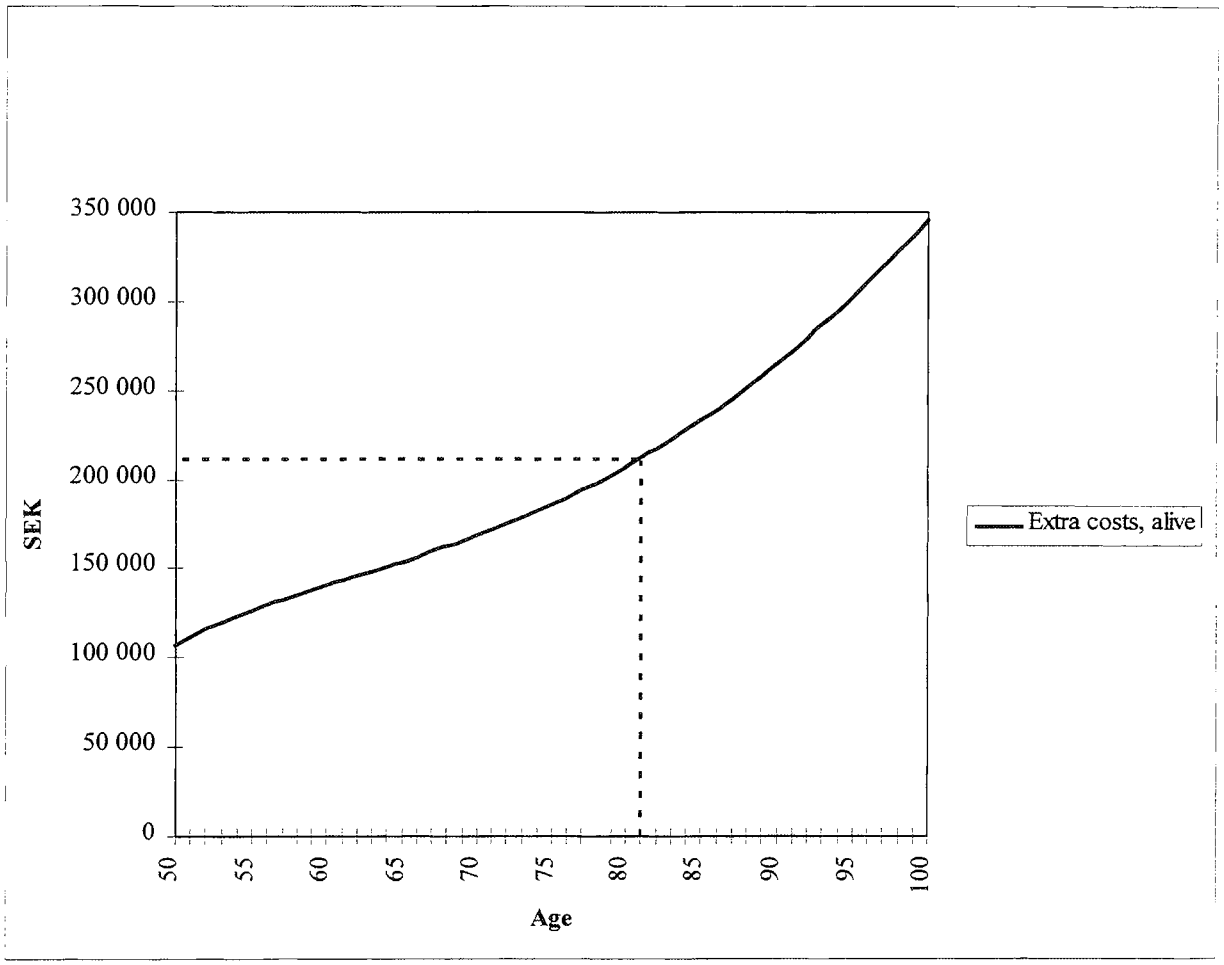
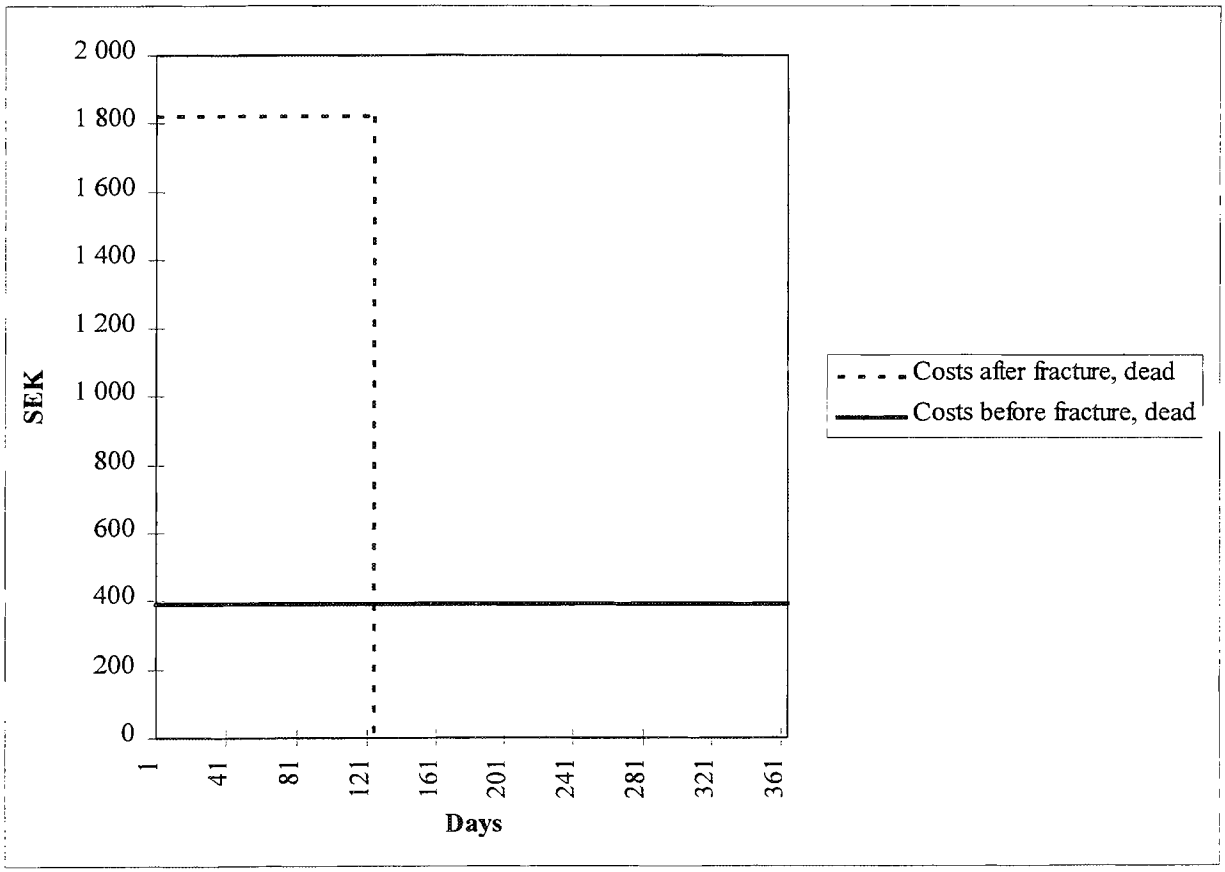


Figure 3: Costs per survival day one year after and one year before a hip fracture for patients dying in the year after the fracture.



Chapter 5

A Computer Model to Analyse the Cost-Effectiveness of Hormone Replacement Therapy*

Niklas Zethraeus, Magnus Johannesson and Bengt Jönsson

Abstract: This paper gives a detailed presentation of a computer model for evaluating the cost-effectiveness (CE) of hormone replacement therapy (HRT). The paper describes the model's design, structure and data requirements. The model needs data specified for costs, quality of life, risks and mortality rates. As an illustration, the CE of HRT in Sweden is calculated. Two treatment strategies were evaluated for asymptomatic women: Oestrogen only therapy and Oestrogen combined with a progestogen. The model produces similar results compared with earlier studies. The CE ratios improve with age and the size of the risk reduction. Further, oestrogen only therapy is associated with a lower cost per gained effectiveness unit compared to combined therapy. Uncertainty surrounding the long-term effects of HRT means that the CE estimates should be interpreted carefully. The model permits the inclusion of indirect costs and costs in added life years, which allows the analysis to be made from a societal perspective, which is an improvement relative to previous studies.

Keywords: Cost-Effectiveness, Evaluation, HRT, Markov Model.

* Comments from participants at the Health Economic Seminars at the Stockholm School of Economics are highly appreciated. We are also grateful to Anja Launila for helping us developing the computer model.

1. Introduction

The computer model is designed to analyse the cost-effectiveness (CE) of hormone replacement therapy (HRT)[1] in the prevention and treatment of postmenopausal women's health problem. The consequences of this intervention in terms of changes in quality of life during treatment and for effectiveness in terms of reduction or increase in cardiovascular events, hip fracture and breast cancer are simulated and related to costs. Alternative interventions that affect one or more of the risks mentioned above can also be studied and compared with HRT. This paper describes the model's structure, data requirements and different outputs it can provide. Estimations based on Swedish cost and risk data are included to illustrate how the model works, its internal consistency, and how the calculated results are related to previous studies. The end of the paper discusses opportunities and data requirements for extending the model to other countries. This section also explains how the model can calculate CE when several interventions are compared simultaneously.

At menopause which occurs on the average at age 50, a majority of women (about 75%) experience menopausal symptoms such as hot flushes, night sweats and atrophy-related symptoms of the urogenital tract. Menopausal symptoms may substantially decrease a woman's quality of life (Daly et al. 1993, Zethraeus et al. 1997). HRT mitigates or eliminates these symptoms and increases quality of life. HRT may also have a cardioprotective effect and offers protection against osteoporosis and related fractures (SBU 1996). Evidence of the effect HRT has on breast cancer is inconclusive; although, the risk is assumed to increase after a long period of treatment (Beral et al. 1997, Colditz et al. 1995, OTA 1995, SBU 1996, Stanford et al. 1995). For non-hysterectomised women taking oestrogens only, an increased risk of endometrial cancer is evident (SBU 1996). The increased risk of endometrial cancer is decreased or eliminated by the addition of a progestogen (Persson et al. 1989, SBU 1996). Combining oestrogen with a progestogen may induce uterine bleedings, however, such bleedings may reduce or vanish if a combined HRT is continuously applied although break through bleeding often occurs in the first few months (Andersson et al. 1992, SBU 1996).

HRT has been used for treating menopausal symptoms for a long time. Also, in the last few years, HRT has been recommended for women at a high risk of osteoporosis-related fractures. Whether HRT can be recommended for asymptomatic women as a preventive treatment has also

[1] If nothing else is said HRT refers both to oestrogen only therapy and oestrogen combined with a progestogen.

been discussed (SBU 1996). From a health economic perspective these and other issues may be considered by using cost-effectiveness analysis (CEA).

CEA is based on maximising health effects, which are subject to a cost constraint (Weinstein and Zeckhauser 1973). In a CEA, costs are measured in monetary units and effects in non-monetary units such as life-years gained (LYG) or quality adjusted life-years (QALYs). The CEA is usually denoted cost-utility analysis (CUA) if the health effects take into account changes in quantity and quality of life. The most frequently used health outcome measures in CUA are QALYs; although, healthy years equivalents (HYEs) have also been proposed (Drummond et al. 1997). CEA must be provided with a useful decision rule such that the price per effectiveness unit must be determined (e.g., the willingness to pay for a QALY or a life year gained). Without information about the price per effectiveness unit, a CEA gives no information about whether a program should be implemented or not unless the intervention is a dominated alternative such that the programme has higher costs and lower effects. Furthermore, if a fixed price is used as a decision rule, the CEA approach can be seen as a special case of cost-benefit analysis where the price per QALY is constant (at all levels of change) and the same for everyone (Johannesson 1996).

