# IN SEARCH FOR MORE CONFIDENCE IN HEALTH ECONOMIC MODELLING:

## Reducing Uncertainty Associated with Modelling Studies

Mark J.C. Nuijten

Nuijten M.J.C.

In Search for More Confidence in Health Economic Modelling:
Reducing Uncertainty Associated with Modelling Studies
Thesis Erasmus University Rotterdam, March 2003



Cover design by Charlotte Nuijten

Printed by PrintPartners Ipskamp

# *In Search for More Confidence in Health Economic Modelling: reducing uncertainty associated with modelling studies*

Op zoek naar grotere betrouwbaarheid in gezondheidseconomische modellen:
de vermindering van onzekerheid inherent aan modelmatige studies

**PROEFSCHRIFT**

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 26 maart 2003 om 13.45 uur

door

Marcus Johannes Catharina Nuijten
geboren te Bergen op Zoom

**PROMOTIECOMMISSIE**

| | |
|---|---|
| PROMOTOR: | Prof.dr. F.F.H. Rutten |
| COPROMOTOR: | Dr. M.J. Al |
| OVERIGE LEDEN: | Prof.dr. M.G. Hunink |
| | Prof.dr. B. van Hout |
| | Dr. M. Koopmanschap |

Aan mijn vrouw en dochter

# PUBLICATIONS

Chapters 2 to 12 are based on the following articles:

| | |
|---|---|
| Chapter 2 | Nuijten MJC, Berto P, Berdeaux G, Hutton H, Fricke F-U, Antonanzas Villar F. Trends in decision-making process for pharmaceuticals in Western European countries: A focus on emerging hurdles for obtaining reimbursement and a price. HEPAC, 2001; 162-169. |
| Chapter 3 | Nuijten MJC. Assessment of clinical guidelines for continuation treatment in major depression. Value in Health, 2001; Jul-Aug 4(4): 281-294. |
| Chapter 4 | Nuijten MJC. The selection of data sources for use in modelling studies. PharmacoEconomics, 1998 March; 3 (3): 305-316. |
| Chapter 5 | Nuijten MJC. Bridging decision-analytic modelling with a cross-sectional study. Application to Parkinson's disease. PharmacoEconomics, 2000 Mar 17(3): 227-236. |
| Chapter 6 | Nuijten MJC, Hutton J. Cost-effectiveness analysis of interferon beta in multiple sclerosis: a Markov process analysis. Value in Health, 2002 Jan-Feb; 5(1): 44-54. |
| Chapter 7 | Nuijten MJC, Rutten F. The incorporation of confounding variables in Markov models. Accepted by PharmacoEconomics, November 2002. |
| Chapter 8 | Nuijten MJC, Hardens M. Measuring sensitivity in pharmacoeconomics studies: An integration of point-sensitivity and range-sensitivity. PharmacoEconomics, 1997 November; 12 (5): 555-564. |
| Chapter 9 | Nuijten MJC. Measuring sensitivity in pharmacoeconomic studies: Refining point-sensitivity and range-sensitivity by incorporating probability distributions. PharmacoEconomics, 1999 July; 16 (1): 33-41. |
| Chapter 10 | Nuijten MJC Al M, Rutten F. Inter-variable uncertainty in decision-analytic modelling: The concept of second-order sensitivity. Submitted to Medical Decision Making. January 2003. |
| Chapter 11 | Nuijten MJC, Rutten F. Combining a budgetary impact analysis and a cost-effectiveness analysis using decision-analytic modelling techniques. PharmacoEconomics, 2002; 20(12): 855-867. |
| Chapter 12 | Nuijten MJC Pronk MH, Brorens MJ, Hekster YA, Lockefeer JH, de Smet PA, Bonsel G, van der Kuy A.. Reporting format part for Economic Evaluation. Part II: Focus on Modelling Studies. PharmacoEconomics, 1998 Sep; 14 (3): 259-268. |

# CONTENTS

Chapter 1

# Introduction

## BACKGROUND AND RATIONALE

Every government is eager to control the increase of expenses by the implementation of central cost containment policies particularly in relation to pharmaceuticals. For the most part, those measures have relied on budgeting or price controls, including negotiated prospective budgets for hospitals, centralized negotiated budgets for ambulatory physicians including drug prescriptions, and limitations on payments for particular medications. Because those traditional central cost containment measures were only partially successful, due to lack of incentives, the health authorities in Europe started to establish incentives for efficient healthcare delivery. Both traditional and recent containment measures focus especially on the pharmaceutical drugs sector in many countries, as these constitute a health technology that is relatively easy to introduce and implement compared to other forms of care.

Financing prescription medicines in ambulatory care has been a central responsibility based on the traditional clinical trial outcomes (efficacy/safety parameters) used for registration. Although there is large variety between the various countries, there are three related trends: decentralization of the healthcare decision-making process, prescription restrictions, and extra data requirements. One can distinguish various extra data requirements which all relate to the use of the drug in real daily practice, while the traditional clinical trial outcomes are only derived from randomised clinical trials. At a central level the demand for cost-effectiveness and budgetary impact data is increasing. The requirement for health economic data resulted to formal reporting requirements in some countries already (e.g. Canada, Australia, The Netherlands, UK, Portugal and Finland). Although the most evident impact of health economic studies is expected to be on central reimbursement audiences, evidence for the use of health economic studies by other audiences is expected to increase (e.g. patients, hospitals, insurers, formulary committees). This background information is described in more detail in **Chapter 2**, which is the introduction to this thesis.

Pricing and reimbursement have been based, until recently, on the traditional clinical trial outcomes used for registration (efficacy, safety and quality parameters), which are called the first three hurdles. Consequently the growing burden on manufacturers to demonstrate the cost-effectiveness of their products is called the fourth hurdle in drug development. This fourth hurdle may have considerable consequences for all players involved. Health economic data (effectiveness and resource utilization) may be collected alongside a randomised clinical trial. (RCT).[1 2 3 4] However, performing a RCT is not always ethical or feasible. For example, in a comparison of surgical techniques with drugs, randomisation can be performed, but blinding is impossible. Another example is that randomisation is not considered ethical in potential life saving interventions. Also the execution of a time consuming and costly RCT for every minor change in technology may not be justified or feasible. Considering the prospective approach, the concept of validity should be addressed: internal validity is the extent to which the analytic inference derived from the study sample is correct for the target population. External validity or generalisability is the extent to which a cost/effectiveness analysis found in the

study sample is also true in the external population. Clinical trials have a low external validity because they have strict inclusion and exclusion criteria and treatments are protocol driven, leading to overestimation of units of healthcare utilization. Therefore, both clinical and economic outcomes may not be typical and do not correspond to usual practice. Hence it should always be considered that, due to its restriction on external validity, the outcomes may not be representative of a target patient population who are going to use the drug. Although randomisation is usually applied to balance confounding variables, inclusion criteria of patients and selection of investigators are fairly rigid, double blind and placebo control is often the rule. In addition treatment patterns are usually driven by the protocol.[2-5] Finally pharmacoeconomic analyses are usually interested in examining the discounted costs and benefits beyond the duration of patient follow-up, for example, when estimating the cost per life years saved in chronic diseases.

An alternative design is the naturalistic economic trial, which is a randomised trial with the primary objective of gathering "real world" or representative cost and effectiveness data. The naturalistic trial is an economic trial with the primary objective of gathering "real world" or representative cost and effectiveness data. The design is comparable with a traditional clinical randomised clinical. Naturalistic prospective studies usually share a number of features with clinical trials, such as parallel design and randomisation. However, they differ on other characteristics: population selection is based on representativeness, and clinical criteria are used instead of academic definitions for inclusions of patients. The protocol of a naturalistic study contains only a limited number of scheduled visits and procedures in order to minimise the impact on the treatment pattern corresponding with daily practice by protocol driven costs or protocol initiated changes in treatment patterns. Naturalistic studies have the additional advantage of providing valuable clinical data obtained in a more representative context (external validity). However since study drugs are usually not approved for registration at the time of performing of economic studies, the use of prospective naturalistic trials is limited, making submission of data on effectiveness and expected costs at the time of reimbursement not feasible. Finally the performing of a naturalistic economic trial may not always be ethical, for example randomisation may not be considered ethical, when clinical trial data show a substantial clinical advantage compared to standard therapy.

Hence in practice it is not always possible to derive information from scientifically sound prospective studies. In these cases decision-analytic models may be used to provide the necessary cost-effectiveness information using various existing data sources for clinical and economic information. Modelling studies are based on decision analysis, which is a well-recognized method for analysing the consequences of decisions that are made under uncertainty.[6] It is an explicit, quantitative, prescriptive approach to healthcare decision-making and allows both clinical and economic consequences of medical actions and attitudes to be analysed under conditions of uncertainty. From treatment algorithms a model can be constructed which considers the timings of actions and their consequences over time. A decision-analytic model consists of a series of branches, each representing different

options (decisions or events), which arise at different points, referred to as nodes (e.g. decision node and chance nodes). In effect, a model shows the consequences and complications of different therapeutic interventions, and it should correspond as much as possible, to the real life situation of the disease. Models may take the form of simple decision-analytic trees or they may be very complex Markov models. While decision-tree models are appropriate for acute episodes, Markov models are the first choice for pharmacoeconomic analysis of chronic diseases.

Projections about a drug's effectiveness and expected costs can be modelled using realistic and explicit assumptions based on data from clinical studies. In addition modelling often helps overcome the practical limitations of prospective studies, particularly for chronic conditions like Parkinson's disease that may require longer-term extrapolations of drug effects and cost implications. Data sources for the variables being used in a model may be meta-analysis, databases, clinical trials and / or Delphi panels.

In this thesis the focus is on the use of health economic models for reimbursement decisions, but models also may be a valuable tool to support the decision-making process in many areas and phases of development of a new drug from clinical program development to the design of a disease management program.[7 8] An example is the use of a model for the assessment of the macro-economic consequences of reimbursement of a new drug on a nation-wide scale, which is included in this thesis **(Chapter 11)** and shows that one model can be developed for various purposes.

In order to increase the scientific quality and integrity of pharmacoeconomic studies, national guidelines have been developed by various countries. The intent of the guidelines is to provide guidance to doers and users of studies, by laying out a general "state of the art" regarding methods, and by providing specific methodological advice on many matters. Although various groups have published recommendations on the good practice of economic evaluations [9 10 11], those guidelines mainly focus on prospective studies and contain only a limited number of recommendations for the execution of modelling studies. This may partly explain why the acceptance of modelling studies has generally been lower than prospective studies, which is especially due to the level of uncertainty associated with modelling studies.[12] In this thesis the role of modelling in economic evaluation is explored by discussing the concerns of models, which mainly relate to the trade-off between internal and external validity: concerns about the inappropriate use of clinical data, concerns in observational data, concerns about the difficulties in extrapolation, concerns about the transparency or validity of the model.