The CE of HRT is often modelled due to uncertainty surrounding the long-term effects of HRT. One model frequently used in the CE literature is the Markov model (Keeler 1995). The Markov model is useful when a decision problem involves risk that is ongoing over time, when the timing of events is important and when important events may happen more than once (Sonnenberg and Beck 1993). This model is defined using a finite number of (health) states in which an individual may be found at any given time. The states are mutually exclusive and collectively exhaustive, meaning that an individual must be in exactly one of the states at any time. The model assumes that all individuals in a specific state are identical and that each individual obtains the same cost or benefit irrespective of which transitions led to the health state; i.e. the model has no memory of prior states (Sonnenberg and Beck 1993). Markov models occur in a discrete time frame and time progresses in units of arbitrary, but fixed length (e.g., one year) called 'cycles'. A transition occurs when an individual moves from one state to the next. Transitions among states occur instantaneously, but often a half cycle correction is included. The transition probability (p_{ij}) is the probability of going from state i to state j; the transition probabilities for exiting a specific state at a particular stage must always sum to one: i.e., $\sum_j p_{ij} = 1$. Three basic methods exist to evaluate a Markov model. The first method is a Monte

Carlo simulation in which an individual is walked through the process many times. The second way is to use a cohort simulation whereby a large group of individuals (e.g., 1,000 individuals) are filtered through the model at the same time, choosing their transitions according to decided distributions. The third method involves matrix algebra and produces an analytical solution (Sonnenberg and Beck 1993). The advantage of using a cohort or Monte Carlo simulation is that transition probabilities, benefits and costs may be viewed as a function of not only the health state, but also other populations characteristics such as age.

Previous studies analysing the cost-effectiveness of HRT suffer from several shortcomings. First the underlying model which the analyses are based upon are seldom explicitly presented or is just briefly explained. Second the analyses never include indirect costs or costs in added life years which means that the analyses are not based on a societal perspective. Third the analyses are often based on assumptions and not on empirical investigations. This paper gives a detailed presentation of the computer model that also allows for the inclusion of indirect costs and costs in added life years. In the empirical application of the model, intervention costs, morbidity costs for hip fracture and coronary heart disease (CHD) are based on empirical studies.

2. The Computer Model

The computer model is programmed in *C++* and built as a Markov model around menus in a Microsoft Windows environment[2]. The computer model is developed to analyse the CE of HRT and is evaluated using a cohort simulation. The model integrates two previously described computer models: one used for cardiovascular-disease prevention and one for fracture prevention (Johannesson et al. 1991, Jönsson et al. 1993, Jönsson et al. 1995). The model also includes a risk function for breast cancer.

2.1 The design and structure of the model

The model's overall structure showing the included health states are illustrated in *Figure 1*. These basic health states are: 1. Healthy; 2. Hip fracture first year; 3. Hip fracture following years; 4. Breast cancer first year; 5. Breast cancer following years; 6. CHD first year; 7. CHD following years; and 8. Death. CHD is subdivided into five health states: 6.1. Recognised acute myocardial infarction, 6.2. Unrecognised acute myocardial infarction, 6.3. Angina pectoris, 6.4. Coronary insufficiency[3]; and 6.5. Sudden death. The health states numbered 2-7 are also denoted

[2] For a more thorough presentation of the model and the menus see Zethraeus et al. (1998).
[3] Coronary Insufficiency or Unstable Angina Pectoris can be used interchangeably.

disease states and their inclusion is motivated from the medical literature showing that HRT may affect these disease risks (SBU 1996). Each disease state is characterised by age-dependent mortality rates, costs and quality of life weights. Hip fractures, breast cancer and CHD are divided into "first" and "second and following years" after a disease event since mortality rates, costs, and quality of life differ between these time periods. When a disease event occurs, the patient will stay in that state or transition until "death". At present, there are no transitions between health states after an event such as hip fracture to CHD or CHD to breast cancer. Solving this problem can be done in two ways: One way is to introduce new states such as a hip fracture after CHD. The problem is that the model becomes very complicated and difficulties with data arise. An alternative is to include the risks and costs of the other two diseases in the sequel after an event. The latter approach has been taken in this model (*Figure 1* in here).

The basic model structure assumes a healthy cohort of individuals in its initial population group (the cohort size can vary between 1- 100,000); whereby, 'healthy' means free from CHD, breast cancer and hip fractures. At each cycle of the process, the cohort is reallocated to health states according to specified transition probabilities. All transitions are assumed to occur instantaneously halfway through each cycle. In the first cycle the cohort is exposed to disease risks of CHD, breast cancer and hip fractures as well as the risk of dying from other causes. A patient experiencing a disease event can only transit to death or "post-disease event". Patients in "post-disease events" can only remain in that state or transit to death. The cohort is followed until age 110. The disease risk function is specified as a logistic distribution function including different risk factors (Gujarati 1988). The disease risk function can be expressed as:

$$p_i = \frac{1}{1+e^{-Z_i}} \tag{1}$$

$$\text{where } Z_i = \alpha_0 X_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \ldots + \alpha_n X_{ni} \tag{2}$$

p_i is the risk of the disease during a cycle; $X_0 \ldots X_n$ are risk factors and $\alpha_0 \ldots \alpha_n$ are parameters to be estimated. The risk of hip fracture and breast cancer are estimated using Swedish incidence data in different age groups (National Board of Health and Welfare 1993, Stockholm County Council 1990); whereas, the risk of CHD is extracted from the Framingham Heart Study in which the results are based on a US population (Kannel et al. 1987). The model can also tabulate the risks instead of using the risk functions.

The CE formula used in the computer model can be expressed as:

$$\frac{\Delta C}{\Delta E} = \frac{C_1 - C_0}{E_1 - E_0} = \frac{\Delta INT + \Delta MORB + \Delta MORT}{\Delta QLE} = \frac{\Delta INT + \Delta MORB + \Delta MORT}{\Delta LE + \Delta LEQ}, \qquad (3)$$

where a subscript 0 (1), referring to C_i and E_i, denotes no intervention (with intervention), where $i = 0,1$.

ΔINT = Intervention costs, direct and indirect.
$\Delta MORB$ = Changes in morbidity costs, direct and indirect, due to the intervention.
$\Delta MORT$ = Changes in mortality costs, direct and indirect, due to the intervention.
ΔLE = Changes in life expectancy due to the intervention.
ΔLEQ = Changes in quality of life measured in years due to the intervention (where 'quality of life' refers to changes in morbidity and side effects).
$\Delta QLE = \Delta LE + \Delta LEQ$

The numerator in the above formula represents the discounted change in costs resulting from an intervention. The denominator is the discounted change in effectiveness generated by the intervention. The change in costs and effectiveness resulting from the intervention is compared to a baseline alternative, i.e., no intervention. The change in cost is based on the sum of changes in intervention, morbidity and mortality costs generated through the intervention; whereas, the change in effectiveness is based on the sum of changes in life expectancy and quality of life measured in years due to the intervention. The model permits the CE ratio to be expressed either as costs per LYG (if ΔLEQ is set to zero) or costs per QALYs gained. As the model incorporates consequences for different diseases, effectiveness measures, such as number of events avoided from an intervention, do not provide meaningful information. Instead a composed outcome measure is needed, which incorporates the interventions effectiveness for different risks.

Intervention costs (ΔINT) are divided into yearly and initial costs. Yearly costs consist of direct and indirect costs. Direct costs for an intervention include: cost of drug, costs for services in hospitals (physician visits), primary health care and travelling costs. Indirect costs reflect resources foregone due to the treatment (e.g., production losses). These costs are particularly relevant for primary prevention when healthy time is used for the interventions (e.g., physician

visits). Initial costs consist of direct and indirect costs and may, for example, be costs for screening patients to be treated.

Changes in morbidity costs (*ΔMORB*) consist of costs saved because of reduced morbidity from CHD and hip fractures and costs added because of increased morbidity from breast cancer. The change in morbidity costs are divided into changes in direct and indirect costs. The model also permits the inclusion of changes in mortality costs (*ΔMORT*). Changes in mortality costs are equal to changes in total consumption minus changes in the total production due to a change in mortality from the intervention (Meltzer 1997). The estimation of consumption and production should in principle be based on a healthy population which are free from hip fractures, breast cancer and CHD.

2.2 Modelling an intervention

An intervention is modelled by its impact on the disease risks (*Figure 2*). The example illustrated in *Figure 2* assumes that treatment duration lasts 10 years. Without treatment, the relative risk (RR) is equal to 1. With treatment, the RR follows the dotted line. The risk reduction is entered as a percentage change in the baseline risk. For example, if the risk of CHD is assumed to be reduced by 40% during HRT, this is equal to multiplying equation (1) for CHD above by 0.6. According to the CHD risk equation, the risk without treatment of CHD for a woman of age 60 years is 6.4 per 1,000. The intervention reduces the risk by 40% and the resulting risk of CHD for this woman with treatment is then equal to: $0.6 \times \frac{6.4}{1{,}000} = \frac{3.84}{1{,}000}$ (*Figure 2* in here).

Different options are available for the user when modelling disease risks affected by the intervention. Start delay is defined as the time prior to when the intervention affects the risk (2 years in *Figure 2*). Rise time is defined as the time it takes from the end of the start delay until the risk reduction has reached its maximum value (2 years in the *Figure 2*). The rise time is specified as a linear function of time. Stop delay and set time are defined analogously to start delay and rise time. The model also permits a remaining effect lasting from the end of set time until the rest of the lifetime. Thus, the model allows the user to make several different assumptions about how the intervention affects the disease risks.

2.3 Data for the model

The model demands data about risks, mortality rates, quality of life weights, and costs.

Risks

First, the base-case risk of CHD, breast cancer and hip fractures without treatment need to be known. Within the model, it is possible to use risks specified as risk functions, risks manually incorporated into tables or a combination of both. The base-case risk of hip fractures can also be elevated in the model. This option makes it easier to analyse cohorts subject to an increased base-case risk of hip fractures where only the relative risk is available (e.g. osteoporotic individuals).

Values have to be identified for the risk factors involved in the CHD risk function: Cholesterol, diastolic blood pressure, smoking status (fraction between 0 and 1), glucose intolerance (fraction between 0 and 1) and left ventricular hypertrophy (fraction between 0 and 1). These may represent mean values (an average woman) in the population that are subject to analysis. By changing the risk factors, it also becomes possible to analyse cohorts subject to an increased risk of CHD. Conditional on sustaining a CHD, a table decides the distribution among the CHD disease states. The age-dependent probability of different CHD disease states must, therefore, also be identified.

When modelling the intervention, the user must identify how the intervention affects the disease risks and specify the different options associated with the intervention as described above (% risk change, start delay, rise time, stop delay, set time and remaining).

Mortality rates

Age-specific annual mortality rates have to be specified for CHD, breast cancer and hip fractures for the first, and the second and following years after the disease event. Mortality rates need to be stated for all ages between the initial age of the cohort and 110 years. CHD mortality rates are divided into four categories: Recognised acute myocardial infarction, unrecognised acute myocardial infarction, angina pectoris and coronary insufficiency. Sudden death is defined as death within 1 hour from the disease's onset.

Age specific annual mortality rates have to be identified also for death for other causes ("death other"). To calculate the mortality rate for "death other", the risk of dying from CHD, breast cancer and hip fractures are subtracted from the normal mortality rate. In the model mortality

rates for "death other" are obtained by multiplying values extracted from normal mortality tables with one minus the fraction of death that breast cancer and CHD constitutes.

Quality of life weights

Age-dependent quality of life weights have to be specified for CHD, breast cancer and hip fractures for the first and following years after an event. The quality of life weight is a number between 0 (dead) and 1 (full health) which reflects health state preference. Quality of life weights also need to be identified for healthy individuals (the quality of life weight may be lower than 1 due to other diseases not included in the model). The model also permits the inclusion of quality of life weights during treatment, which takes into account potential side effects associated with the intervention.

Costs

Costs necessary for the model can be divided into: intervention, morbidity and mortality costs. Their inclusion is based on a societal perspective meaning that all costs are incorporated into the analysis no matter who pays the costs (see also the CE formula in equation (3) above). The model permits other perspectives as well, e.g. a health care budget perspective.

Age-specific annual morbidity costs must be specified for the first, and the second and following years after a disease event. Morbidity costs are those associated with the treatment of CHD, breast cancer and hip fractures and are also divided into direct and indirect costs as follows: direct costs are those related to the patients' treatment and rehabilitation; whereas, indirect costs are equivalent to a decrease in the value of production caused by the disease. The morbidity costs are interpreted as the extra costs related to morbidity (i.e. the increase in costs due to the disease compared to being "Healthy").

Finally age-specific costs in added life years may be included. Costs in added life years, or mortality costs, are equal to total consumption minus total production in these years (Meltzer 1997).

2.4 Output from the model

Cost per gained life-year and QALY

At the top of the intervention result menu the change in life years resulting from the intervention is shown. This change is calculated as the change in expected survival for the cohort generated by the intervention (all results are presented per individual). Subsequently, the change in quality of life due to morbidity and side effects, measured in QALYs, are shown. Adding the change in life expectancy with the change in quality of life gives the change in QALYs resulting from the treatment.

The change in total costs are presented as changes in intervention, morbidity and mortality costs. The costs are also presented as direct and indirect costs. At the bottom of the intervention menu the model presents CE ratios expressed as costs per change in life years and costs per change in QALYs.

Diseases; lifetime risk; life expectancy

The model shows the distribution of individuals with or without intervention in the different health states (i.e. Death other, Healthy, CHD, Hip fracture and Breast cancer) for a given cycle after treatment onset (0-60 years if the cohort is followed from 50 years). It is also possible to calculate the *lifetime risk* of different diseases for an individual at a certain age. For example the life-time risk of hip fracture for an individual at a certain age is the number of individuals who sustained a hip fracture during the remaining lifetime divided by the number of individuals at risk (the initial cohort). These figures may be compared to estimates on lifetime risks in the general population to check the model's credibility.

Life expectancy is defined as the average future lifetime of the cohort. The total number of cycles (years) for each health state is divided by the size of the original cohort. The total life expectancy is the sum of cycles for the included health states. The life expectancy, conditional on a certain disease state, is calculated as the number of cycles in the disease state divided by the number of women that end up in the disease state, assuming that the cohort starts in the healthy state.

Treatments

There is an option that enables the cohort of individuals to be followed through the model by varying the time horizon for the analysis. It is possible to view the number of individuals in

different disease states and number of cycles the individuals have been in the disease state. In addition to following the cohort at any time horizon, this option can be used as an aid for controlling the model's calculations.

3. Empirical Application - Estimations Based on Swedish Data

A sample simulation applied to a hypothetical cohort of average asymptomatic women at the age of 50, 60 and 70 years with an assumed treatment duration of 10 years. At each age group two indications were analysed: Women with an intact uterus and women with a hysterectomy. Women with a hysterectomy is given oestrogen only therapy whereas women with an intact uterus is given oestrogen combined with a progestogen. For each indication and age the treatment strategy were compared to no treatment. Thus six independent patient groups are considered, two at each age group. The stated question is whether there is good value for money to treat asymptomatic women with an intact uterus or hysterectomised with HRT compared to no treatment.

The costs are collected using a societal perspective including intervention costs, morbidity costs and costs in added life years (Meltzer 1997). The intervention cost include costs for the drug, travel/time and physician visits. Morbidity costs include both direct and indirect costs. Reduced morbidity costs occur when the risk of hip fractures and CHD decrease from using HRT. Increased morbidity costs occur due to increases in the risk of breast cancer from using HRT. Costs in added life years are calculated as the difference between total consumption (i.e. private and public) and total production (Johannesson et al. 1997, Meltzer 1997). The hip fracture risk reduction is assumed, during treatment, to be 40 or 50 percent (OTA 1995, SBU 1996). The risk of hip fractures is assumed to gradually adjust to the base case risk at 10 years after HRT (SBU 1996). The CHD risk reduction is assumed, during treatment, to be 20 or 50 percent (OTA 1995, SBU 1996). The decrease in the risk of CHD is assumed to be the same for oestrogen only therapy and oestrogen combined with a progestogen (Falkeborn et al. 1992). The increase in the risk of breast cancer is assumed to be 0 or 35 percent respectively (OTA 1995). The increased risk of breast cancer is assumed to start instantaneously after 5 years of HRT and remains elevated during the rest of treatment (OTA 1995, SBU 1996). Costs (given in 1995 prices) and effects are discounted at the rate of 3% (Gold et al. 1996). A detailed presentation of the assumptions made and the included data is presented in a working paper (Zethraeus et al. 1998).

Table 1 demonstrates the cost per life-year gained and QALY for different risk reductions and ages. Treating hysterectomised women with oestrogen only therapy is associated with lower CE ratios compared to treating non-hysterectomised women with combined therapy for all ages and risk reductions (Daly et al. 1992, Tosteson and Weinstein 1991). This is explained by a higher intervention cost associated with the combined therapy (*Table 1* in here).

Assuming a 20% reduction in the risk of CHD the CE ratios improves with age at treatment onset. The improved CE ratios are mainly explained by an increased absolute risk of CHD as age increases and, therefore, also a larger decrease in the number of CHD events and related mortality compared with HRT at younger ages. There is also an increased age-related absolute risk of breast cancer and hip fractures, but not as large as CHD.

Assuming a 50% reduction in the risk of CHD the CE ratios increases with age in some cases. With low CE ratios at the age of 50 years, increases in the effectiveness of starting treatment in older ages results in higher CE ratios due to costs in added life years. Thus, the increase in the CE ratio is due to increases in the costs in added life years which are large in the ages above 65 years (i.e. increases in life expectancy are outweighed by increases in costs in added life years).

By adding the risk of breast cancer, the CE ratio generally increases (i.e. worsens). This is due to a lower life expectancy and that savings in morbidity costs are decreased. However, due to the lower life expectancy, costs in added life-years decreases which implies a decrease in total costs and that the CE ratio is unaffected. In some cases a slight decrease in the CE ratio is observed.

The CE ratio is sensitive to changes in the assumptions about the relative-risk reduction in CHD. This sensitivity is confirmed in *Table 1*, which shows a substantial decrease in the cost per gained effectiveness unit for the ages 50 and 60 years when the risk reduction is 50% instead of 20%. Also note that the CE ratio is sensitive to the inclusion of breast cancer risk, for 50-year-old women with a 20% reduction in the risk of CHD. This is explained by the fact that such women have a rather small increase in the life expectancy (when breast cancer is excluded) compared to other groups. The inclusion of breast cancer then has a large influence on life expectancy and CE ratios for these subgroups

If a 5% discount rate for costs and effects are used the CE ratios increases (Zethraeus et al. 1998). This is due to the fact that benefits of the treatment in terms of increased life expectancy and avoided morbidity occur in the distant future whereas the costs mainly arise in the near future (intervention costs) implying that benefits are given lesser weight (more heavily discounted) compared to the costs. The CE of HRT is very sensitive to the presence of side effects. Assuming any side effects during the entire treatment period implies that HRT is dominated by the no intervention alternative in all patient groups. On the other hand if it assumed that the HRT increases the quality of life during the treatment period the CE ratios improves substantially. The CE ratios also improves if the women is assumed to be subject to an increased risk (a doubled risk compared to an average woman) of hip fractures, i.e. if the women are osteoporotic. This is explained by an increased number of avoided fractures due to the increased base risk of hip fractures.

As said above without information about the price per effectiveness unit, a CEA gives no information about whether a program should be implemented or not unless the intervention is a dominated alternative, e.g. when side effects are prevalent according to above. Under some restrictive conditions it is possible to conclude that if the WTP exceeds the costs per gained QALY the treatment programme should be carried out. Recently the WTP per gained QALY is estimated at about SEK 160,000, which may be interpreted as a lower bound for the WTP per gained QALY (Zethraeus 1998). Using this lower bound and comparing it with the costs per gained QALY as estimated in *Table 1* the conclusion is that oestrogen only therapy is cost effective for all ages (50, 60 and 70 years) if the assumed risk reduction of hip fractures and CHD amount to 50%.

The life expectancy calculated from the model is 31.9 years for a woman at the age of 50 years which is near the expected survival extracted from Statistics of Sweden (1995) amounting to 32.3 years. The model predicts a life time risk of having CHD of 25.6%. The predicted life time risk of hip fractures and breast cancer is 16.4 and 7.3% respectively (see e.g. Torgerson and Reid 1997).

4. Conclusion and Discussion

The model, constructed to be as general and flexible as possible, may be theoretically used for any population. However, the default data used for the model in empirical applications are assumed to be valid only for Swedish populations. To make accurate conclusions using the

model in other countries, the data must be valid for the specific setting to which the model is applied. Below, opportunities and data needs for extending the model to other countries are discussed.

Direct and indirect costs must be determined for each country subject to analysis. Using Swedish cost data, multiplied with an appropriate exchange rate, implicitly assumes that the absolute and relative price level is the same as in Sweden. It also assumes that medical and social care patterns are equivalent. These are very strong assumptions and can only be recommended as a very first preliminary analysis. Country-specific costs should ideally be collected. Yearly direct intervention costs, including the costs of pharmaceuticals, physician visits as well as time and travelling costs, can be estimated empirically by following patients during a year of treatment.

Direct and indirect disease costs must be collected for the first 12 months following an event and for the second and following years. Direct morbidity costs are interpreted as the extra costs of the disease compared with no disease occurrence and can be estimated by, for example, subtracting the costs during one year before a disease event from the costs during one year after the disease event (e.g. see Zethraeus et al. 1997). Another alternative is to estimate the costs without the disease by using a matched cohort. The direct costs include all costs associated with the treatment during the initial hospital stay, as well as rehabilitation in aftercare. Indirect morbidity costs can be estimated by subtracting the production value the year after disease onset from its value the year prior to disease.

Quality of life weights may differ between countries and the data should be based on empirical studies. Different methods exist for estimating the quality of life weights (Drummond et al. 1997). For example, the rating-scale, time trade-off and standard gamble methods are commonly used to estimate weights to construct QALYs.

The risk of disease may differ between countries and should be based on country-specific data. The model permits the default values to be changed for the estimated parameters in the risk equations. With these changes, it is then possible to estimate country-specific risk equations for the same risk factors and use these parameter estimates in the model. The risk of breast cancer and hip fractures may be estimated using country-specific incidence data. In the absence of such epidemiological data in Sweden at the moment the risk equation of CHD is extracted from the Framingham Study; whether the results of this study can be extrapolated to other populations is

uncertain. One alternative to verify whether the results are applicable is to compare them with incidence data. Instead of using the risk equations, tabulations may be used. Data on mortality after disease events should be based on country-specific empirical studies. Mortality data is referring to the first year and subsequent years after an event. General mortality may be estimated from national registers involving statistics of mortality rates from the general population.