Over the next several years, most submissions for reimbursement in countries like the Netherlands will likely be based on modelling since pharmacoeconomic data were generally not required or included in the clinical research programs of products now approaching launch. If reimbursement of pharmaceuticals will be based predominantly on economic data derived from modelling studies, it is vital to scrutinise and refine the modelling approach carefully, especially the handling of uncertainty. Therefore the objective of this thesis is to identify and explore various types of uncertainty in

modelling studies. Also methodologies were presented, which may reduce the level of uncertainty associated with modelling, and which may consequently increase the reliability of health economic outcomes of modelling studies. This may improve the acceptance of modelling studies, even in the absence of formal guidelines. The concepts are empirically illustrated using Markov models in chronic diseases: depression, Parkinson's disease and MS. Published real data were used, whenever possible, but extrapolation methods were used in absence of real data in order to illustrate the relevant issues.
The thesis presents various methodologies, which are illustrated using a Markov modelling study in depression (**Chapter 3**). The primary objective of this study was to assess the appropriateness of the existing Dutch clinical treatment guidelines for GP's in depression from a health economic perspective. The assessment was based on a Markov model based on decision-analytic techniques. The primary perspective of the study was that of the third party payer, while the secondary perspective was that of the society in 1999. A scenario analysis was performed to test if an extension of the continuation treatment to maintenance treatment might result in a less or more favourable cost-effectiveness outcome of the treatment guideline.

## UNCERTAINTY ASSOCIATED WITH THE DATA SOURCES YIELDING THE INPUT DATA FOR THE MODEL

The reliability of the estimates depends on the choice of the data sources (selection criteria, external validity). The data may come from a variety of sources and are subject to varying degrees of uncertainty. The following issues relating to the selection of data sources are addressed in **Chapter 4**.

- Classification of data being used in modelling studies
- An assessment of the various data sources, including advantages and disadvantages
- A general strategy on how to determine the appropriateness of a data source for a model
- Recommendations on a strategy for data source selection and a transparent reporting format on data source selection.

A main concern on the use of a modelling study is the use of a Delphi panel to gather data not available from actual existing data sources. There is no guarantee that the panel assessment of resource utilisation and utilities is an accurate reflection of reality. The use of a Delphi panel may particularly be controversial to estimate QALYs, which may become the primary final outcome for pharmacoeconomic studies. The existing Canadian, Australian and Dutch guidelines and the draft UK guidelines strongly recommend an estimate of the incremental cost per QALY gained.[13 14 15 16] **Chapter 5** presents a cross-sectional study as an alternative data source for a modelling study. It can be used to determine resource utilization and utilities for Markov health states. The overall design may be considered a hybrid between a naturalistic prospective study and a modelling study by maximising the pros and minimising the cons of both types of design. This hybrid design was based on bridging the probabilities derived from literature and clinical trials with information on costs and utilities from a cross-sectional study. This design was illustrated using a Markov model for Parkinson's disease.

**Chapter 6** presents a health economic modelling study for multiple sclerosis (MS), which is an application of the two previous chapters on the handling of data from various secondary data sources, which allowed us to develop a life time Markov model for MS in order to estimate the cost-effectiveness of interferons. The classification of disease severity in this study corresponds with semi-Markov health states in our model. The probabilities of clinical events were based on interferon beta clinical trial data in respectively relapse remitting and secondary progressive MS and observational data from the literature. The utilities and costs were derived from two separate cross-sectional studies. This study showed that a health economic model can be completely based on published data from various sources.

## UNCERTAINTY ABOUT CONFOUNDING VARIABLES

The costs and utilities may be not only a function of the defined health states in a Markov model, but also of other variables, which may act as confounding variables when they are not taken into account. In **Chapter 7** a strategy is presented for the incorporation of a confounding variable in a Markov health states by means of health state specific relationships between the confounding variable and costs as well as time-dependent values of the confounding variable. The objective is to show whether the outcomes of a health economic model are severely biased, when a confounding variable is not taken into account. The concepts were illustrated using a hypothetical Markov model for Parkinson's disease.

## UNCERTAINTY REDUCTION THROUGH SENSITIVITY ANALYSIS

Sensitivity analysis is currently the most widely used method to deal with uncertainty in economic evaluations. A sensitivity analysis is based on modification of the basic clinical and economic estimates of parameters to judge the effect on study results of alternative assumptions for the range of potential values for uncertain parameters.[17] Most recent pharmacoeconomic publications contain sensitivity analyses on only a limited number of variables, without justification of the choice of selected variables or the chosen range of each variable. In addition, the interpretation of the results of a sensitivity analysis remains subjective, as there is no scale for measurement of sensitivity and consequently there is no threshold value for discriminating between "sensitive" and "non sensitive" variables.

In **Chapter 8** a more objective method is presented for sensitivity analyses to minimize the amount of potential subjectivity. This new procedure, employing point-sensitivity and range-sensitivity, allows an objective judgment of the sensitivity of all variables of a model, permitting the variables to be ranked according to the degree of sensitivity. However, a limitation of this method is that the sensitivity measurement is based on a uniform distribution of the variables, which may actually have different distributions. A second limitation is that the overall sensitivity measure is based on a

subjectively chosen range, which excludes the impact of values outside the range on the overall sensitivity. **Chapter 9** presents a refinement of the method by the incorporation of probability distributions, which allow a more accurate assessment of the level of uncertainty in the model. In addition, a bootstrap method is used to create probability distribution for fixed input variables based on a limited number of data points.

A constraint of the above-mentioned and other methods for conducting sensitivity analyses is that those methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taking into account the existing relationships between those variables. In **Chapter 10** the concept of inter-variable uncertainty is introduced and explored. A methodology was presented considering this type of inter-variable uncertainty and showing the practical implications of it on the execution of sensitivity analysis. The initial method was based only on uniform distributions, and subsequently various methods were presented incorporating the real distributions.

## UNCERTAINTY ASSOCIATED WITH EPIDEMIOLOGY

**Chapter 11** presents a methodology for an appropriate assessment of the budgetary impact of a new drug, which simultaneously can be used for traditional cost-effectiveness analysis. A model was constructed for a new hypothetical drug in Parkinson's disease, which allowed us to determine the budgetary impact and the cost-effectiveness of this new anti-Parkinson drug (AP) from the perspective of the society in The Netherlands. The methodology consists of two steps: 1) a simple population model (Markov model) was constructed to validate the epidemiological data by proving the consistency between prevalence and incidence of Parkinson's disease for the Dutch population; 2) this model was extended to a more complex model (semi-Markov model) by incorporation of disease progression for Parkinson's disease and all relevant economic and clinical measures.

## UNCERTAINTY ASSOCIATED WITH THE TRANSPARENCY OF THE MODEL

The acceptance of modelling studies is generally lower than prospective studies not only because of the use of secondary data, but also because the reports of modelling studies do not always have sufficient transparency. Hence a standardised reporting format for modelling studies may improve the acceptance of those studies. **Chapter 12** presents an example of a reporting format for economic evaluation based on a modelling design. Since, contrary to clinical trial data, there are no accepted methods for data collection and analysis for modelling studies, a more disaggregate reporting presentation is required for modelling studies. The report must address the various types of uncertainty in modelling studies and contain a justification of the choices in the selection of data sources, model design and its assumptions, managing confounding variables and execution and interpretation of

sensitivity analysis. The proposed format consists of a number of headings, which are followed by a brief recommendation on their content.

In the discussion (**Chapter 13**) the use of health economic models for reimbursement purposes is addressed taking into consideration the levels of uncertainty, which were explored in this thesis.

## References


1 Office. Robsinson R. economic evaluation and health care. What does it mean? British Medical Journal. 1993; 307: 670-3
2 Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomised controlled trials. Medical Care 1992; 30 (3): 231-43
3 Drummond MF, Davies L. Economic analysis alongside clinical trials. Revisiting the methodological issues. International Journal of Technology Assessment in Health Care. 1991; 7 (4): 561-573
4 Drummond MF, Ludbrook A, Lowson K, Steele A. Studies in economic appraisal in health care 1991; 7 (4); 561-573
5 Sloan FA, valuing health care. Costs, benefits and effectiveness of pharmaceuticals and other medical technologies. Cambridge: Cambridge University Press, 1995
6 Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-65
7 Bowine Michel. Modelling in Medical Technology Assessment. Thesis. January 24, 1996
8 Nuijten MJC, Starzewski J. Applications of Modelling Studies. PharmacoEconomics, March 1998; 3 (3): 289-291
9 Clemens K, Townsend R, Luscombe F, Mauskopf J, Osterhaus P, Bobula J. Methodological and conduct principles for pharmacoeconomic research. Pharmacoeconomics 1995; 8 (2): 169-74
10 Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996; 313: 275-83
11 Jefferson T, Demichelli V. Are guidelines for peer-reviewing economic evaluation necessary ? A survey of current editorial practice. Health Econ 1995; 4: 383-8
12 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 Econ 1997 May-Jun;6(3): 217-27
13 NICE guidelines. NHS National Institute for Clinical Excellence (NICE). Guidance for Manufacturers and Sponsors. Available from: URL: http://www.nice.org.uk. June 2001
14 Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Commonwealth Department of Human Services 1995
15 Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. 2st ed. Ottawa: CCOHTA. 1997
16 National Health Insurance Council. Report on pharmaco-economic research guideline. March 2. 1999.
17 Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health care Technology: A Report on Principles. Ann Intern Med. 1995; 122: 61-70

Chapter 2

# Trends in Decision-making Process for Pharmaceuticals in Western European countries: A Focus on Emerging Hurdles for Obtaining Reimbursement and a Price


## SUMMARY

Healthcare financing has become a topic on the political agenda in Western Europe in recent decades. For every government it has become a subject of continuous concern because the costs of health services and healthcare are an increasing important part of the collective burden of the economy. Most cost containment measures have relied on budgeting or price controls. Because those traditional central cost containment measures were only partially successful, due to lack of incentives, the health authorities in Europe started to establish incentives for efficient healthcare delivery by means of decentralisation of the healthcare decision-making process and the implementation of market mechanisms. Both traditional and recent containment measures focus especially on the pharmaceutical drugs sector in many countries. Recently there have been three parallel trends showing increasing data requirements at a central level, more decentralisation of the responsibilities and decision-making process and prescription restrictions. We address especially the increasing central data requirements and decentralisation of the pricing and/or reimbursement decision-making process. At a central level the demand is increasing for cost-effectiveness and budgetary impact data, which has already resulted in formal reporting requirements in some countries. The findings are based on the literature and expert opinion of local health policy experts in the various European Union countries, who co-authored this manuscript

# INTRODUCTION

Healthcare financing has been a topic on the political agenda in Western Europe in recent decades. For every government it has become a subject of continuous concern because the costs of health services and healthcare are an increasing important part of the collective burden of the economy. There are a number of reasons, which may explain the increase in healthcare costs. The ageing of the population may be associated with an increase in morbidity and associated healthcare costs, although a recent study shows that population ageing may contribute much less to future growth of the healthcare sector than claimed by most observers. This study showed that costs may especially depend on remaining life time.[1] In addition to that health technology is also contributing to increases in health expenditures. In contrast to other economic sectors, new health technology scarcely reduces costs (e.g. personnel, energy), at least in the short term. Another reason is that the patient has become more knowledgeable. The asymmetry of information between physician and patient has become much less pronounced, because the patient has become more knowledgeable than in the past by means of better education and media. In addition the patient has become aware more demanding and is claiming maximum quality of life regardless of the costs. Because insurance is generally compulsory in the European Union, the patient is usually insured, and because third party payers pay the majority of the costs there is generally no direct demand control for a patient.

Every government is eager to control the increase in expenses by the implementation of central cost containment policies, which in particular relate to pharmaceuticals. For the most part these measures have relied on budgeting or price controls, including negotiated prospective budgets for hospitals, centralised negotiated budgets for ambulatory physicians including drug prescriptions, and limitations on payments for particular medications. Because traditional central cost containment measures were only partially successful, due to lack of incentives, the health authorities in Europe started to establish incentives for efficient healthcare delivery by means of decentralisation of the healthcare decision-making process and the implementation of market mechanisms.