The model in its original setting evaluates a treatment compared to a baseline alternative (i.e. no treatment). However, the model also permits comparisons between two or more treatments for a given population. The incremental CE ratios between these alternatives must then be calculated. This is made by first calculating the change in costs and effects for the treatments separately (e.g. in the case of two treatments (1 and 2), calculate (C_1 - C_0) and (E_1 - E_0) for treatment 1 compared with the baseline alternative (0) and then calculate (C_2 - C_0) and (E_2 - E_0) for treatment 2 compared with the baseline alternative).

The marginal CE ratio for treatments 1 and 2 is calculated as:

$$\frac{(C_2 - C_0) - (C_1 - C_0)}{(E_2 - E_0) - (E_1 - E_0)} = \frac{C_2 - C_1}{E_2 - E_1} \tag{4}$$

Compliance is the extent to which the patient follows a physician's treatment recommendations. Non-compliance is present if the patient does not follow these recommendations and may be one of two types. First the patient may not buy the drug the physician has prescribed such that no costs or effects associated with the treatment are present. Second, the patient may buy the drug, but diverge from the physician's recommendations, in which case costs for the drug are incurred and only a fraction of the full effect (or no effect) from using the drug is incurred. The second definition of non-compliance necessitates information about how the drug's effect is altered by non-compliance. Note that this type of compliance should be reflected in the estimates of the costs and effects used such that estimates of costs and effects in a clinical trial are based on actual compliance within the trial. To analyse the effect of compliance which differs from compliance within the trial, this necessitates information on how to adjust the effects.

5. References

Andersson K, Mattsson L-Å, Rybo G, Stadberg E. Intrauterine release of levonorgestrel, a new way of adding progestogen in hormone replacement therapy. *Obstetrics and Gynecology* 1992; **79**: 963-967.

Beral V. et al. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *The Lancet* 1997; **350**: 1047-1059.

Colditz GA, Hankinson SE, Hunter DJ, Willett WC, Manson JE, Stampfer MJ, Hennekens C, Rosner B, Spezer FE. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *The New England Journal of Medicine* 1995; **332**: 1589-1593.

Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the Impact of Menopausal Symptoms on Quality of Life. *British Medical Journal* 1993; **307**: 836-840.

Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: An analysis of benefits, risks and costs. *British Medical Bulletin* 1992; **48**: 368-400.

Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes. Second Edition.* New York: Oxford University Press, 1997.

Falkeborn M, Persson I, Adami H-O, Bergström R, Eaker E, Lithell H, Mohsen R, Naessén T. The risk of acute myocardial infarction after oestrogen and oestrogen-progestogen replacement. *British Journal of Obstetrics and Gynaecology* 1992; **99**: 821-828.

Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine.* New York: Oxford University Press, 1996.

Gujarati DN. *Basic Econometrics*. Singapore: McGraw-Hill international editions, 1988.

Johannesson M, Hedbrant J, Jönsson B. A computer simulation model for cost-effectiveness analysis of cardiovascular disease prevention. *Medical Informatics* 1991; **16**: 355-362.

Johannesson M, Meltzer D, O'Conor RM. Incorporating future costs in medical cost-effectiveness analysis: Implications for the cost-effectiveness of the treatment of hypertension. *Medical Decision Making* 1997; **17**: 382-389.

Johannesson M. *Theory and methods of economic evaluation in health care*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

Jönsson B, Christiansen C, Johnell O, Hedbrant J. Cost-effectiveness of fracture prevention in established osteoporosis. *Osteoporosis International* 1995; **5**: 136-142.

Jönsson B, Hedbrant J, Johnell O. *A Computer Simulation Model to Analyse the Cost-effectiveness of Fracture Prevention of Osteoporosis*. EFI Research paper Nr 6525, 1993.

Kannel WB, Wolf PA, Garrison RJ, editors. *The framingham study: an epidemiological investigation of cardiovascular disease. Section 37: the probability of developing certain cardiovascular diseases in eight years at specified values of some characteristics*. Springfield: US Department of Commerce National Technical Information Service, 1987.

Keeler E. *Decision trees and markov models in cost-effectiveness research.* In Sloan, Frank A. (ed.) Valuing health care. New York: Cambridge University Press, 1995.

Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997; **16**: 33-64.

National Board of Health and Welfare (Socialstyrelsen), Sweden, 1993.

OTA (Office of Technology Assessment), Congress of the United States. *Effectiveness and Costs of Osteoporosis Screening and Hormone Replacement Therapy*. Background paper. Volume 1: Cost-effectiveness analysis, 1995.

OTA (Office of Technology Assessment), Congress of the United States. *Effectiveness and Costs of Osteoporosis Screening and Hormone Replacement Therapy*. Background paper. Volume 2: Evidence on Benefits, Risks, and Costs, 1995.

Persson I, Adami H-O, Bergkvist L, Lindgren A, Pettersson B, Hoover R, Schairer C. Risk of endometrial cancer after treatment with oestrogens alone or in conjunction with progestogens: results of a prospective study. *British Medical Journal* 1989; **298**: 147-151.

SBU (The Swedish Council on Technology Assessment in Health Care). SBU report no. 131, Stockholm, Sweden, 1996.

Sonnenberg FA, Beck JR. Markov models in medical decision making. *Medical Decision Making* 1993; **13**: 322-338.

Stanford JL, Weiss NS, Voigt LF, Daling JR, Habel LA, Rossing MA. Combined estrogen and progestin hormone replacement therapy in relation to risk of breast cancer in middle-aged women. *Journal of the American Medical Association* 1995; **274**: 137-142.

Statistics Sweden. *Statistical yearbook of Sweden.* Stockholm: Norstedts tryckeri AB, 1995.

Stockholm County Council. Stockholm inpatient register. 1990.

Torgerson DJ, Reid DM. The economics of osteoporosis and its prevention. A review. *PharmacoEconomics* 1997; **11(2)**: 126-138.

Tosteson A, Weinstein MC. Cost-effectiveness of hormone replacement therapy after the menopause. *Baillière's Clinical Obstetrics and Gynaecology* 1991; **5(4)**: 943-959.

Weinstein MC, Zeckhauser R. Critical Ratios and Efficient Allocation. *Journal of Public Economics* 1973; **2**: 147-157.

Zethraeus N, Johannesson M, Henriksson P, Strand R. The impact of hormone replacement therapy on quality of life and willingness to pay. *British Journal of Obstetrics and Gynaecology* 1997; **104**: 1191-1195.

Zethraeus N, Johannesson M, Jönsson B. *A computer model to analyse the cost-effectiveness of hormone replacement therapy.* EFI Research Paper No 6578 January 1998.

Zethraeus N, Strömberg L, Jönsson B, Svensson O, Öhlén G. The cost of a hip fracture. Estimates for 1,709 patients in Sweden. *Acta Orthopaedica Scandinavica* 1997; **68(1)**:13-17.

Zethraeus N. Willingness to pay for hormone replacement therapy. Forthcoming in Health Economics, 1998.

Tables

Table 1: Cost (SEK thousand) per gained life-year and QALY (QALY in parenthesis) assuming different risk reductions for CHD, hip fracture (Hip) and breast cancer (Cancer). The treatment duration is 10 years for women aged 50, 60 and 70 years.

	Oestrogen (Hysterectomised women)			Oestrogen+Progestogen (Intact uterus)		
Riskchange	50	60	70	50	60	70
Hip-40%, CHD-20%	320 (240)	190 (190)	140 (160)	490 (370)	250 (240)	170 (200)
Hip-40%, CHD-50%	130 (110)	160 (170)	150 (180)	210 (180)	190 (200)	160 (200)
Hip-50%, CHD-20%	280 (200)	170 (160)	120 (140)	440 (320)	220 (210)	150 (170)
Hip-50%, CHD-50%	120 (100)	150 (150)	140 (160)	190 (160)	170 (180)	150 (180)
Hip-40%, CHD-20%, Cancer+35%	4900 (380)	210 (180)	140 (160)	8650 (670)	290 (250)	180 (200)
Hip-40%, CHD-50%, Cancer+35%	150 (100)	160 (160)	150 (180)	270 (190)	190 (190)	170 (200)
Hip-50%, CHD-20%, Cancer+35%	2230 (270)	170 (140)	120 (130)	4210 (520)	240 (210)	150 (160)
Hip-50%, CHD-50%, Cancer+35%	120 (80)	140 (140)	140 (160)	240 (160)	170 (170)	150 (180)

Figures

Figure 1: The basic model structure for evaluating the CE of HRT.

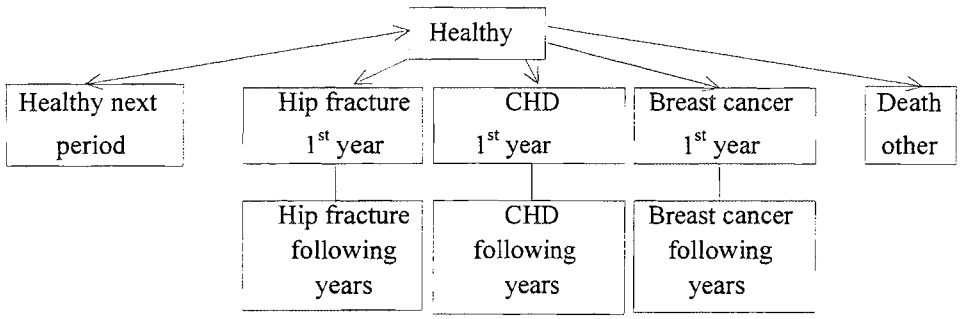
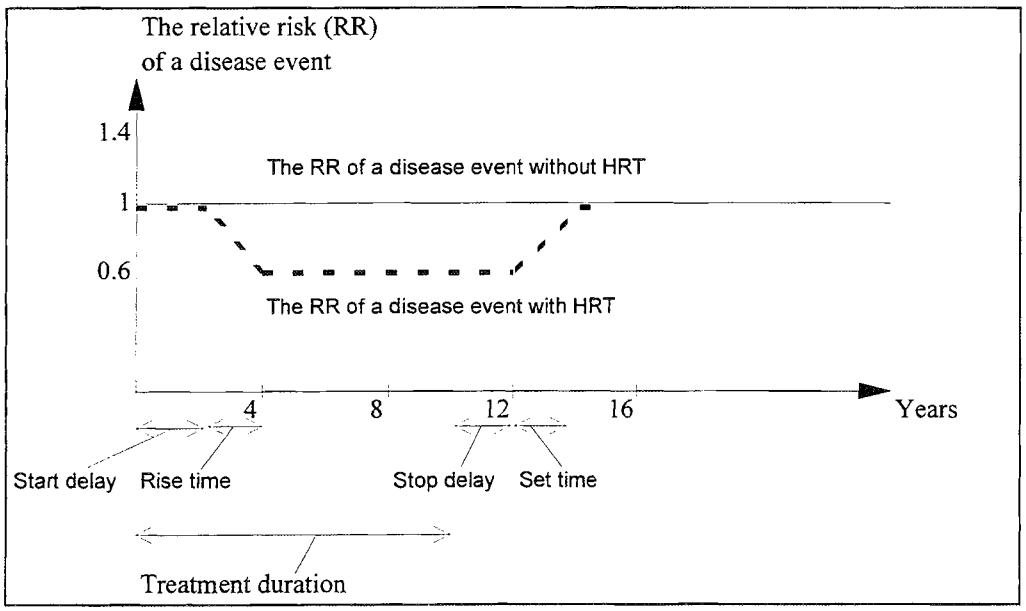


Figure 2: Modelling an intervention.