## Decentralisation

Most European healthcare systems, except that of Germany, have been based on the so-called Beveridge model, a central system of care consisting of public institutions financed by state budget. Local authorities councils have recently obtained increasing authority to implement policies and freedom to structure local healthcare organisation (e.g. Italy, Sweden) [2] (in Italy: Decreto Legislativo, 19 June 1999, no. 229 [3]). In France in 1996 a reform created the Agence Régionale d'Hospitalisation (ARH) and the Union Régionale des Caisses d'Assurances Maladies (URCAM). Their main rules are to organise the hospital supply according to a fixed budget and regional healthcare needs. They must also implement locally healthcare priorities that are fixed at a national level and monitor the quality of healthcare delivery.[4] While local health authorities in most countries usually receive funding from the central health authorities based on a per capita basis, in Italy and Sweden they can supplement this

funding with local taxes and health service charges, which may vary locally depending on local budgetary needs (Decreto Legislativo. 19 June 1999, no. 229).

In Germany the principle of statutory social insurance is called the "Bismarck system". Main features of the system are the financing via contributions by employees and employers and a mix of service supply by private and public services. Although in Germany there are no formal layers below the state level (*Bundesländer*), statutory sickness funds are trying to implement local healthcare organisations, such as practice networks including hospitals.

## Market Mechanisms

A growing number of countries have recently adopted some form of purchaser-provider split, although the others have retained the more conventional budget setting structure. The basic idea is to create a demand side (purchasing agency) that is separate from the supply side (providers). The reforms split the system into purchasers and providers. For example, in the United Kingdom health authorities and general practitioner fundholders became purchasers, while the Trust hospitals and directly managed units became the providers. General practitioner fundholders are now grouped into primary care groups, which will become trusts (PCTs). Primary care trusts will control the whole budget (except for some specialist services); they will provide primary care and purchase hospital care and other services including drugs. Also in Sweden and Italy there are signals at local level of establishing purchaser-provider split by local healthcare councils. As stated, for example, in the two Italian laws on the "reordering of the National healthcare system" (Decreto Legislativo 30 December 1992, no. 502; Decreto Legislativo 7 December 1993, no. 517.[5] [6] In other countries, with the more conventional budget setting structure, market mechanisms may be limited to drugs only and especially hospital drugs. In France private (and also public) hospitals buy drugs through tenders. Private hospitals negotiate with the local sick fund payer according to their activities and receive a fixed budget per patients for drugs.

## Other Related Strategies

The decentralisation of the healthcare decision-making process and the implementation of market mechanisms resulted in various other related strategies.

### *Shift to Primary Care*

Throughout the 1980s and 1990s the trend has been away from hospital-based consultations and towards primary care. The function of the general practitioner has switched from a physician treating patients to a gatekeeper of the healthcare system, whose main responsibility is avoiding inappropriate optimal referring of patients to secondary care in order to save costs. For example, in Germany the so-called *Hausarztmodelle*, meaning that the family physician should navigate the patient through the system, emerged and to avoid hospitalisation out-patient surgery was encouraged through special contracts by sick funds together with statutory health insurance physicians associations to induce a

shift from in-patient to out-patient care. Further, health authorities have encouraged the development of practice networks and integrated care models integrating in-patient and out-patient care. The respective legal requirements have been implemented into social security laws in recent years (http://www.bmgesundheit.de/engl/healthcare.htm; German social security law, SGB 140a-h).[7 8] Due to the legal framework of the German healthcare system these measures had only limited success up to date. On the other hand, a country such as France is still a pure fee-for-service country, and there is almost no barrier to any health goods. French citizens can access either specialist or hospital directly without any inputs from the general practitioner. In Denmark, for instance, in the past decade it was considered an improvement in quality of health services that the patient was allowed to choose the type of care, primary or specialised. However, a co-payment was introduced for those patients using directly specialised care. After some years of follow-up the results show that the majority of patients keep attending the primary care services and only a lower than 10% uses specialised care. This might be considered a compromise between freedom to choose by the patient and respecting the principles of cost containment.

### *Co-payment*

The health systems of the Western European economies have generally offered universal coverage with a comprehensive benefits package. Recent financial pressures have prompted most countries to look for ways to limit public sector financial liability by limiting benefits and influencing the demand side of the healthcare market, which resulted in an increase of co-payment by the patient. The bulk of co-payment is limited to the ambulatory healthcare setting, especially pharmaceuticals, although in Germany and France there exists also co-payments for in-patient care (German social security law, SGB 31:39 V). The objective of co-payment is to establish financial incentives for patient's demand control, which in many countries fail because of complementary insurance for co-payments. In France most French citizens have complementary private insurance, which is paid both by the employee and the employer, at least in large companies. This complementary insurance covers a significant part of the patient co-payment left by the Sécurité Sociale.[9]

### *Clinical Guidelines/Disease Management*

Clinical guidelines outline the proper care of medical conditions and performance of clinical procedures. The intended goal of guidelines is to reduce inappropriate care and to improve patient outcomes. In addition, these guidelines are potential tools for reducing the costs of healthcare and for improving medical education.[10]

## Pharmaceuticals

Pharmaceuticals expenditures have risen in Europe and the United States during the 1990s. In most countries pharmaceutical spending has reached 10-15% of the total healthcare budget (in some countries, such as Greece, Portugal and Spain, this proportion is higher than 20%; OCDE health data

2000). Both traditional central containment measures and the cost containment measures resulting from the above strategies (decentralisation/free market regulation) focus especially on the pharmaceutical drugs sector in many countries, as these constitute a health technology that is relatively easy to introduce and implement compared to other forms of care. Recently there have been three parallel trends showing increasing data requirements at a central level, more decentralisation of the responsibilities and decision-making process and prescription restrictions.

The objective of this manuscript is to give an overview of the current pricing and reimbursement environment for pharmaceuticals in Western European countries. We address especially the increasing central data requirements and decentralisation of the pricing and/or reimbursement decision-making process. The information was obtained from published literature, local published or available policy documents. A survey was conducted with local health policy experts in the key countries to validate the findings from the literature, add missing information and update the information, if necessary. The local health policy experts co-authored this manuscript and are involved in pricing and reimbursement issues in their domestic markets. We focussed on the key countries (United Kingdom, Germany, France, Italy and Spain), but also included relevant information from other countries (e.g. The Netherlands, Sweden and Denmark).

# TRADITIONAL CENTRAL MANAGED POLICIES

## Pricing and Reimbursement

Registration and pricing/reimbursement decisions are currently distinct processes: registration of a new drug is based on quality of manufacturing and efficacy/safety data from randomised clinical (phase III) trials and a product has market approval after registration. The drug is available on the market but does not have reimbursement status, and consequently the patient must pay for the drug alone. Therefore registration is only the first entry barrier for a new drug, which is followed by hurdles due to pricing and reimbursement procedures. Although it is difficult to disentangle pricing and reimbursement decisions, a recent overview by our group for Europe clearly indicated that reimbursement and pricing may be considered as two separate procedures: decisions are made by different bodies, different laws apply, different reporting data are required, and reimbursement and pricing are sequential decisions.[11 12 13 14] For example, in France the Transparency Commission decides on reimbursement, while the price is negotiated with the Comité Economique des Produits de Santé (CEPS). An exception is Italy, where pricing and reimbursement are strictly related to, the Commissione Unica del Farmaco (CUF), which is the Italian drug regulatory agency, in collaboration with the Comitato Interministeriale per la Programmazione Economica (CIPE; a body of the Ministry of the Treasury.[15]

## Cost Containment Measures

Although there does not seem to be any systematic approach in selecting policies to curb the costs of pharmaceuticals, governments in general more and more have resorted to central demand and supply-oriented policies to limit drugs expenditures. An example of a supply-oriented policy constitutes direct price control, whereas a demand-oriented policy would imply the introduction of a limited list of reimbursable drugs (i.e. a formulary). Hence governments have opted for limiting demand as well as supply of pharmaceuticals available under public reimbursement schemes.

### *Drug Pricing*

Pricing of drugs is often limited by either price negotiation or price laws, which define an upper limit based on a reference price basket consisting of prices in neighbouring countries or a European average price, except for the United Kingdom, Germany and France. In the United Kingdom and Germany pharmaceutical companies are free to price their drugs that have received marketing authorisation. The only limitation is that medicines may only be sold at one price sold throughout the country. However, in the United Kingdom, pricing is constrained by pharmaceutical company's total profit in the domestic market, which can vary between 17% and 21% according to United Kingdom based research and development activity and exporting. France is a price state regulated country: unit price is negotiated with the Comité Economique des Produits de Santé on the basis of Amélioration du Service Médical Rendu (ASMR) and the drug budget impact (Décret du CEPS, 2000). In Italy both pricing procedures apply, which depend on registration: (a) price negotiation, which applies to any innovative drug approved by European Agency for the Evaluation of Medicinal Products (EMEA) or introduced by a mutual recognition procedure and (b) directly calculated price, by means of the so-called "European average", which applies when market drug authorisation is allowed directly from national government (Ministry of Health) instead of the European Agency for the Evaluation of Medicinal Products.[15] Pricing of in-patient drugs is less centrally regulated and corresponds more with a market model. Pharmaceutical companies are free to set drug prices, and drugs can be sold at different prices to various hospitals based on a negotiation process. The governments do not intervene in regards to the price that is negotiated, although the results of the tender are deeply influenced by an out-patient price. Discounts can widely vary and better discounts are obtained for widely used drugs. Competition between manufacturers has the strongest influence on drug price negotiations, followed by the volume of sales and package deals. Because in most countries in-patient drugs are within the hospital budget, hospitals may not always be able to finance premium priced drugs. Central authorities or local health insurers have recently taken over financing of those drugs. Drugs are excluded from the traditional hospital budget and prescriptions are often limited to selected centres in order to control the number of prescriptions. An example is Remicade, an expensive drug for the treatment of severe rheumatoid arthritis, which requires day-care because of the intravenous route of administration. The use of this drug results in an average annual costs of 30,000 euros per treatment, which exceeds the budget of

most hospitals. Dutch health authorities took over financial responsibility for this drug to guarantee it is available for the patients, especially because of shortage of Enbrel, an out-patient drug for the same indication.

### *Reference Price System and Co-Payment*

Drugs are usually grouped according to chemical structure and the reimbursement level depends on necessity of a treatment. Essential drugs may be fully reimbursed, while complementary drugs may be partly reimbursed, and non-essential drugs may be fully paid by the patient. Often usually a number of categories, including chronically ill, and pensioners do not have to pay co-payment for drugs. This system is called the reference price system, which is applied usually only to drugs when there are several brands with the same compound available.[16] [17] This system does not set drug prices; rather, it sets the reimbursement levels at which the sickness funds pay for each out-patient prescription drug (consumers pay the amount by which the product prices exceed the reimbursement levels). Drugs in each group are all reimbursed on a fixed amount. Reimbursement decisions regarding new innovative drugs will be based on judgement of a clinical benefit compared with standard therapy in the target patient population (indication). The judgement of the clinical benefit is based on traditional clinical outcomes derived from phase III clinical trials used for registration: efficacy, safety and quality. In addition other clinical criteria are taken into consideration: route of administration, or other relevant clinical information. When the evaluation of a drug is positive from a clinical point of view, the drug price has been the critical factor for final decisions on reimbursement until recently. The reference price system has two primary functions: first, to lower the prices of drugs by inducing price competition, and, second, to encourage greater use of generic drugs by making consumers pay a greater share of the cost of higher-price brand-name drugs. Patients usually can have co-payment for drugs refunded through additional private insurance. Therefore this demand-oriented policy is usually not an effective cost containment measure, because patients remain relatively insensitive to the co-payment system, when co-insurance is possible. There is usually no or limited co-payment for drugs supplied during hospitalisation. It is important to note that while we describe general mechanisms, that there is a wide variation across EU.

### *Positive and Negative Lists*

There is an increasing trend to the development of positive and negative lists, which aim at reducing the number of reimbursement drugs as well as total spending on pharmaceuticals. The selection of drugs for a list is mainly on the basis of efficacy/safety parameters. At central level a positive list contains drugs which will be reimbursed, while drugs on a negative list (e.g. Black List in the United Kingdom and the former List 1B in The Netherlands) must be paid fully by the patient. The Black List contains drugs, which are not reimbursed, although licensed. These are mainly old and ineffective products for which better and cheaper alternatives are available. In the United Kingdom there is also a grey list of drugs for use in limited circumstances, for example, Viagra. In Germany there is an

established negative list [social security law, SGB 34(3) V] which is technically updated by a federal committee (Bundesausschuss Ärzte/Krankenkassen). Further, the first draft of a positive slist is due for submission on 30 June 2001, which is set up by a special committee (German social security law, SGB 33a V).[18] The increase in pharmaceuticals expenditures in Europe up to 10-15% of the total healthcare budget [9], shows that the traditional central cost containment measures were not sufficient to control the drug costs leading to drug policy reform, which is described below.

## CURRENT RESTRUCTURING OF DRUG POLICIES

Recently there have been three parallel trends showing increasing data requirements at a central level, more decentralisation of the responsibilities and decision-making process and prescription restrictions.

### Data Requirements

Pricing and reimbursement have been based until recently on the traditional clinical trial outcomes (efficacy, safety and quality parameters) used for registration. We can distinguish various data requirements which all relate to the use of the drug in real daily practice, while the traditional clinical trial outcomes are only derived from randomised clinical trials.

#### *Effectiveness*

There is an increasing demand for effectiveness data. Efficacy and effectiveness are two different concepts. Both, however, have an impact on the effect of a drug. In the case of efficacy, the effect is examined under ideal conditions in a homogeneous group of patients, and usually with the assistance of intermediate (surrogate) end-points. Effectiveness data, on the other hand, offer a clearer picture of the actual value of a drug because the effect is examined under more realistic conditions using a more heterogeneous group of patients. This information about use in common practice also provides more insight into whether the aim of the treatment will ultimately be achieved. Effectiveness research is therefore oriented towards definitive outcomes such as a reduction in morbidity and mortality. A recent Italian law allows phase III clinical experimentation (both randomized controlled trials and uncontrolled, observational studies) at present conducted only in hospitals and/or specialty clinics, to be conducted also in an out-patient setting. This kind of experimentation provides data from a large sample of population in a "naturalistic setting", therefore providing good insight into the "effectiveness" of the pharmaceutical technologies (SOLE 24 Ore Sanità, no. 3, 23-29 January 2001).[19] In France the Transparency Commission considers the public health value of the drug, which is called an Amélioration du Service Médicale Rendu (ASMR), which may be considered as a comprehensive effectiveness measure.

#### *Cost-Effectiveness Data by a Health Economic Analysis*

The fourth hurdle in drug development is the growing burden on manufacturers to demonstrate the cost-effectiveness of their products before acceptance for reimbursement or, less relevant, pricing, may have considerable consequences for all players involved. Health economic data should permit reliable,

reproducible and verifiable insight into the effectiveness of a drug, the costs that will result from its use, and the possible savings that will be made compared with other drugs and/or treatments. Health economic studies are already being used for the reimbursement of new drugs in Australia and Canada. Those countries have official requirements for submission of health economic data since the early 1990s.[20 21] There is currently a trend towards an increasing demand for health economic data in the decision-making process information in Europe and in several countries to formal reporting requirements now (the United Kingdom, Finland and Portugal) or in the near future (The Netherlands and Norway).[11 21] This is also true in Italy. In fact, a recent provision (February 20001) by the Comitato Interministeriale per la Programmazione Economica (CIPE; the Ministry's Board of Economic Planning) claims that a new drug is admitted to reimbursement when its cost-effectiveness ratio is favourable in comparison to other drugs already admitted for the same indication, or it is useful to prevent and treat symptoms and pathologies not already treatable by others drugs (Delibera CIPE 1 February 2001, no. 3/2001).[22] In France health economic data help at supporting premium price, as part of the negotiation process.

### *Budgetary Impact Data by a Financial Analysis*

In addition to the cost-effectiveness of a new drug, reimbursement decisions will also be based on the budgetary impact of a new drug on the annual national healthcare budget, especially the impact of a new drug on the drug budget. Therefore the authorities are requiring an assessment of the impact of a new drug on the annual drug budget. For the financial analysis, data on the following subjects will be required: descriptive epidemiology (data on incidence and prevalence). The patient group that is indicated for the drug and the anticipated substitution effects (i.e. the extent to which the existing treatment will be replaced); the use of the drug (e.g., posology, length of the treatment), the price of the drug; the expected market share plus the variables that would facilitate or slow down the drug sales and the total treatment costs. On the basis of the cost-effectiveness analysis and financial analysis an advice can then be drawn up as to whether the drug should be reimbursed; examples include The Netherlands, United Kingdom and Italy [23] (in Italy: Delibera CIPE 1 February 2001 [22]). For example, in Italy the recent Comitato Interministeriale per la Programmazione Economica (CIPE) provision (see above) specifies that the drug price is negotiated also on the basis of appropriate economic evaluations of the drug in its market and competition context. However, the budgetary impact analysis is not clearly defined and no formal guidelines exist, for example, it is not clear yet what is taken into account, drugs costs only or also other medical costs. In France the price is negotiated with the Comité Economique des Produits de Santé (CEPS) on the basis of the Amélioration du Service Médical Rendu (ASMR), but also incidence/prevalence of the disease, public health concern and drug budget in order to assess the budgetary impact (Décret du CEPS, 2000).[24] In Spain the budget impact is also considered during the price and reimbursing negotiation. The royal decree 271/1990, which regulates these processes, requires a forecast of the sales as an element for the final decision.

## Decentralisation and/or Prescription Guidelines

Financing prescription medicines in ambulatory care has been a central responsibility. The central authorities have recently begun often to shift the responsibility for development of prescription lists to the local authorities.

### *Decentralisation*

The responsibility for financing prescription medicines in ambulatory care is moving from the central to local level (e.g. Sweden, The Netherlands, United Kingdom). For example, the Dutch government considers giving the local health insurers the responsibility for the purchasing process for drugs, which means that they will directly bargain and negotiate with pharmaceutical companies. In the United Kingdom drug financing is no longer a central funding mechanism. Primary care trusts now have "devolved" budgets and can set their own drug budgets. In Sweden from the beginning of 1998, the responsibility for financing prescription medicines in ambulatory care was transferred from the National Social Insurance Board (RFV) to the county councils over a 3-year period. Under the new structures local consumption patterns and total cost of medicines are more transparent to regional administrators and prescribers. As a result, treatment should be more responsive to local needs and cost containment measures should become more effective. Although the local authorities are becoming more involved in influencing the prescribers to reduce volumes and switch to cheaper drugs (supply side), decisions on a reimbursement status level and reimbursement price in general is still the responsibility of the central authorities.

### *Drug Formularies*

There is a tendency to shift the development of lists to local authorities leading to local lists. Formularies have been used already by hospitals for in-patient drugs, but key actors in local healthcare (e.g. specialists, general practitioners, pharmacists and insurers) are now developing formularies for ambulatory drugs. In general drugs are added to the formulary when there are no similar drugs available. Drugs are substituted mainly when major clinical improvements and/or drug cost savings are expected. The use of expensive drugs is usually rationed by means of prescription guidelines or delivery restriction.

### *Prescription Restrictions*

While traditionally reimbursement decisions applied to the officially registered indication, which was usually a broad indication, authorities have recently been imposing restrictions on the claim made for the drug. These restrictions usually relate to follow a treatment protocol, to limit the prescribers or to limit the range of indications (United Kingdom, Germany and France). In Germany prescriptions are more and more restricted by the prescribing guidelines given by the federal committee of physicians and sickness funds (*Arzneimittelrichtlinien*). For example, these prescribing guidelines restrict drug therapy to the approved indications from clinical trials. The Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices) tends to approve drugs only for

indications which were part of the clinical trials. Specific to Italy is the issue of the so-called "Note CUF". This set of criteria for reimbursement of some selected active principles are specifically meant at reducing drug availability by identifying specific drug indications for which the drug is reimbursed. For example, ondansetron, granisetron and all 5-hydroxytryptamine type 3 antagonists are limited by Nota 57 to "prevention and treatment of nausea and vomiting in patients undergoing emetogenic chemotherapy", thus excluding all other potentially emetogenic indications (Decreto Ministero della Sanità, 22, December 2000).[25] The narrowing of the indication especially depends on the efficacy, but also the results of the above health economic analysis and financial analysis may be taken into consideration, which may suggest that within the registered range of indications a further limitation of the field of application must be made from the point of view of cost-effectiveness and budgetary impact. In the United Kingdom the National Institute for Clinical Excellence (NICE) is producing clinical guidelines incorporating technology and economic results, which include specific advice on targeting drugs, although it is not legally binding.

### *Financial Liability*

Prescription guidelines may be also used to audit physicians in terms of quality assurance to determine whether these guidelines are being correctly followed. Financial sanctions may be imposed if the recommendations are not followed. Prescription behaviour (or prescriptions patterns) of physicians may be also be directly linked with financial liability. We may distinguish between budget and envelop responsibility, budget being individual and envelop being collective responsibility. In Germany physicians have a target budget for pharmaceutical expenditure and exceeding the budget may result in financial sanctions. This target budget is calculated by applying the so-called *Richtgrößen* per member or *Richtgrößen* per retiree multiplied by the number of treated members or retirees per quarter. The sum per quarter times four results in the yearly target budget. The *Richtgrößen* take criteria such as physician specialty and status (member, retiree) into account. In addition, some products and some indications are considered as 'essential' and exempted from these guidelines (*Praxisbesonderheiten, Wirkstoffausnahmen*), which means that they are not included in the total prescribing bill of the physician. In Germany the Kassenärztliche Bundesvereinigung (National Association of Statutory Health Physicians) and the leading *Krankenkassen* associations form a federal framework, the Bundesausschuss Ärzte/Krankenkassen, which is responsible for setting up and updating prescribing guidelines (*Arzneimittelrichtlinien*) which should guide the physicians prescribing behaviour with regard to economic viability. The introduction of these guidelines means that the physicians are individually liable for mal-compliance with these guidelines, and that this mal-compliance may result in financial sanctions. Through the creation of these guidelines for certain drugs, the Bundesausschuss can exercise control over physicians prescribing. In general, there seems to be a mix between prescribing guidelines (*Arzneimittelrichtlinien*) and physicians individual budgets based on *Richtgrößen*. In the United Kingdom the individual prescriber is monitored by hospitals and

primary care trusts. "Overprescribing" is a problem only if budget constraints are broken. "Underprescribing" can become a political issue, for example, not following the National Institute for Clinical Excellence (NICE) guideline. Each trust must decide on prescribing policy: peer pressure is used to make individuals conform. In France groups of healthcare provider have their own envelop. In case they exceed the value of the "key-letter" might be reduced or its increase may be lower than inflation. In Spain during the past decade the position of primary care pharmacist has been created in several regions. This person is committed to develop recommended lists of save, effective and cheap drugs that are mailed to physicians, and also to advice prescribers. The pharmacist is also performing an ex post control of the type and budget impact of prescriptions and interviewing physicians to keep the expenditure in the established limits.