Chapter 6

Bootstrap Confidence Intervals for Cost-Effectiveness Ratios: Some Simulation Results*

Magnus Tambour and Niklas Zethraeus

Abstract: Recently, a number of papers have brought up the issue of how to make cost-effectiveness (CE) studies stochastic, i.e. how to obtain confidence intervals for CE ratios. In this paper we present a bootstrap procedure for estimating bias-corrected confidence intervals for CE ratios. The bootstrap procedure is tested in a simulation study based on the assumptions made in a recent paper by Wakker and Klaassen (1995). We test two variants of CE ratio bootstrap confidence intervals. The first is a bootstrap analogue of the parametric method proposed by Wakker and Klaassen which gives results similar to those obtained with the parametric method. However, computing bootstrap confidence intervals directly for the CE ratio produces results closer to the predetermined significance level.

Keywords: Bootstrap Confidence Intervals, Cost-Effectiveness Ratios.

* We thank Magnus Johannesson, Andrew Briggs and two anonymous referees for the *Health Economics* for providing valuable comments. The paper has also benefited from comments by participants at the Health Economic Seminars at the Stockholm School of Economics.

1. Introduction

Recently several authors have suggested bootstrap techniques as one possible approach to obtain confidence intervals for cost-effectiveness (CE) ratios (Drummond and O'Brien 1993, O'Brien et al. 1994). The purpose of this paper is to apply bootstrap techniques for calculating bias-corrected confidence intervals for CE ratios and to assess the performance of the bootstrap method. The assessment is made by a simulation study based on the assumptions made in Wakker and Klaassen (1995). Wakker and Klaassen describe how to obtain confidence intervals for CE ratios based on a parametric approach, and as a way to illustrate and test their method they perform a simulation where it is assumed that costs and effects are bivariate normally distributed. In this paper a Monte Carlo resampling algorithm as well as a method for calculating bias-corrected bootstrap confidence intervals are presented. Both are fairly general and have been applied in other contexts for bootstrapping sample means (Atkinson and Wilson 1995, Efron and Tibshirani 1995).

2. Confidence Intervals for CE Ratios

The health economic issue that is considered in this paper is whether a new treatment T_1 should replace a customary treatment T_0. Assume that costs $(C_{i,n})$and effects $(E_{i,n})$ are *observed* for N_1 and N_0 patients, respectively. Let C_i and E_i denote sample means of costs and effects and γ_i and ε_i denote the corresponding *unobserved* true means. The statistic of interest is $\rho = \frac{\gamma_1 - \gamma_0}{\varepsilon_1 - \varepsilon_0}$ which denotes the true incremental CE ratio. Since ρ is unknown it has to be estimated from the samples as $\hat{\rho} = \frac{C_1 - C_0}{E_1 - E_0}$, signifying the *observed* CE ratio. If ρ is less than some critical value ρ^* the new treatment should be implemented. To investigate whether $\hat{\rho}$ is significantly lower than ρ^* a confidence interval for ρ is wanted.

Wakker and Klaassen (1995) show how to compute a 95 percent confidence interval for ρ, which they claim avoids some common ratio estimation problems. The confidence interval is

obtained by traditional statistical methods by first constructing a 97.5 percent one-sided (upper bound) confidence interval for γ and then a 97.5 percent one-sided (lower bound) confidence interval for ε. Finally, they argue that dividing the upper bound in the confidence interval for γ by the lower bound in the confidence interval for ε gives the upper bound for a 95 percent confidence interval for ρ.

The bootstrap procedure

An alternative approach to obtain confidence intervals for CE ratios is to use bootstrap techniques. In each bootstrap resample, a set of "pseudo" cost and effect observations are obtained by simply using the originally estimated values in a Monte Carlo resampling. That is, in each resample N_i integers between 1 and N_i are drawn with replacement, and then the corresponding patient specific cost and effect estimates are used to obtain a pseudo-sample. These N_i estimates are then used to estimate a pseudo-sample mean for each variable. This procedure is repeated B times in order to obtain a set of bootstrap sample means. Various methods for calculating bootstrap confidence intervals can be applied in order to test hypotheses on costs, effects and CE ratios. More formally the bootstrap Monte Carlo algorithm is given by five steps as below, in which θ denotes the statistic of interest, $\hat{\theta}$ denotes an estimate of this statistic and $\hat{\theta}^{*b}$ denotes the b:th bootstrap estimate.

(I) Compute sample means for costs and effects

$$\hat{\theta}_{\gamma_i} = C_i = \frac{1}{N_i}\sum_{n=1}^{N_i} C_{i,n} \text{ and } \hat{\theta}_{\varepsilon_i} = E_i = \frac{1}{N_i}\sum_{n=1}^{N_i} E_{i,n}, \qquad i = 0,1. \qquad (1)$$

(II) Compute small-sample corrections

$$\tilde{C}_{i,n} = C_{i,n}\sqrt{\frac{N_i}{N_i - 1}} + \hat{\theta}_{\gamma_i}\left(1 - \sqrt{\frac{N_i}{N_i - 1}}\right), \; i = 0,1 \qquad (2a)$$

and

$$\tilde{E}_{i,n} = E_{i,n}\sqrt{\frac{N_i}{N_i - 1}} + \hat{\theta}_{\varepsilon_i}\left(1 - \sqrt{\frac{N_i}{N_i - 1}}\right), \; i = 0,1 \qquad (2b)$$

This step in the algorithm is motivated by Atkinson and Wilson (1995) as a necessary step to avoid type-*I* errors in small samples.

(III) Resample N_1 and N_0 integers with replacement from the index sets $I_1 = \{1,2,...,N_1\}$ and $I_0 = \{1,2,...,N_0\}$. Let I^{*b} denote the vector of resampled patient identification numbers in bootstrap resample b. I^{*b_1} thus gives the index of the patients undergoing treatment 1 in resample b. Note that the same index is used for costs and effects for treatments $i = 0,1$. This is motivated by the possibility of a non-zero correlation between costs and effects for each patient. A possibility of a non-zero correlation between patient specific costs and effects means that costs and effects for a given patient may be related in some way. A positive (negative) correlation exists if there is a positive (negative) relationship between costs and effects. A positive correlation may exist when a dose regime is increased (and the pharmaceutical costs increase), up to a certain level, at the same time as the effects increase. A negative correlation may exist when side effects are associated with a treatment. If the possibility of a non-zero correlation between costs and effects is excluded an unnecessary restriction is imposed. Ultimately, this may lead to incorrect confidence intervals and erroneous conclusions about the true CE ratio.

(IV) Compute bootstrap sample difference and CE-ratio means

$$\hat{\theta}_{\gamma}^{*b} = \frac{1}{N_1}\sum_{n=1}^{N_1}\tilde{C}_{1,n}^{*b} - \frac{1}{N_0}\sum_{n=1}^{N_0}\tilde{C}_{0,n}^{*b}, \qquad \hat{\theta}_{\varepsilon}^{*b} = \frac{1}{N_1}\sum_{n=1}^{N_1}\tilde{E}_{1,n}^{*b} - \frac{1}{N_0}\sum_{n=1}^{N_0}\tilde{E}_{0,n}^{*b} \quad (3a,b)$$

and

$$\hat{\theta}_{\rho}^{*b} = \left(\frac{1}{N_1}\sum_{n=1}^{N_1}\tilde{C}_{1,n}^{*b} - \frac{1}{N_0}\sum_{n=1}^{N_0}\tilde{C}_{0,n}^{*b}\right) \Big/ \left(\frac{1}{N_1}\sum_{n=1}^{N_1}\tilde{E}_{1,n}^{*b} - \frac{1}{N_0}\sum_{n=1}^{N_0}\tilde{E}_{0,n}^{*b}\right), \qquad (4)$$

where a tilde denotes a small-sample corrected value.

(V) Steps III and IV are repeated B times in order to obtain a set of bootstrapped cost and effect differences and incremental CE-ratio means.

Bootstrap confidence intervals

The simplest method of calculating bootstrap confidence intervals is to arrange the B bootstrap statistics in increasing order and then to cut off, say, five percent of the observations in the upper tail. This procedure gives a 95 percent, one-sided, *percentile* confidence interval. Since this method does not take the original estimate into consideration, the issue of bias arises. We use the bias-corrected accelerated (BC_a) method (Efron and Tibshirani 1995) to compute bias-corrected confidence intervals. The BC_a method has advantages over the percentile method. For example, it corrects for potential bias and is second order accurate - as opposed to the percentile method which is only first order accurate (Efron and Tibshirani 1995). Second order accuracy mean that the error in not covering the true value approaches zero at the rate $1/N$ in terms of the sample size (N). For the percentile method the error approaches zero at the rate $\frac{1}{\sqrt{N}}$. For a more thorough description of the properties of BC_a, see Efron and Tibshirani (1995).

A one sided - upper bound - $(1-\alpha)\times 100$ percent BC_a confidence interval for the statistic of interest is bounded by the percentile

$$\hat{\theta}^{*b}(\alpha_U), \tag{5}$$

where α_U is given by

$$\alpha_U = \Phi\left(\hat{\sigma} + \frac{\hat{\sigma} + \Phi^{-1}(1-\alpha)}{1-\hat{a}\left(\hat{\sigma} + \Phi^{-1}(1-\alpha)\right)}\right) . \tag{6}$$

$\Phi(\cdot)$ denotes the standard normal cumulative distribution function and $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal distribution. In order to compute the bounds of the BC_a confidence intervals, both the bias adjustment and the acceleration term have to be estimated. The bias-adjustment term $\hat{\sigma}$ is given by

$$\hat{\sigma} = \Phi^{-1}\left(\frac{\#\left\{\hat{\theta}^{*b} < \hat{\theta}\right\}}{B}\right). \tag{7}$$

The bias-correction term, $\hat{\sigma}$, measures the median bias of $\hat{\theta}^*$, i.e. the discrepancy between the median of $\hat{\theta}^*$ and $\hat{\theta}$ in normal units (Efron and Tibshirani 1993). If $\hat{\sigma}$ is equal to zero, exactly half of the $\hat{\theta}^{*b}$ values are less than (or equal to) $\hat{\theta}$ and no bias-correction is made. Finally, the acceleration constant, $\hat{a}$, can be estimated in terms of the jack-knife values

$$\hat{\theta}_s = \frac{1}{K-1}\sum_{\substack{i=1 \\ i\neq s}}^{K}\hat{\theta}_i \;,\; K = N_1 + N_0 \tag{8}$$

as

$$\hat{a} = \frac{\sum_{s=1}^{K}\left(\hat{\theta} - \hat{\theta}_s\right)^3}{6\cdot\left[\sum_{s=1}^{K}\left(\hat{\theta} - \hat{\theta}_s\right)^2\right]^{\frac{3}{2}}} \;,\; K = N_1 + N_0 . \tag{9}$$

The acceleration term, $\hat{a}$, reflects the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter value, θ, measured on a normalised scale (Efron and Tibshirani 1995). If both $\hat{a} = 0$ and $\hat{\sigma} = 0$ the BC_a confidence interval is the *percentile* interval. Thus, the percentile interval is a special case of the BC_a interval. In that case the upper bound of the BC_a and percentile confidence interval is equal to the $(1-\alpha)\times B : th$ highest bootstrap estimate.