*Clinical Guidelines/Disease Management*

There is a trend to development and implementation of clinical guidelines, which usually are only prescription guidelines for physicians. These guidelines usually contain a recommendation for prescribing generic drugs and encourage the optimal use of drugs. Through the creation of guidelines for certain conditions, the healthcare authorities can exercise control over physicians' prescribing. These recommendations are seen as an extension of the non-reimbursement list. However, as described above, most of the prescribing guidelines enforce the use of drugs in the approved indication instead of the optimal use of drugs. These prescribing guidelines should strictly be distinguished from clinical guidelines or treatment recommendations from specialty associations or opinion leader. In France the Agence Nationale d' Accréditation et Evaluation en Santé (ANAES) issues clinical guidelines. The Réference Médical Opposable (RMO) is the official prescription guidelines for physicians (http://www.anaes.fr/). The French guidelines are determined by a scientific committee who consider general themes (prescription, diagnostic tests, biology, surgery) and/or different pathologies. Included in these guidelines are regulations indicating what the physicians should or should not prescribe following a certain diagnosis. All drugs were revisited by the Transparency Commission to check again their public health value. The development of guidelines is usually a decentralised local process with involvement of representatives of physicians and local insurers. Another example is the United Kingdom; one of the objectives of the National Health Service is to promote rational prescribing amongst general practitioners. The Department of Health pays for general practitioners' subscription to the Drug and Therapeutic Bulletin, an independent publication from the Consumers' Association, which contains evaluations of treatments and pharmaceutical products. However, this is not always in line with the official guideline, for example Relenza.

In summary, we have three recent parallel trends showing additional data requirements at a central level, more decentralisation of the responsibilities and decision-making process and prescription restrictions.

## DISCUSSION

The objective of this contribution was to present an overview of the current pricing and reimbursement environment for pharmaceuticals. Financing prescription medicines in ambulatory care has been a central responsibility, which was based on the traditional clinical trial outcomes (efficacy/safety parameters) used for registration. Although there is large variety between the various countries, there are three related trends: decentralisation of the healthcare decision-making process, prescription restrictions, and additional data requirements. Decentralisation and prescription limitations are not independent processes: Central authorities often shift the responsibility for development of prescription lists to the local authorities. At a central level the demand for cost-effectiveness and budgetary impact data is increasing, which has already resulted in formal reporting requirements in some countries (e.g. The Netherlands). Although the most evident impact of health economic studies is expected for central reimbursement audiences, evidence for the use of health economic studies by other audiences is expected to increase (e.g. patients, hospitals, insurers, formulary committees).[26] We already notice that this recent decentralisation process is adopting some of the economic criterion to better inform medical decisions on prescription. For instance, in Spain that probably is the most decentralised EU country, a new type of staff has been incorporated, in some regions, for primary care: the primary care pharmacists. These are committed to report on safety, efficacy and effectiveness of drugs as well as on the prices and rough cost-effectiveness values. Their recommendations are published and handed out to primary care physicians. A close follow-up of the prescriptions is carried out and an evaluation of the outcomes after the pharmacists reports are implemented. This approach to prescription has contained costs where applied, and it is observed by other Spanish regions with interest.

Below we address first general potential limitations of decentralisation and market mechanisms in the healthcare market and then focus on drug-specific issues.

We distinguish two potential limitations of decentralisation and market mechanisms in the healthcare market limitation. Firstly, in the current free market developments healthcare remains free at the point of delivery or is paid for indirectly through insurance premiums. Although there is an increase in co-payment, private health insurance is taken out for those services that are not provided free of charge. Hence customers will not shop around for the lowest price, and consequently there is still little pressure on producers to keep prices down. Instead, they may compete with one another by providing more attractive services, which may in fact lead to price increases rather than reductions; this is referred to as non-price competition. Secondly, the decentralisation of the healthcare decision-making process by broadening the role of health insurers from financial controllers to purchasers increased competition by increasing the number of buyers from one central body to more potential buyers. There are two recent developments which may inhibit the favourable consequences of this competition. Firstly, the increasing opportunities of administrative databases may lead to mergers between

purchasers: The use of databases allows management at a larger scale (e.g. larger number of patients) and may be beneficial only when benefitting from economies of scales. As a consequence mergers may lead to an oligopoly inhibiting the intended efficiency of purchaser-provider split.

On the other hand, the use of clinical guidelines outlining proper care will increase the homogeneity of the healthcare services, and therefore the products can compete on price and be comparable with one another. The increasing information technology (e.g. decision support tools for physicians) may facilitate implementation of more standardised treatment from prescription guidelines to more comprehensive disease management. The use of evidence based medicine will increase the consensus in clinical decision-making, increase the homogeneity and improve the assessment of appropriateness and quality. The increasing information technology (e.g. databases) may contribute to data collection of evidence-based data associated with different treatments.

In addition to the above general trends, there are other drug-specific developments, which need to be considered. While registration procedures for Europe are becoming centralised with central European decision-making by the European Agency for the Evaluation of Medicinal Products (EMEA), pricing and reimbursement decisions are still national affairs. Although opponents argue that drug registration and reimbursement must be kept strictly separate, there are signs that economic evaluation may become part of the registration dossier in the near future. Secondly, harmonisation of guidelines and further moves towards Europe-wide decisions on drug pricing and reimbursement are likely, given the increasing interdependency of European markets and regulatory authorities. A consequence is that the role of the central authorities will further decrease. On the one hand, centralisation of pricing/reimbursement decisions shifts to the European level, and, on the other hand, decentralisation of the other healthcare decisions (e.g. prescription restrictions, local formularies, clinical guidelines) shifts to the local level. However, the second part of this conclusion needs to be considered with prudence because the local sickness funds are financed differently in the various European countries, which will affect the option of the decentralised decision process in each country.

Another consideration is how cost-effectiveness data could be used at a central European level for pricing and reimbursement decisions. A key principle in health economics is that cost-effectiveness is based on the country-specific healthcare setting: local treatments patterns and local financing system determine the clinical and economic outcomes. Consequently a European cost-effectiveness ratio (e.g. 10,000 EURO per quality-adjusted life year) cannot be determined or is meaningless. Hence the use of health economic data at a central European level needs an in-depth examination. Some other important difficulties for European pricing and reimbursement are drug price differences between countries and parallel import. Furthermore, cost-effectiveness needs to be a particularised not only at the country level but also at the regional level. The reason is that once there are decentralised budgets, regions have political power to adopt decisions that consider more efficient given their health policy, costs (for instance, salaries are not equal across Spanish regions), patient management and epidemiological conditions. Trying to foresee the future, due to international trade and patent agreements, similar

pricing will become a more and more common policy throughout EU countries, but reimbursement policies - and hence adaptation of cost-effectiveness studies to inform on reimbursement and prescription practices -will have to be tailored more specifically to a difficult range of medical decisions.

Finally, we address the option of temporary reimbursement, which is being considered by Dutch and French authorities. Health economic evaluations consider efficacy and especially effectiveness. However, effectiveness data are usually not available at time of reimbursement procedures. Therefore temporary acceptance of an innovative drug to the reimbursement package might be considered. A conditional acceptance would permit initial decision-making on reimbursement based on the cost-effectiveness of the new drug derived from modelling data, followed by validation through subsequent prospective data collection. This would minimise the logistical and methodological concerns related to current policy. It would also reduce the concern of industry that health economic evaluation guidelines would delay product launch, shortening the period of useful patent life and the return on research and development investment. New drugs would be made available more quickly if prospectively data collection were not required prior to reimbursement. On the other hand, removal of a drug from a reimbursement package after additional prospective data were evaluated might have ethical concerns and lead to some social unrest. Regardless of the issue of temporary reimbursement there will surely be a need for collection of real-life data after the introduction of a new expensive drug, which accords perfectly with the concept of evidence-based medicine. The principle of evidence-based medicine is that clinical encounters should be supported by scientific conclusions based on data as much as possible.

Although the objective of this manuscript is to address the increasing central data requirements in the pricing and/or reimbursement decision-making process, the decision will be at least partially political. First, guidelines prescribe proper execution of health economic evaluations but not cut-off points for approving reimbursement. The decision about how much society will pay for increased effectiveness is political. Second, the weight of all data (e.g. data of the health economic evaluation and the financial impact analysis) in the decision-making process is currently not defined. Therefore we may conclude that the decision-making must become more transparent for a successful implementation of the new drug policies.

## References

1 Zweifel P; Felder S; Meiers M . Ageing of population and healthcare expenditure: a red herring? Health con 1999 Sep;8(6):485-96

2 J. Rovira, M. Craig. Consolidation, co-payment and selective financing: Spain remodels its healthcare system: Spectrum healthcare delivery (Decision resources); 50:1-9: 1993

3 Decreto Legislativo - 19/06/1999 n. 229 "Norme per la razionalizzazione del Servizio sanitario nazionale, a norma dell'articolo 1 della L. 30 novembre 1998, n. 419

4 Loi Juppé: PMSI. Manuel des DRG, version 1. Paris : Ministère des Affaires Sociales et de l'Emploi, 1992. Bulletin officiel n° : 92-2 bis

5 Repubblica Italiana. Decreto Legislativo 30 dicembre 1992, n. 502. Riordino della disciplina in materia sanitaria, a norma dell'articolo 1 della legge 23 ottobre 1992, n. 421. Supplemento ordinario alla Gazzetta Ufficiale n. 305 del 30 dicembre 1992)
6 Repubblica Italiana. Decreto Legislativo 7 dicembre 1993, n. 517. Modificazioni al Decreto Legislativo 30 dicembre 1992, n. 502. Riordino della disciplina in materia sanitaria, a norma dell'articolo 1 della legge 23 ottobre 1992, n. 421. Supplemento ordinario alla Gazzetta Ufficiale n. 293 del 15 dicembre 1993
7 http://www.bmgesundheit.de/engl/healthcare.htm
8 §§ 140a-h SGB V. (social security law in Germany)
9 OECD Healthcare data 99. A comparative analysis of 29 countries. Paris 1999
10 Berto. P, Lopatriello S. "Per chi non suona la campana delle linee guida" Sole24ore Sanità, 4(18):20.21, 2001
11 Nuijten MJC, "Pharmacoeconomics in European Decision-Making".Value in Health. 1999, 2(5): 319-322
12 G. Baum. Healthcare and medical devices. Challenges for bringing new technlogies to the German market Hepac1(2): 131-133; 2000
13 A.Guhl. Pricing and reimbursement systems in Europe. Hepac. 8-11; 2000
14 F. Pelen. Reimbursement and pricing of drugs in France. Hepac. 19-23.2000
15 Serra, Lucioni, Pharmacoeconomics and the pharmaceutical marketplace: an Italian view, Cronache Farmaceutiche 2:79-84, marzo-aprile 2000)
16 Antoñanzas F, Juárez C, Portillo F (2001). El mercado único de medicamentos en Europa. Farmacia Profesional 15(1):14-18; 2000
17 Juárez C, Portillo F, Antoñanzas (2000). Los precios de referencia: un cambio cultural. Farmacia Profesional: 14(9): 15-24
18 § 33a SGB V. (social security law in Germany)
19 SOLE24Ore Sanità, n.3 23-29 January 2001
20 Canadian coordinating office for health technology assessment (1997). Guidelines for economic evaluation of pharmaceuticals. 2nd edition. Ottawa
21 Drummond MF et al (1999). Current trends in the use of pharmacoeconomics and outcomes research in Europe. Value in Health 2:323-332
22 Delibera CIPE 01/02/2001 n° 3/2001 Individuazione dei criteri per la contrattazione del prezzo dei farmaci. Gazzetta Ufficiale 28/03/2001
23 Wiseman V, Mooney G (1998). Burden of illness estimates for priority setting: a debate revised. Health Policy 43:243-51
24 Décret du CEPS. 2000
25 Decreto Ministero della Sanità 22/12/2000. Revisione delle "note" riportate nel provvedimento 30/12/93 di riclassificazione dei medicinali e successive modificazioni. Supplemento Ordinario Gazzetta Ufficiale 10/01/2001
26 J.M. Graaf von der Schulenburg (ed). "The influence of economic evaluation studies on healthcare decision-making" IOS Press, Berlin, 2000