3. Simulation of CE Ratios

To test the bootstrap method in terms of degree of accuracy, a simulation study is performed in which the bootstrap method is compared to the Wakker and Klaassen (1995) approach (*Table 1*). Costs and effects are assumed to be bivariate normally distributed with negative correlations in the default values. The simulation is based on the same default values as in Wakker and Klaassen (1995). Thus, $N_0 = 60$; $N_1 = 40$; $\gamma_0 = 30{,}000$; $\gamma_1 = 40{,}000$; $\varepsilon_1 = 60$; $\varepsilon_0 = 50$; Hence $\rho = 1{,}000$. The standard deviations are $\sigma_0 = \sqrt{Var(\gamma_0)} = \frac{5{,}000}{\sqrt{2}}$; $\sigma_1 = \sqrt{Var(\gamma_1)} = \frac{5{,}500}{\sqrt{2}}$; $\tau_0 = \sqrt{Var(\varepsilon_0)} = \frac{1}{\sqrt{2}}$; $\tau_1 = \sqrt{Var(\varepsilon_1)} = \frac{1.2}{\sqrt{2}}$. Finally, the correlations between costs and effects are set equal to: $\lambda_0 = -0.3$ and $\lambda_1 = -0.2$.

We estimate two bootstrap confidence intervals called the ***"WK" bootstrap*** and the ***"direct" bootstrap***. The first confidence interval is a bootstrap analogue of the method developed by Wakker and Klaassen (1995) which means that we first compute *separate* confidence intervals for the cost difference (γ) and the effect difference (ε) using expression (3a, b). Then the upper (bootstrap confidence interval) bound for ρ is obtained by taking $\hat{\theta}_{\gamma}^{*b}(\alpha_U)/\hat{\theta}_{\varepsilon}^{*b}(\alpha_L)$. The second confidence interval is based on expression (4), which is the same as the bootstrap estimate suggested by Stinnett (1996). In this case the upper bound of the confidence intervals is estimated directly for ρ by computing $\hat{\theta}_{\rho}^{*b}(\alpha_U)$. Note that $\hat{\theta}_{\gamma}^{*b}(\alpha_U)/\hat{\theta}_{\varepsilon}^{*b}(\alpha_L)$ is not necessarily equal to $\hat{\theta}_{\rho}^{*b}(\alpha_U)$. The simulation results are shown in *Table 1*, which is essentially a replication of *Table 1* and *Table 2* in Wakker and Klaassen (1995) (*Table 1* in here).

Each row represents 10,000 simulations and in each of these simulations a bootstrap with 1,000 resamples was performed. The first row shows the results using the default values. In the next seven rows the left-hand column shows which default values are changed. The results from the simulation show that the WK bootstrap produces very similar results with respect to upper bound (95%-CI-ub), standard deviation (S.D.) and error percentage (Err.%) compared to Wakker and Klaassen (1995). The Err.% gives the percentage of the 10,000 repetitions of the simulation where ρ was not contained in the confidence interval for ρ. The S.D. and the Err.% are always somewhat higher for the WK bootstrap compared to the Wakker and Klaassen method. Still, the Err.% is far below the 5 percent significance level for the WK bootstrap. This indicates that the WK bootstrap is conservative in the same way as the parametric WK method, i.e. the risk of not rejecting a false null hypothesis may be high (type II error). However, the direct bootstrap produces different results as compared to Wakker and Klaassen (1995). The upper bound for the confidence interval is always lower in the direct bootstrap as compared to Wakker and Klaassen while the opposite holds for the standard deviations. The Err.% for the direct bootstrap is always close to the 5 percent significance level. The Err.% differs most from the 5 percent significance level for the case when $N_0 = 15, N_1 = 10$.

4. Summary and Conclusions

In this paper we presented an approach to obtain bias-corrected bootstrap confidence intervals for cost-effectiveness ratios. To assess the performance of the bootstrap method a simulation study was performed based on the assumptions made in a paper by Wakker and Klaassen (1995). Two different bootstrap confidence intervals for the incremental CE ratio (ρ) were estimated. First a WK bootstrap analogue was computed where *separate* confidence intervals were constructed for the cost difference (γ) and the effect difference (ε). The upper bound in the confidence interval for γ where then divided by the lower bound in the confidence interval for ε resulting in an upper bound for the confidence interval for ρ. Secondly, a direct bootstrap was computed where a confidence interval was constructed directly for ρ.

The results from the simulations show that the WK bootstrap produced very similar results as compared to the method presented by Wakker and Klaassen. Wakker & Klaassen conclude that their method is conservative, i.e. it 'gives up power' by taking large confidence intervals, especially in the case of few observations. The implications of this may be that the Wakker & Klaassen method tends to under-reject, given the significance level, i.e. too few false null-hypotheses may be rejected. These conclusions are also valid for the WK bootstrap that we performed. However, the direct bootstrap produces narrower confidence intervals compared to the other two methods. The direct bootstrap performs best in terms of error percentage when the sample variation is small and worst for small sample sizes. The error percentage is however always close to the 5 percent significance level in all simulations. Thus, on the basis of our results we conclude that the direct bootstrap is a more accurate method, in terms of producing error percentages close to the stated significance level, compared to the WK bootstrap and the parametric Wakker & Klaassen method.

5. References

Atkinson SE, Wilson PW. Comparing Mean Efficiency and Productivity Scores from Small Samples: A Bootstrap Methodology. *Journal of Productivity Analysis* 1995; **6(2)**: 137-152 .

Drummond M, O'Brien B. Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Economics* 1993; **2**: 205-212.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability, No. 57. New York: Chapman and Hall, 1993.

O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; **32**: 150-163.

Stinnett AA. Adjusting for Bias in C/E Ratio Estimates. *Health Economics* 1996; **5**: 470-472.

Wakker P, Klaassen MP. Confidence intervals for cost/effectiveness ratios. *Health Economics* 1995; **4**: 373-381.

Tables

Table 1: Bootstrap methods versus Wakker and Klaassen.

Non-default	**95%-CI-ub**			**S.D.**			**Err. %**		
	Wakker & Klaassen	*WK Bootstrap*	*Direct Bootstrap*	*Wakker & Klaassen*	*WK Bootstrap*	*Direct Bootstrap*	*Wakker & Klaassen*	*WK Bootstrap*	*Direct Bootstrap*
	1186.62	1186.55	1135.26	87.05	87.65	85.11	1.47	1.54	5.34
$\tau_0=5/\sqrt{2};\tau_1=5.5/\sqrt{2}$	1365.44	1365.85	1231.07	173.83	174.73	154.88	0.67	0.68	5.24
std/8	1022.97	1022.71	1016.82	10.29	10.33	10.26	1.32	1.38	4.78
$N_0=N_1=200$	1089.80	1089.82	1066.09	40.66	40.87	40.32	1.25	1.27	4.90
$N_0=15;N_1=10$	1389.01	1390.69	1278.79	187.12	190.50	178.25	1.57	1.68	5.73
$\lambda_1=\lambda_0=-0.95$	1188.62	1188.66	1156.37	99.47	100.16	98.50	2.77	2.90	5.19
$\lambda_1=\lambda_0=0$	1188.97	1189.09	1131.02	82.64	83.18	80.40	0.88	0.92	5.07
$\lambda_1=\lambda_0=+0.95$	1186.39	1186.24	1097.38	63.09	63.64	59.85	0.17	0.26	5.38

Chapter 7

Non-Parametric Willingness to Pay Measures and Confidence Statements*

Magnus Tambour and Niklas Zethraeus

Abstract: Willingness to pay (WTP) for a health care program can be estimated in contingent valuation (CV) studies by a non-parametric approach suggested by Kriström (1990). The non-parametric approach is free from distributional assumptions, which is a strength compared to parametric regression based approaches. However, using a non-parametric approach it is not clear how to obtain confidence statements for WTP estimates, for example when testing hypotheses regarding differences in mean WTP for different subsamples. In this paper we propose a procedure that allows statistical testing and confidence interval estimation by employing bootstrap techniques. The method is easy to implement and has low computational costs due to the performance of modern PCs. The method is applied to data from a CV study where the WTP for hormone replacement therapy (HRT) was investigated. The mean WTP was estimated for the full sample and separately for women with mild and severe menopausal symptoms. One conclusion that can be drawn using the proposed method is that the mean WTP is significantly higher in the group with severe symptoms.

Keywords: Bootstrap, Economic Evaluation, Hormone Replacement Therapy, Non-Parametric, Willingness to Pay.

* We thank Magnus Johannesson for providing comments. We are also grateful to the editor for *Medical Decision Making* and two anonymous referees for providing comments.

1. Introduction

Different methods for eliciting monetary values of health care programs have been presented in the literature (Johannesson 1996, Johansson 1995). The expressed preference approach, or contingent valuation (CV) method, is one way to obtain benefit measures of health care programs. In the CV method, survey methods are used to investigate the willingness to pay (WTP) for a good or a service. The CV method was originally developed in the environmental field to measure the value of changes in the environment, but recently a number of health care applications have been presented (Appel et al. 1990, Berwick and Weinstein 1985, Donaldson 1990, Estaugh 1991, Golan and Shechter 1993, Johannesson et al. 1993, Johannesson and Johansson 1996, Kartman et al. 1996, Lauraine et al. 1996, O'Brien and Gafni 1996, Zethraeus et al. 1997).

CV questions can be divided into open-ended questions and closed-ended questions. Due to problems with open-ended questions (e.g. starting point bias when using bidding games (Stålhammar 1996)) the current recommendation is to use closed-ended CV questions (Johannesson et al. 1996, NOAA 1993). The CV method based on closed-ended questions (and open-ended questions) is not without problems and different issues remain to be solved. For example, the relationship between hypothetical and real money payments has to be tested further. Another methodological issue, which we address in this paper, is how to obtain confidence statements for the mean WTP.

Parametric or non-parametric approaches can be used to estimate the mean WTP from dichotomous choice (closed-ended) CV questions. A parametric approach necessitates assumptions regarding functional form (Bishop and Heberlein 1979, Cameroon 1988, Hanemann 1984). The main advantage of non-parametric estimators is that they are robust against functional misspecification. The non-parametric approach also has the virtue that the mean WTP is usually simple to estimate (Kriström 1990). For the parametric approach different methods have been proposed to account for uncertainty in the mean WTP measure due to sample variation (Cooper 1994, Park et al. 1991). However, it is perhaps less clear how to obtain confidence statements for mean WTP estimates using a non-parametric approach. The purpose of this paper is to propose a procedure based on bootstrap techniques that allows statistical testing and confidence interval estimation for the mean WTP where the estimator is

based on the non-parametric approach developed by Kriström (1990). This procedure makes it possible not only to estimate confidence intervals for the mean WTP in the whole sample in a CV study, but also to test whether there are significant differences in mean WTP between different subsamples. The approach requires no assumptions regarding parametric functional form, but due to the computational intensity of the bootstrap technique the method cannot be applied using "back of the envelope" calculations. However, due to the performance of modern PCs the computational cost is still low.

2. A Non-Parametric WTP Measure

The non-parametric approach suggested by Kriström (1990) has the same underlying theoretical foundation as the parametric methods (Johansson 1995). The difference lies in the estimators of the (mean) WTP. Following Hanemann (1984) and Kriström (1990) it is assumed that a bid of p_i dollars for a change in health status $(h^0$ to $h^1)$ is accepted if

$$V(h^1, I - p_i; C) - V(h^0, I; C) - \varepsilon \geq 0, \qquad (1)$$

where $V(.)$ denotes a utility function, I is the net income, ε is an error term which is assumed to be an identical and independent distributed random variable and C is a vector of household characteristics (Cameron 1988). The price for which the equality holds defines the individual's maximum WTP for the change in health status. In other words, this price is such as to keep the individual at her initial level of utility following a change in the individual's health status. The representation in (1) reflects the information about the individual utility function that is available to the researcher, i.e. the utility difference contains some unobservable elements to the researcher. However, it is assumed that the individual knows his or her utility function with certainty. The probability that an individual accepts a given bid of p_1 dollars can then be written as

$$\text{Prob}\{\text{individual accepts } p_1\} = F_\eta(\Delta V(.)) = 1 - G(.), \qquad (2)$$

where $F_\eta(.)$ is the cumulative distribution function of the error term $\eta = \varepsilon$ whose expected value is equal to zero. $\Delta V = V(h^1, I - p_1; C) - V(h^0, I; C)$ is the change in utility measured as

the difference between the utility in health state 1 (with treatment and payment) and health state 0 (without treatment) and $G(.)$ is the cumulative distribution function of WTP.

The expected (mean) WTP can then be expressed as[1]

$$E[WTP] = \int_0^b (1 - G(p))dp - \int_a^0 G(p)dp, \qquad (3)$$

where $G(p) = prob\{WTP \leq p\}$ is the cumulative distribution function of WTP, $a \leq 0$ and $b > 0$ are the lower and upper limits of integration, and where $G(b) = 1$ and $G(a) = 0$. Thus $1 - G(p)$ yields the probability that the individual is willing to pay at least p. Note that if $a = 0$ the possibility of negative bids is excluded. The main advantage of using a non-parametric estimator is that it is robust against distributional misspecifications (in $G(.)$) (Kriström 1990).

In a closed-ended CV study J prices (bids), $(p_1, p_2, ..., p_J)$, are offered to J subsamples each consisting of N_j individuals.[2] Each respondent is offered one price only and the respondent either rejects or accepts the price. The observed proportion of "yes" answers (acceptance probabilities) in each group of respondents, $(\hat{r}_1, \hat{r}_2, ..., \hat{r}_J)$, can be used to construct an empirical survival-function (a so called Ayer-curve) that shows the relationship between the price and the proportion of yes answers (Ayer et al. 1955). This curve can be interpreted as a market demand curve for the treatment. The mean WTP can then be estimated by simply calculating the area below the curve for non-negative prices less the area below the line $r(p) = 1$ and above the curve for negative prices. This is referred to as a non-parametric estimator of the mean WTP (Kriström 1990).

Let x_{ij} denote the answer from the i:th patient in the j:th bid-group, $(i = 1,2,..., N_J)$. Each respondent can reject the bid $(x_{ij} = 0)$ or accept the bid $(x_{ij} = 1)$. Let $\hat{r}_j = \sum_i x_{ij} / N_j$ denote

[1] See Johansson (1995), pp 96, for a derivation of this expression.
[2] We use the convention that prices (and the corresponding bid-groups) are denoted in increasing order, i.e. $p_1 < ... < p_J$.

the observed ratio of yes answers in group j and let $\hat{\omega}$ denote the estimated mean WTP. The formula for the non-parametric mean WTP proposed by Kriström (1990) can be written as

$$\hat{\omega} = \sum_{j=0}^{J}\left(\hat{r}_{j+1}\left(p_{j+1}-p_j\right)+\frac{1}{2}\left(p_{j+1}-p_j\right)\left(\hat{r}_j-\hat{r}_{j+1}\right)\right) = \frac{1}{2}\sum_{j=0}^{J}\left(p_{j+1}-p_j\right)\left(\hat{r}_{j+1}+\hat{r}_j\right). \quad (4)$$

It should be noted that there are some ambiguities in how to determine the choice of endpoints (corresponding to a and b in equation (3) above) in the tails of the distribution. In the empirical application we first assume that r_0 is equal to 1 when $p_0 = 0$ which implies that a is assumed to be equal to 0. In other words it is assumed that all individuals accept the offer at a zero price and that we exclude negative WTP. This seems reasonable in the evaluation of health care programs since they almost entirely concern private goods or services which people may freely elect or not elect to buy. It may be the case that the individual perceives non-positive marginal benefits from the consumption of a medical drug (e.g. due to side-effects) in which case he or she either interrupts the treatment or switches to a different drug. If some individuals report a zero WTP our first assumption implies that the mean WTP according to (4) is overestimated. Second, it is assumed that r_{J+1} is equal to 0 when $p_{J+1} > p_J$, i.e. it is assumed that no one accepts to pay a higher price than p_J (i.e., $b = p_J$). This (or a similar) assumption has to be made if some individuals accept to pay the maximum price p_J in the price vector. Presuming that some individuals are willing to pay a higher price than p_J our second assumption implies that the mean WTP according to (4) is underestimated. Valuable information about the endpoints can be obtained from a pilot study. This information can be used to design the price vector in the main study (Johannesson 1996). Between the co-ordinates $\left(\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_{J+1}\right)$ we use linear interpolation. It can be noted that it would of course be possible to choose other endpoints in the tail of the distribution and other interpolation methods, but we leave this discussion aside as it is not important for the bootstrap procedure per se.

Note that the proportion of yes answers must be non-increasing in p (Kriström 1990). That is, $\hat{r}_{j+1} \leq \hat{r}_j$ for all $j = 0,1,\ldots,J,$ must hold. If this is not satisfied we follow Kriström (1990) and calculate adjusted ratios according to the algorithm

$$\tilde{r}_j = \tilde{r}_{j+1} = \left(\sum_i x_{i,j} + \sum_i x_{i,j+1}\right) / \left(N_j + N_{j+1}\right), \qquad (5)$$

for each pair, triplet etc. until monotonicity is obtained. These adjusted ratios are then used in the estimation of the mean WTP. It can be shown that the adjusted ratios $\tilde{r}_j$ have the desired consistency property, i.e. the adjusted ratios converge in probability to the true ratios (Ayer et al. 1955).

The bootstrap approach

The bootstrap technique is a computationally intensive method for obtaining measures of accuracy in statistical estimates (see Efron and Tibshirani (1993) for an introduction). The idea of the bootstrap approach as it is used here is to obtain pseudo-samples of ones and zeros from each of the J subgroups and estimate a large number of bootstrap mean WTP measures. With repeated non-parametric WTP estimates it is possible to calculate bootstrap confidence intervals and to test various hypotheses. The pseudo-samples are obtained by drawing N_j random numbers between $\left(1,\ldots,N_j\right)$ (with replacement) in each j bid-group. Each element (or index number) corresponds to a zero or one in each original sample bid-group. For each of the bootstrap pseudo-samples it is possible to obtain a bootstrap estimate of the ratio of yes-answers. These bootstrap ratios are then used to compute a non-parametric mean WTP bootstrap estimate ($\hat{\omega}^{*b}$) by using (4). This procedure is repeated B times to obtain the same number of non-parametric mean WTP estimates. The B bootstrap estimates can then be used to compute bootstrap confidence intervals. A more formal presentation of the bootstrap algorithm is given in Appendix 1.

Different methods to obtain bootstrap confidence intervals can be employed (Efron and Tibshirani 1993). The simplest method of calculating bootstrap confidence intervals is to arrange the B bootstrap statistics in increasing order and then to cut off, say, 2.5 percent of the observations in each tail. This procedure gives a 95 percent, double-sided, *percentile* confidence interval. However, in cases where the distribution of $\hat{\omega}^{*b}$ is not symmetrical the percentile method leads to incorrect intervals, i.e. the endpoints are biased. An improved method is the bias-corrected (BC) method (Efron and Tibshirani 1993) which adjusts for this

type of bias. The algorithm for estimating the BC confidence interval for the mean WTP is given in Appendix 1.

3. Empirical Application

At menopause, around the age of 50, about 80 percent of all women experience menopausal symptoms, which have negative impacts on the quality of life. The loss in quality of life may be substantial. HRT may alleviate these symptoms and increase the quality of life for women. To investigate whether HRT is an efficient use of scarce health care resources in a cost-benefit framework the costs and benefits of HRT have to be measured in the same units (monetary units). One way to assess the value of the change in quality of life due to HRT is to measure the WTP for HRT. Then it becomes possible to directly compare the benefits with the costs in monetary terms.

Data

To analyse how much symptomatic women are willing to pay for HRT a questionnaire was consecutively administered to 104 women recruited from the Department of Gynaecology at a Swedish hospital (Södertälje Hospital) during the period February 6, 1995 to March 18, 1996[3]. A nurse at the clinic interviewed the women after their consultation with the clinic doctor. In order to classify each woman as having mild or severe symptoms she was asked to read a description of mild and severe symptoms used in a study by Daly et al. (1993), and to choose the alternative which best corresponded to her own symptoms before taking HRT. The interview consisted of three parts. In the first part the woman was asked to indicate her health status before starting HRT and her present health status with HRT, on a rating scale (RS) between 0 (dead) and 100 (full health). In the second part the woman was asked to indicate her health status before starting HRT and her present health status with HRT, based on the time trade-off (TTO) method. In the third part of the interview the WTP for HRT was investigated by the CV method based on a closed-ended approach, i.e., each individual was asked if she would be willing to pay a specific price or not.

[3] For a more detailed description of the data collection, see Zethraeus et al. (1997).

In the questionnaire the woman was asked if she would continue her current HRT if she had to pay SEK p (August 1997: 1£= SEK 12.0, 1$=SEK 8.0) per month out of her own income. The price (p) was randomly varied between SEK 100 and SEK 10,000 in eight different subsamples ($J = 8$), and each individual was offered one of these prices. The eight different prices were SEK 100; 500; 1,000; 1,500; 2,000; 3,000; 5,000 and 10,000. The formulation of the WTP question is given in Appendix 2.

In the estimation of the mean WTP we assume, as mentioned above, that no one accepts to pay a higher price than the maximum of SEK 10,000 used in the study ($b=10,000$). We also assume that each woman would continue her HRT if the price was 0 ($a=0$). The mean WTP was estimated for the whole sample and separately for the two subsamples with mild and severe menopausal symptoms, respectively. To obtain a non-increasing function of the yes answers in p smoothing was necessary for the entire sample as well as for the groups with mild and severe symptoms. *Table 1* shows the original as well as the adjusted ratio values after smoothing was carried out (*Table 1* in here).

The mean treatment duration at the time for the interview was 3 years and the response rate was 100%. Eighty-five women were treated with oestrogen in combination with a progestogen, while 19 women were treated with oestrogen alone. The mean age of the entire patient group was 52.2 years (range 45-65 years) whereas the mean age for women with mild and severe menopausal symptoms was 52.0 (range 45-60) and 52.4 (range 45-65) years respectively (*Table 2*). *Table 2* shows a large increase in the quality of life scores in the two groups in terms of increases in RS and TTO. The socio-economic factors income, age and household size are almost the same in the two groups, whereas women with mild symptoms have a higher education level (*Table 2* in here).

Based on the Ayer-curve represented by the curve in *Figure 1* the mean WTP was estimated (*Figure 1* in here). The mean WTP for all the women (full sample) was SEK 3,508. For women suffering mild and severe menopausal symptoms the mean WTP for HRT was 2,346 and SEK 4,838 respectively. These are point estimates on the mean WTP for HRT, which indicates a large difference in mean WTP between the two groups. One explanation of the high WTP for HRT is that there is a considerable increase in the quality of life from using HRT in terms of changes in the TTO and RS, as indicated in *Table 2*.

To account for uncertainty due to the sample variation, a confidence interval for the true mean WTP can be constructed using the outlined bootstrap approach. The bootstrap approach also makes it possible to test whether there is a statistically significant difference in the mean WTP between the groups with severe and mild menopausal symptoms.

Bootstrap results

In this subsection we report confidence intervals for mean WTP. First, a confidence interval for the mean WTP was estimated using the whole sample. The same procedure was then used for the two subsamples with women with severe and mild symptoms, respectively. Finally, to investigate whether there is a significant difference in mean WTP between women suffering from mild and severe symptoms, we estimated a bootstrap confidence interval for the difference in mean WTP. The results are reported in *Table 3* (*Table 3* in here).