Chapter 3

# Assessment of Clinical Guidelines for Continuation Treatment in Major Depression from a Health Economic Perspective


## SUMMARY

OBJECTIVE: The primary objective of this study was to assess the appropriateness of the existing Dutch clinical guidelines for the treatment of depression from a health-economic perspective. The existing guidelines recommend continuation treatment for a period up to 9 months. METHODS: The assessment was based on a Markov model using decision-analytic techniques. For this analysis we defined six mutually exclusive states defined by the existence of depression and type of treatment. The outcomes for the model were defined as: time without depression (TWD), quality-adjusted life years (QALYs), direct medical costs, and cost of lost productivity. The primary perspective of the study was that of the third-party payer, while the secondary perspective was that of the society in 1999. The probabilities of clinical events and therapeutic choices as well as the utilities were based on published literature. The medical resource use related to each state was abstracted from published literature and expert opinion. The associated 1999 unit costs of the used medical resources were derived from official Dutch tariff lists of allowable reimbursements. Indirect costs in this model were based on lost productivity only. RESULTS: The results of the primary analysis showed that the use of the guidelines is not cost-effective. Continuation treatment for a period of 9 months increases the total direct medical costs (NLG 1276 vs. NLG 474), decreases the costs resulting from lost productivity (NLG 304 vs. NLG 909), increases total costs (NLG 1580 vs. NLG 1383) and increases TWD (96.9% vs. 86.4%). However, continuation treatment does not change the utility outcomes (0.60 vs. 0.61 QALYs) for both treatment strategies. Hence continuation treatment is not cost-effective from either a third-party payer perspective or a societal perspective. A scenario analysis showed that an extension of the continuation treatment to maintenance treatment might result in a favorable cost-effectiveness outcome of the treatment guideline. CONCLUSION: In conclusion, based on the assumptions used in the model, the current Dutch treatment guidelines for depression are only appropriate from a health-economic perspective if continuation treatment is extended to maintenance treatment.

# INTRODUCTION

## Epidemiology

Depression is a common debilitating illness that exerts a large social, medical and economic impact on society. Epidemiological studies show a prevalence varying between 1% and 4% in women and between 2% and 3% in men.[1 2 3] The highest prevalence of the disease appears to be in adults between the ages of 18 and 44.[4 5] A study by Ormel yields a prevalence of 6.7% for persons older than 18 years in the Netherlands, which is the study country in this manuscript.[6]

## Economic Impact of Depression

Although there is some variation due to the diversity of healthcare systems, as a rule of thumb most of the developed countries spend about 10% of their national income on healthcare.[7] About 10% of total healthcare expenditures is spent on mental disorders treatment.[8] Several important findings have emerged with regard to mental disorders. Most recent studies point out that the economic burden of depression consists of healthcare expenditures (direct costs) and costs related to loss of productivity.[9] [10] There are a number of other factors that may explain, why depression is such a significant problem of public health, leading to long-term morbidity and costs: 1) depression has a high risk of recurrence and chronicity; 2) the fact that depression is mainly prevalent among young adults (between 18 and 44 years) means that the patients involved represent the active population, increasing the costs to society through productivity losses due to work days lost [3]; 3) severe depression can lead to frequent and lengthy hospitalisations, which have an enormous impact in economic terms [11] and; 4) major depression is associated with significant social morbidity, decreases in health status and well being, and impacts dramatically on the ability of patients to function normally in the work place.[12 13 14]

## Treatment of Depression

The drugs most often used in current treatment patterns for depression include the tricyclic antidepressants (TCAs) and the newer selective serotonin reuptake inhibitors (SSRIs), which are associated with fewer side effects.[15,16] Depressive episodes usually respond well to short-term treatment: studies have shown that approximately 65% to 80% of depressive patients can be treated effectively with anti-depressant drugs, while placebo response rates have often been in the range of 20% to 40%.[17] The treatment of the acutely depressed patient is referred to as short-term or episodic treatment. Although patients usually responded well to short term treatment, relapse and recurrence are common problems in clinical practice.[18 19 20] A new episode may be either a relapse or recurrence, which are now recognised as two distinct events, the former being a re-emergence of the original depressive episode and the latter the emergence of a new episode, unrelated to the original.[18] A relapse has been loosely defined as a depressive event that takes place within 4 to 6 of the original event, whereas a recurrence occurs after this period.[15] Patients who achieve a stable asymptomatic state with full restoration of psychosocial function and who are symptom free for six months or more

following an episode of depression are said to have recovered.[21] [22] A review of studies with a follow-up of more than 1 year has shown consistently high rates of recurrence: all patients experienced at least 2 episodes in Perris' and Angst's study; respectively 96% in Carlson's and 78% in Lehman's.[23] [24] [25] [26] Studies indicate that 50% to 85% of patients with a diagnosis of major depression will experience at least one more episode of depression in their lifetime,[27] [28] and individuals with a history of two or more depressive episodes have almost a 90% chance of experiencing a third.[29]

Several studies have indicated that long-term treatment of depression may be beneficial for patients with major depression. [30] [31] [32] The findings of those studies resulted in recommendations for prolongation of the antidepressant medication. Accordingly, the WHO (WHO Organisation Mental Health Collaborating Centers, 1989) [33], the Agency for Health Care Policy and Research (AHCPR) [34] in the US and numerous other consensus groups (Paykel and Priest [35], Montgomery et al. 1993 [36], American Psychiatric Association [37], Commission de Transparence 1994 [38], Swedish Board of Health [39]) have established depression treatment guidelines recommending at least 4 to 9 months of antidepressant therapy beyond initial symptom resolution for an initial depressive episode, with longer antidepressant treatment for subsequent episodes. The clinical treatment guidelines for Dutch GPs ["NHG-standaard" by the Dutch association of GP's ("Nederlands Huisartsgenootschap"[1])] recommend a period of continuation treatment up to 9 months after response to medication.[40]

As mentioned above, a distinction should be made between relapse and recurrence. The decision whether to continue treatment after a positive response and for how long should therefore take into consideration both the risk of relapse and the risk of recurrence. Hence, prolongation treatment may be considered to consist of two components, one aimed at preventing a relapse, and one aimed at preventing recurrence. Those components have been labelled continuation and maintenance treatment, respectively, and for the rest of this article we will use this terminology.

Although the clinical efficacy of the SSRIs is comparable to that of the TCAs [41], their major advantage in clinical practice is the lower incidence of side effects and increased patient compliance.[15] [42] [43] The use of antidepressive medication in The Netherlands has been studied using a local database of GPs over the period 1994 to 1995. This study showed that nearly all patients using SSRIs, received the minimal effective dosage versus only 61% of the patients using TCAs.[44] Hence the increased patient tolerability of SSRIs clearly makes them ideal agents for the long-term maintenance phase of treatment, which has been confirmed by a number of studies.[20] [32] [45] [46] In those studies, responders to an acute treatment with an SSRI were randomised to continuation of the treatment or switched to placebo, and relapse and/or recurrence were assessed during the continuation and/or maintenance period.

[1] NHG-standaard: guideline by the Dutch Association of GPs ("Nederlands Huisartsgenootschap").

## Assessment of Clinical Guidelines from a Health-Economic Perspective

A growing number of organisations have issued clinical guidelines for prolongation of antidepressant medication following response to an initial antidepressant medication.[34 35 37 36 38 39 40] Clinical guidelines outline the proper care of medical conditions and performance of clinical procedures. The intended goal of guidelines is to reduce inappropriate care and to improve patient outcomes. In addition, these guidelines are potential tools for reducing the costs of healthcare, and for improving medical education.[47] While the principal goal of clinical guidelines is to define what is best for the individual patient, an assessment of those guidelines from a health economic perspective is performed from the broader society perspective. This approach consists of relating cumulative measures of costs over time with cumulative measures of clinical benefit. Economic measures such as direct medical costs and cost of lost productivity must be considered, while costs play a minor role in the decision-making for clinical guidelines.[48 49] Another difference is that the majority of clinical guidelines have been based on efficacy outcomes of randomised clinical trials - relapse and recurrence rates for continuation and maintenance treatment, respectively. Ideally, health-economic studies should not be based on efficacy, but on effectiveness in terms of morbidity and mortality: [e.g. quality-adjusted life years (QALY)]. If no effectiveness data are available, then appropriate modelling techniques may be used to model efficacy data into what is expected in practice (effectiveness). All assumptions behind such modelling techniques should be evaluated with the help of sensitivity analysis.

Economic evaluation is a valuable tool allowing strategic choices to be made in the context of healthcare policy. This approach appears to be particularly relevant for chronic disorders that are common and make use of many healthcare resources. A health-economic assessment of clinical guidelines is especially relevant when the disease leads to a high burden on society in terms of costs and social morbidity (quality of life). Consequently the necessity for a health-economic assessment of clinical guidelines may be based on following criteria: 1) prevalence or incidence and 2) cost-of-illness. According to both criteria and considering the epidemiological and economic data presented on depression, it is obvious that a health-economic assessment of clinical guidelines for treatment in depression is justified.

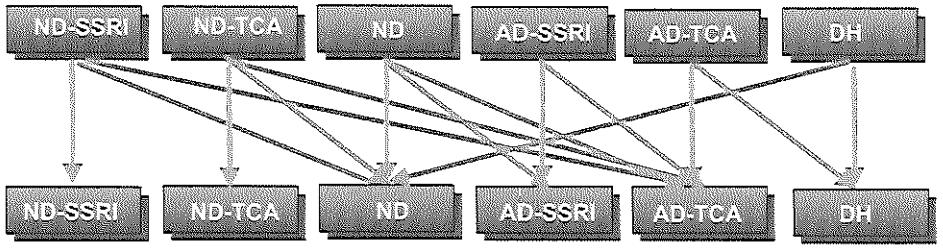
The primary objective of this cost-effectiveness study was to determine the appropriateness of the Dutch "NHG-standard" recommending continuation treatment from a health economic perspective; a secondary objective was to determine the optimal period of prolongation of an antidepressant medication.

The underlying hypothesis of this study was that a prolongation of an antidepressant treatment will lead to a favourable health-economic outcome compared with short-term treatment only. A reduction of relapse and/or recurrence will result in: 1) an increase of Time without depression (TWD), in quality of life, and consequently, in QALYs; and 2) a reduction of total costs (direct medical costs and costs of lost productivity), which will lead to a favourable incremental cost-effectiveness outcome.

## DESIGN AND METHODS

Decision analytic techniques were used to specify the potential health-economic benefits of the clinical guidelines. This application of decision analytic techniques to the development of guidelines is relatively new, and thus there is limited experience.[50 51] In that sense, this study is exploratory in nature. Its main aim is to show how health-economic analysis can be used alongside clinical guideline development. This cost-effectiveness analysis compared prolongation of antidepressant medication versus no prolongation after response to an initial treatment with an antidepressant in the Dutch healthcare setting. The Netherlands was chosen for the purpose of illustration, because this country was the first European country to have developed pharmacoeconomic research guidelines.[52] Hence, this offered the opportunity of conducting a health-economic analysis according to these standardised research guidelines. The primary perspective of the study was that of the third party payer, while the secondary perspective was that of the society in 1999.

**Figure 1. Markov states and possible transition used in the model.**




### Markov Models

Health-economic modelling studies are based on decision analysis, which allows for the analysis of both clinical and economic consequences of medical actions and attitudes.[53] From treatment algorithms a model can be constructed which considers the timings of actions and their consequences over time. In effect, a model shows the consequences and complications of different therapeutic interventions, and it should correspond as much as possible, to the real life situation of the disease. Models may take the form of simple decision analytic trees or they may be very complex Markov models. Whereas decision tree models are appropriate for acute episodes, Markov models are the first choice for pharmacoeconomic analysis of chronic diseases, like depression. The data being used in a modelling study can be categorised into transition probabilities (e.g. response, relapse and recurrence), healthcare utilisation (e.g. drugs, consultations), prices and tariffs, and utilities.[54] The data may come from a variety of sources and is subject to varying degrees of uncertainty. Data sources for the variables being used in a model may be clinical trials, literature (e.g. meta-analysis), databases, medical records, and official tariff lists of allowable reimbursements. These data sources will yield, for each variable, a fixed input value and a range. The standard analysis will be based on the fixed

input value for all variables. The range of each variable is then used to determine the sensitivity of the outcome to the analysis, when the input value is varied within its range.

## Description of the Markov Model

A Markov process represents a convenient way of modelling the long-term evolution of health states over successive time periods.[55] A Markov process model describes several discrete states of health in which a person can be at time $t=n$ as well as the states of health into which the person may move at time $t=n+1$. The progression from $t=n$ to $t=n+1$ is called a cycle. Probabilities are associated with each change from one health state to another; these are termed transition probabilities (P). Each transition probability is a function of the health and of the time spent in each health state. The Markov model used in this study is shown in Figure 1; each Markov state is assigned a utility, and the overall contribution of this utility depend on the length of time spent in the state. A utility can be a clinical as well as an economic parameter.

For this analysis we defined six mutually exclusive states based on the existence of depression and the type of treatment applied:

- no depression and no prolongation of treatment (ND)
- no depression and prolongation with SSRI (ND-SSRI)
- no depression and prolongation with TCA (ND-TCA)
- depression in ambulatory setting treated with SSRI (AD-SSRI)
- depression in ambulatory setting treated with TCA (AD-TCA)
- treatment failure in severe depression requiring hospitalisation (DH).

Patient progression through these states was divided into cycles of 2 months. A 2-month cycle time was chosen, because this interval closely approximates the time of the sequential therapeutic stages: e.g. the initial 2-month trial of anti-depressant medication is followed by a continuation treatment for another 6 months in case the patient responded positively.

## Description of the Transitions

In the model all patients begin in the state identified above as "no depression", either with or without medication, and during each subsequent cycle they either remain in the same state or progress to another state depending on the allowed transitions and their associated probabilities. The transition probabilities between the states are based on relapse and recurrence rates for responders with or without prolonged treatment and subsequent response rates to treatment.

The model is based on the following assumptions:

<u>No Prolongation and prolongation of treatment:</u>

- The low patient compliance with TCAs makes them less than ideal agents for prolongation treatment. Therefore the health-economic assessment was based on prolongation of first-line treatment of depression with an SSRI.

- A patient will be hospitalised after three subsequent treatment failures with antidepressants, including increase of dosage of antidepressants.
- An increase of dosage of antidepressant medication (SSRI or TCA) was not been included in the model. The assumption is that an increase of the dosage of an antidepressant will occur within the cycle time of 2 months.
- We have excluded the possibility of suicide in the model. Although patients with depression may have a significant risk for suicide over the long-term, previous studies showed that the incidence of suicide over a period of 18 months was negligible, being less than 1%.[56] For the same reason, suicide was not included in a previous publication by our group nor was it included in the study by Jonsson, as the cost associated with suicide is minimal due to the low incidence and the relatively low impact on cost.[57 58]

No prolongation of treatment:

- Patients not currently on prolongation treatment will be treated again with an SSRI after a relapse or recurrence. In the event of no response to this treatment, the patient will be switched to a TCA.
- Patients who do not respond to a TCA will be switched to another type (class) of TCA.
- Responders to a SSRI or a TCA will discontinue this treatment after the short-term period.

Prolongation of treatment:

- Patients receiving prolonged treatment with a SSRI, who experience a relapse or recurrence, will be switched to a TCA. The probability of a response will correspond to the probability of response to a first-line antidepressant treatment.
- Responders to a SSRI will continue this treatment as prolongation treatment after the short-term period.
- Responders to a TCA will, in the base case analysis, terminate this treatment as prolongation treatment after the short-term period because of the low level of compliance with TCAs; in a scenario analysis, they will continue this treatment according to the Dutch clinical guidelines.
- Patients, who do not respond to a TCA, will be switched to another type (class) of TCA.
- The patients who experience a relapse or recurrence after termination of an antidepressant treatment, consist of two subpopulations: 1) patients who would also have experienced a relapse or recurrence on prolonged treatment; and 2) patients who would have not experienced a relapse or recurrence on prolonged treatment. Both groups of patients will be treated with the initial antidepressant that resulted in the response. The first group of patients will not respond to this medication; the second group of patients will have a probability of response corresponding to an initial antidepressant treatment.
- Patients receiving prolonged treatment with the first TCA who experience a relapse or recurrence will be switched to another TCA. The probability of a response will correspond to the probability

of response to a first-line antidepressant treatment. Patients receiving prolonged treatment with the second class of TCA who experience a relapse or recurrence will be hospitalised.

**Table I. Clinical data and sources.**

| Clinical data | Transition probability | Reference |
|---|---|---|
| **SSRI** | | |
| Response to 1-line treatment | 0.600 | Nuijten, 1995 |
| Response to 2-line treatment | 0.300 | Nuijten, 1995 |
| Relapse during continuation treatment | 0.016 | Montgomery, Dunbar 1993 |
| | 0.040 | Montgomery, Rasmussen 1993 |
| Mean | 0.028 | |
| Recurrence during maintenance treatment | 0.036 | Montgomery, Dunbar 1993 |
| | 0.049 | Montgomery 1988 |
| Mean | 0.042 | |
| **TCA** | | |
| Relapse during continuation treatment | 0.092 | Nuijten 1998 |
| Recurrence during maintenance treatment | 0.018 | Frank 1993 |
| Relapse after termination of medication | 0.112 | Montgomery 1993 |
| | 0.100 | Montgomery, Rasmussen 1993 |
| Mean | 0.106 | |
| Recurrence after termination | 0.084 | Montgomery, Dunbar 1993 |
| | 0.131 | Montgomery 1988 |
| Mean | 0.108 | |
| **Response to hospitalization** | | |
| TCA | 0.581 | Geretsegger 1995 |
| SSRI | 0.643 | Geretsegger 1995 |
| TCA | 0.612 | Gershon 1981 |
| SSRI | 0.627 | Clerc 1994 |
| Mean | 0.616 | |

## Clinical and Economic Outcomes

The clinical outcomes for the model were defined as Time without depression (TWD) and QALYs; the economic outcomes were the direct medical costs and costs resulting from lost productivity.

The primary perspective of the study was that of the third-party payer, while the secondary perspective was that of the society in 1999. From a scientific point of view, the society perspective should include all indirect costs, including the costs of time spent. However, the Dutch pharmacoeconomic research guidelines state that "the focus is on production losses, when including indirect costs".[52] The costs were determined from the perspective of the Dutch society in 1999. Utilities and costs were discounted at 4% according to the Dutch pharmacoeconomic research guidelines.

The cost assessment was based on the assignment of costs to defined health states associated with a cycle time of 2 months. The direct medical cost of each health state was determined by the healthcare utilisation associated with each health state, including medication, consultations, procedures and

number of days of hospitalization. The cost resulting from lost productivity was based on working days lost over a period of 2 months.

**Table II. Utilities for the health states.**

| Health state | Utility |
|---|---|
| Treatment depression: SSRI | 0.70 |
| Treatment depression TCA | 0.64 |
| In remission SSRI, prolongation | 0.80 |
| In remission TCA, prolongation | 0.72 |
| In remission, off drug therapy | 0.86 |
| Severe depression | 0.30 |

## Data Sources

Different types of data can be distinguished in modelling studies:

- probabilities of clinical events, generally are beyond the control of the physician (e.g. probability of relapse, recurrence and response);
- probability of therapeutic choices (e.g. therapy after a relapse or during continuation treatment or maintenance treatment).
- utilities of different Markov health states; and
- direct medical costs, consisting of the units of healthcare utilisation and their prices/tariffs. Costing of direct medical costs involves estimating the number of units of healthcare utilisation and multiplying this by the corresponding prices/tariffs. Costing for lost productivity involves estimating the number of working days lost and the opportunity cost to society.

The probability of clinical events and therapeutic events was based on published literature (Table I). The response to first- and second-line antidepressant medication was derived from a study by Nuijten.[58] The percentage of patients relapsing during continuation treatment with an SSRI was based on the average percentages of patients relapsing in two randomised double-blind clinical trials by Montgomery and Dunbar [32] and Montgomery and Rasmussen.[45] The percentage of patients not on continuation treatment who relapsed was based on the percentage of patients receiving placebo who relapsed in those clinical studies.

The percentage of recurrence during a maintenance treatment with an SSRI was based on the percentage of recurrence in the two clinical studies published by Montgomery in 1988 [46] and 1993.[32] The percentage of recurrence among patients not on maintenance treatment was based on the percentage of recurrence among placebo-treated patients in these clinical studies. The relapse rate for patients treated with TCAs was derived from a literature review by Nuijten[56] which was based on data from various studies. [59 60 61]

**Table III. Cost data and sources.**

| Costs | | | |
|---|---|---|---|
| **Direct costs (NLG)** | **Costs *** | **Range** | **Reference** |
| SSRIs | 2.60 | | Taxe Nov. 1999 |
| TCA- amytriptiline | 0.25 | | Taxe Nov. 1999 |
| TCA- anafranil, clomipramine | 0.94 | | |
| Consultation | 40.00 | | COTG '99 |
| Per diem | 289.00 | 280-295 | COTG '99 |
| **Cost of lost productivity** | | | **Reference** |
| Working days lost (days) (over two months) | 5.2 | | Broadhead |
| GDP per capita (NLG) | 42729 | | CBS, 1999 |
| GDP per capita per working day | 244 | | CBS, 1999 |

*: drugs: daily treatment costs.