The first row shows the results for the full sample. The original estimate shows a mean WTP of about SEK 3,500 per month that Zethraeus et al. (1997) concluded is well above the estimated treatment costs. The bootstrap results show that even the lower bound of the mean WTP exceeds the treatment costs. An analysis of the uncertainty due to sampling variation as reported in *Table 3* could therefore strengthen the results in a cost benefit study. For the two subsamples the original estimates are as expected, in the sense that women with severe symptoms are willing to pay a larger amount than women with mild symptoms. The original point estimates indicate that women with severe symptoms are willing to pay twice as much as women with mild symptoms. As the last row in *Table 3* shows, this difference is statistically significant. It should be noted that it is not enough to use the confidence intervals for the two groups and compare the upper bound for the mild symptoms group with the lower bound for the severe symptoms group in order to conclude whether or not the difference between the groups is significant (Poe et al. 1994).

Finally, it should also be noted that other factors not controlled for may explain some of the difference in the WTP between women with mild and severe symptoms. The background variables that differ significantly between the two groups are the education, RS and TTO variables, which indicate a higher education level and less gain in quality of life for women with mild symptoms. Socio-economic factors such as income, age and household size do not differ significantly between the two groups.

4. Summary and Conclusions

Cost-benefit analysis with WTP estimates used as benefit measures is one approach to the evaluation of health care programs. Estimates of mean WTP can be obtained from CV studies by parametric or non-parametric techniques. In this paper we proposed a bootstrap procedure that allows statistical testing and confidence interval estimation for non-parametric mean WTP estimates.

The bootstrap approach can also be applied using other assumptions regarding the tails of the distribution and the behaviour between the prices in the bid groups. Instead of using linear interpolation, cubic splines could have been used. Other upper and lower limits of integration could also have been used.

The method was applied to data from a Swedish CV study of HRT. It was possible to conclude that the lower bound of the confidence interval was well above the treatment costs for this program. In a complete stochastic analysis one should of course also estimate confidence intervals for the cost measure, but this was not possible in the study from which the data were taken. The results also showed that there was a significant difference in mean WTP between women with severe and mild symptoms respectively. Such conclusions cannot be drawn using the non-parametric estimates alone. We therefore believe that bootstrap techniques can offer a comprehensive tool for stochastic analysis for the non-parametric WTP measure.

5. References

Appel LJ, Steinberg EP, Powe NR, Anderson GF, Dwyer SA, Faden RR. Risk reduction from low osmolality contrast media. What do patients think it is worth? *Medical Care* 1990; **28**: 324-334.

Ayer M, Brunk HD, Ewing GM, Silverman E. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics* 1955; **26**: 641-647.

Berwick DM, Weinstein MC. What do patients value? Willingness to pay for ultrasound in normal pregnancy. *Medical Care* 1985; **23**: 881-893.

Bishop RC, Heberlein TA. Measuring values of extra goods: Are indirect methods biased? *American Journal of Agricultural Economics* 1979; **61**: 926-930.

Cameron TA. A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics* 1988; **15**: 355-379.

Cooper JC. A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. *Land Economics* 1994; **70**: 111-122.

Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the impact of menopausal symptoms on quality of life. *British Medical Journal* 1993; **307**: 836-840.

Donaldson C. Willingness to pay for publicly-provided goods: A possible measure of benefit? *Journal of Health Economics* 1990; **9**: 103-118.

Efron B, Tibshirani RJ. *An introduction to the bootstrap.* Monographs on Statistics and Applied Probability, No. 57. New York: Chapman and Hall, 1993.

Estaugh SR. Valuation of the benefits of risk-free blood. Willingness to pay for hemoglobin solutions. *International Journal of Technology Assessment in Health Care* 1991; **7**: 51-57.

Golan EH, Shechter M. Contingent valuation of supplemental health care in Israel. *Medical Decision Making* 1993; **13**: 302-310.

Hanemann WM. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 1984; **66**: 332-341.

Johannesson M, Johansson P-O, Kriström B, Gerdtham U-G. Willingness to pay for antihypertensive therapy: further results. *Journal of Health Economics* 1993; **12**: 95-108.

Johannesson M, Johansson P-O. To be, or not to be, that is the question: An empirical study of the WTP for an increased life expectancy at an advanced age. *Journal of Risk and Uncertainty* 1996; **13**: 163-174.

Johannesson M, Jönsson B, Karlsson G. Outcome measurement in economic evaluation. *Health Economics* 1996; **5**: 279-296.

Johannesson M. *Theory and methods of economic evaluation in health care*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

Johansson P-O. *Evaluating health risks. An economic approach*. Cambridge University Press, 1995.

Kartman B, Andersson F, Johannesson M. Willingness to pay for reductions in angina pectoris attacks. *Medical Decision Making* 1996; **16**: 248-253.

Kriström B. A non-parametric approach to the estimation of welfare measures on discrete response valuation studies. *Land Economics* 1990; **66**: 135-139.

Lauraine G, Chestnut MA, Keller LR, Lambert WE, Rowe RD. Measuring heart patients' willingness to pay for changes in angina symptoms. *Medical Decision Making* 1996; **16**: 65-77.

NOAA (National Oceanic and Atmospheric Administration). Report of the NOAA panel on contingent valuation. *Federal Register* 1993; **58**: 4602-4614.

O'Brien B, Gafni A. When do the 'dollars' make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making* 1996; **16**: 288-299.

Park T, Loomis JB, Creel M. Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Economics* 1991; **67**: 64-73.

Poe GL, Severance-Lossin EK, Welsh MP. Measuring the difference (X-Y) of simulated distributions: A convolutions approach. *American Journal of Agricultural Economics* 1994; **76**: 904-915.

Stålhammar NO. An empirical note on willingness-to-pay and starting-point bias. *Medical Decision Making* 1996; **16**: 242-247.

Zethraeus N, Johannesson M, Henriksson P, Strand R. The impact of hormone replacement therapy on quality of life and willingness to pay. *British Journal of Obstetrics and Gynaecology* 1997; **104**: 1191-1195.

Appendix 1

Formally the bootstrap algorithm is as follows:

(I) Compute the original sample's mean WTP

$$\hat{\omega} = \frac{1}{2}\sum_{j=0}^{J}\left(p_{j+1} - p_j\right)\left(\tilde{r}_{j+1} + \tilde{r}_j\right), \tag{A1}$$

where a tilde denotes a (possibly) "corrected" ratio estimate.

(II) Resample N_j integers with replacement from the index sets $I_j = \{1,2,\ldots,N_j\}, j = 1,\ldots,J$, where J is the number of bid-groups, e.g. 8 as in our empirical example. Let I_j^{*b} denote the vector of resampled patient identification numbers in the bootstrap resample b for bid-group j. I_1^{*b} thus gives the index of the patients in bid-group 1 in resample b. Each index number corresponds to either a zero or a one in the original bid-group sample.

(III) Compute bootstrap ratios of yes-answers in each bid-group

$$\hat{r}_j^{*b} = \frac{1}{N_j}\sum_i x_{ij}^{*b}, j = 1,\ldots,J. \tag{A2}$$

If the bootstrap ratios, $\hat{r}_j^{*b}\, j = 1,\ldots,J$, are not non-increasing in p: correct according to (5).

(IV) Estimate bootstrap sample's mean WTP

$$\hat{\omega}^{*b} = \frac{1}{2}\sum_{j=0}^{J}\left(p_{j+1} - p_j\right)\left(\tilde{r}_{j+1}^{*b} + \tilde{r}_j^{*b}\right), \tag{A3}$$

where a tilde again denotes a possibly adjusted (bootstrap) ratio.

(V) Steps II - IV are repeated B times in order to obtain a set of (bootstrap) non-parametric mean WTP estimates.

A double sided $(1-2\alpha)\times 100$ percent BC confidence interval for the mean WTP is estimated by (Efron and Tibshirani 1993)

$$\left(\hat{\omega}^{*b}(\alpha_L), \hat{\omega}^{*b}(\alpha_U)\right), \tag{A4}$$

where the percentiles α_L and α_U are given by

$$\alpha_L = \Phi\left(2\hat{v} + \Phi^{-1}(\alpha)\right) \text{ and}$$

$$\alpha_U = \Phi\left(2\hat{v} + \Phi^{-1}(1-\alpha)\right). \tag{A5}$$

$\Phi(\cdot)$ denotes the standard normal cumulative distribution function. The bias-adjustment term $\hat{v}$ is given by

$$\hat{v} = \Phi^{-1}\left(\frac{\#\left\{\hat{\omega}^{*b} < \hat{\omega}\right\}}{B}\right). \tag{A6}$$

If $\hat{v} = 0$, i.e. $B/2$ of the bootstrap replications are less (and larger) than the original estimate, the confidence interval is the *percentile* interval. Thus, the percentile interval is a special case of the BC interval.

Appendix 2

The formulation of the willingness-to-pay question:

This question focuses on how much you value the continuation of your hormone replacement therapy. Assume that you have to pay the majority of the treatment costs for drugs and physician visits by yourself. Would you choose to continue your current treatment for menopausal symptoms if you had to pay SEK 1,000 each month[4] as patient charges for the treatment? Be aware that the money is taken from your own disposable income and hence decreases your private consumption.

Alternatives:

Yes ☐

No ☐

Statements of your motives:

_____

_____

Follow-up question:

Are you sure or uncertain that you want to pay SEK 1,000 for continuing the hormone replacement therapy?

Certain ☐

Uncertain ☐

[4] The bid varies from SEK 100 to SEK 10,000 in eight sub-samples.

Tables

Table 1: Share of yes answers by price; original values and adjusted values after smoothing.

Price	Total Original	Total Adjusted	Severe Original	Severe Adjusted	Mild Original	Mild Adjusted
100	0.88	0.88	1.00	1.00	0.80	0.80
500	0.80	0.80	0.86	0.92	0.75	0.75
1,000	0.70	0.70	1.00	0.92	0.40	0.50
1,500	0.60	0.60	0.75	0.75	0.55	0.50
2,000	0.50	0.50	0.50	0.50	0.50	0.50
3,000	0.23	0.28	0.25	0.39	0.20	0.20
5,000	0.33	0.28	0.50	0.39	0.17	0.17
10,000	0.18	0.18	0.50	0.39	0.00	0.00

Table 2: Mean values of background variables for all women and for women with mild and severe symptoms respectively.

	N	ΔTTO[a]	ΔRS[a]	Age	Income[b]	Household Size	Education[c]
All women	104	0.29	0.37	52.2	27,840	2.15	0.55
Mild symptoms	56	0.18	0.26	52.0	28,839	2.14	0.71
Severe symptoms	48	0.42	0.5	52.4	26,649	2.17	0.35
p-values[d]		0.00	0.00	0.77	0.36	0.90	0.00

[a] The difference in the quality of life score with and without HRT.
[b] Per month pre-tax household income.
[c] Coded 0 for primary education and 1 for secondary and university or higher education.
[d] Based on t-tests of mean differences between the mild and severe symptom groups.

Table 3: 95-percent confidence intervals for the mean WTP. Lower and upper bootstrap BC confidence intervals and original mean WTP estimates.

	Lower bound	Original	Upper bound
All women ($N = 104$)	2,539	3,508	4,813
Women with mild symptoms ($N = 56$)	1,444	2,346	3,870
Women with severe symptoms ($N = 48$)	3,092	4,838	6,767
Difference in WTP. Severe – Mild	313	2,492	4,409

Figures

Figure 1: The relationship between the price level and the proportion of women willing to pay each price. Full sample and subsamples with mild and severe symptoms. *N* (full sample) = 104 (*N*(Mild) = 56, *N*(Severe) = 48).