The Actuarial Method was used to adjust all relapse and recurrence percentages to transition probabilities corresponding to a cycle time of 2 months.[62] It was remarkable that the transition probabilities for relapse and recurrence for patients on placebo were similar in both studies.

The utilities for the different health states were derived from a study by Revicki (Table II).[63] All utilities were determined for ambulatory treatment. We assumed that the utility for the health state corresponding to hospitalisation was equal to the utility for severe depression, as an assumption was validated by expert opinion (DA Revicki, MEDTAP International, Bethesda, Washington, USA).

The medical resource use for each state was abstracted from published literature (Table III).[64] The treatment patterns and resource utilisation were adjusted for 1999 values and validated by expert opinion. The associated 1999 unit costs of medical resources used (e.g., professional services, healthcare services) were derived from official Dutch tariff lists of allowable reimbursements (Taxe; November 1999) and other sources [College Tarieven Gezondheidszorg (CTG, 1999), central tariffs in healthcare for consultations and procedures, Utrecht,]. The daily treatment cost (DTC) of a treatment with an SSRI was an average of the DTCs of the four most prescribed SSRIs in The Netherlands (IMS). Dosages of drugs were derived directly from the official Dutch Formulary (Farmacotherapeutisch Kompas) (Table III). The base-case analysis was based on the cheapest TCA, amytriptiline.

Costs due to lost productivity in this model were based on the value of working days lost, according to the gross domestic product (GDP) per-capita figure. Because lost productivity is frequently excluded from pharmacoeconomic analyses, it was difficult to find sources directly applicable to The Netherlands for this study, and so we used data from the study by Broadhead et al.[14] We assumed that lost productivity for depressed patients is consistent, whether they are treated with an SSRI or a TCA.

When the results of this study were adjusted to our cycle time of 2 months, depression was found to increase the number working days lost to 5.8 days over a period of 2 months. The cost of lost productivity resulting from hospitalisation was calculated by assuming 100% productivity losses. Assessment of lost productivity was based on the friction method: the Dutch pharmaco-research guidelines recommend the use of the Friction Cost Method over the more traditional Human Capital Approach (HCA).[52] The use of an appropriate method of assessment is especially relevant for chronic disease like depression, where the use of HCA would have lead to an overestimation of the cost of lost productivity.

Sensitivity analyses were performed on the relevant variables in our model. Sensitivity analyses were performed on relapse and recurrence probabilities under prolongation treatment because both figures were derived from a limited number of clinical trials with limited external validity. The response rate to hospitalisation was based on the average of the response rates reported in studies that assessed the efficacy of intramural administration of antidepressants.[65 66 67] The lower and upper ranges were based on the minimum and maximum values that we found in the literature (Table I). In addition, we performed sensitivity analyses on the DTC for a TCA and the *per diem*. The DTC for TCA was varied between minimum and maximum values; the *per diem* was varied between plus and minus 10%.

Sensitivity analysis is based on the modification of the basic clinical and economic estimates of input variables over a plausible range of values to evaluate the effect on study results of alternative assumptions for uncertain variables.[68] In contrast, scenario analysis is based on the modification of the underlying therapeutic strategies of the model. In this study, a scenario analysis was performed on the prolongation of a treatment with a TCA assuming full compliance. Another scenario analysis was performed on an extension of the continuation treatment as recommended by the guidelines to an extra year of maintenance treatment. Finally a scenario analysis was performed on compliance during prolonged antidepressant treatment with an SSRI.

**Table IV. Results: discontinuation of TCAs.**

| Outcome | Costs (NLG) | | | Time without depression | | QALY |
|---|---|---|---|---|---|---|
| | Direct Costs | Cost of lost productivity | Total Costs | % | Months | Years |
| **Base case: 9 months** | | | | | | |
| - Prolongation | 1276 | 304 | 1580 | 96.9% | 8.72 | 0.60 |
| - Termination | 474 | 909 | 1383 | 86.4% | 7.78 | 0.61 |
| - Difference | 803 | -606 | 197 | 10.4% | 0.94 | -0.01 |
| - ICR (Cost/month) | 854 | | 210 | | | |
| - ICR (Cost/QALY) | Termination Dominant | | Termination Dominant | | | |
| **Scenario: 21months*** | | | | | | |
| - Prolongation | 3831 | 1512 | 5343 | 94.9% | 17.09 | 1.31 |
| - Termination | 4681 | 4282 | 8963 | 82.9% | 14.92 | 1.31 |
| - Difference | -850 | -2770 | -3620 | 12.0% | 2.17 | 0.00 |
| - ICR (Cost/month) | Prolongation Dominant | | Prolongation Dominant | | | |
| - ICR (Cost/QALY) | Prolongation Dominant | | Prolongation Dominant | | | |

*: discounting at 4%.

**Table V. Cost distribution of direct medical costs: discontinuation of TCAs.**

| Outcome | Costs (NLG) | | | |
|---|---|---|---|---|
| | Drugs | Consultations | Hospitalisation | Total |
| **Base case: 9 months** | | | | |
| - Prolongation | 736 | 309 | 232 | 1276 |
| - Termination | 75 | 53 | 346 | 474 |
| - Difference | 661 | 255 | -114 | 803 |
| **Scenario: 21 months*** | | | | |
| - Prolongation | 1514 | 642 | 1675 | 3831 |
| - Termination | 178 | 129 | 4374 | 4681 |
| - Difference | 1336 | 513 | -2698 | -850 |

## RESULTS

The Markov model was built using the decision-tree software (DATA™ 3.5 by TreeAge Software, Williamstown, MA, 1996). The model was run to giver an expected value for TWD, QALYs and cost per patient. Sensitivity analyses were performed on the main clinical and economic input variables in order to ascertain the robustness of the conclusions. The key results of the analysis are shown in Table IV, and show that the use of guidelines (prolongation treatment) increases the total direct medical costs substantially from NLG 474 to NLG 1276, while it reduces substantially the cost of lost productivity from NLG 909 to NLG 304.

**Table VI: Results: continuation of TCAs.**

| Outcome | Costs (NLG) | | | Time without depression | | QALY |
|---|---|---|---|---|---|---|
| | Direct Costs | Cost of lost productivity | Total Costs | % | Months | Years |
| **Base case: 9 months** | | | | | | |
| - Prolongation | 1278 | 302 | 1580 | 96.9% | 8.72 | 0.59 |
| - Termination | 474 | 909 | 1383 | 86.4% | 7.78 | 0.61 |
| - Difference | 804 | -607 | 197 | 10.5% | 0.94 | -0.02 |
| - ICR (Cost/month) | 852 | | 209 | | | |
| - ICR (Cost/QALY) | Termination Dominant | | Termination Dominant | | | |
| **Scenario: 21 months*** | | | | | | |
| - Prolongation | 3801 | 1415 | 5216 | 95.2% | 17.13 | 1.27 |
| - Termination | 4681 | 4282 | 8963 | 82.9% | 14.92 | 1.31 |
| - Difference | -880 | -2867 | -3747 | 12.3% | 2.22 | -0.04 |
| - ICR (Cost/month) | Prolongation Dominant | | Prolongation Dominant | | | |
| - ICR (Cost/QALY) | Termination Dominant | | Termination Dominant | | | |

*: discounting at 4%.

**Table VII: Cost distribution of direct medical costs: continuation of TCAs.**

| Outcome | Costs (NLG) | | | |
|---|---|---|---|---|
| | Drugs | Consultations | Hospitalisation | Total |
| **Base case: 9 months** | | | | |
| - Prolongation | 736 | 309 | 232 | 1276 |
| - Termination | 75 | 53 | 346 | 474 |
| - Difference | 662 | 256 | -114 | 803 |
| **Scenario: 21 months*** | | | | |
| - Prolongation | 1539 | 663 | 1599 | 3801 |
| - Termination | 178 | 129 | 4374 | 4681 |
| - Difference | 1361 | 534 | -2775 | -880 |

**Table VIII. Sensitivity analysis: base case analysis (continuation treatment).**

| Sensitivity Analysis | Range | ICR- Medical costs | ICR- Total costs |
|---|---|---|---|
| Relapse | 0.0160 | Termination-dominant | Termination-dominant |
| | 0.04 | Termination-dominant | Termination-dominant |
| Response hospitalisation | 0.518 | Termination-dominant | Termination-dominant |
| | 0.643 | Termination-dominant | Termination-dominant |
| Per diem (NLG) | 260 | Termination-dominant | Termination-dominant |
| | 318 | Termination-dominant | Termination-dominant |
| DTC-TCA (NLG) | 0.94 | Termination-dominant | Termination-dominant |
| No discounting | 0% | Termination-dominant | Termination-dominant |

*: discounting at 4%.

**Table IX. Sensitivity analysis: scenario analysis (maintenance treatment).**

| Sensitivity Analysis | Range | ICR- Medical costs | ICR- Total costs |
|---|---|---|---|
| Relapse | 0.0160 | Prolongation-dominant | Prolongation-dominant |
| | 0.04 | Termination-dominant | Termination-dominant |
| Recurrence | 0.036 | Prolongation-dominant | Prolongation-dominant |
| | 0.04 | Prolongation-dominant | Prolongation-dominant |
| Response hospitalisation | 0.518 | Prolongation-dominant | Prolongation-dominant |
| | 0.643 | Prolongation-dominant | Prolongation-dominant |
| Per diem (NLG) | 260 | Prolongation-dominant | Prolongation-dominant |
| | 318 | Prolongation-dominant | Prolongation-dominant |
| DTC-TCA (NLG) | 0.94 | Prolongation-dominant | Prolongation-dominant |
| No discounting | 0% | Prolongation-dominant | Prolongation-dominant |

*: discounting at 4%.

The use of the guideline increases TWD substantially from 86.4% to 96.9%, while it does not increase the number of QALYs (0.60 versus 0.61). This first incremental cost-effectiveness analysis (cost/QALY) shows that the use of the guidelines is not cost-effective either from the third-party payer (medical costs), or from the societal perspective; termination of treatment is dominant versus prolongation, because prolongation increases the costs without an increase in QALYs. The results of the incremental cost-effectiveness analysis on TWD are expressed in extra costs per extra month without symptoms of depression; the results vary from NLG 854 to NLG 474 for the third-party payer and society perspective, respectively.

Prolonged treatment until 21 months (continuation treatment followed by maintenance treatment) reduces the medical costs from NLG 4,681 to NLG 3,831when compared with termination of treatment. Maintenance treatment increases TWD from 82.9% to 94.9%, while it does not increase the number of QALYs (1.31). Hence, maintenance treatment is cost-effective from both the third-party payer and societal perspective because prolongation decreases the costs without a decrease in QALYs.

Table V provides a breakdown of direct costs by components: drugs, consultations and hospitalisations. As would be expected, the costs of drug therapy and associated extra consultations were higher for prolongation treatment, whereas the cost of hospitalisation is lower due to fewer relapses. Only an extension to a maintenance treatment is cost saving, since the higher drug and consultation costs are offset by the reduction in hospitalisation costs.

**Table X. Compliance with SSRIs: base case of 9 months.**

| Outcome | Costs (NLG) | | | Time without depression | | QALY |
|---|---|---|---|---|---|---|
| | Direct Costs | Cost of lost productivity | Total Costs | % | Months | Years |
| 20% | | | | | | |
| - Prolongation | 634 | 788 | 1422 | 88.5% | 7.96 | 0.61 |
| - Termination | 474 | 909 | 1383 | 86.4% | 7.78 | 0.61 |
| - Difference | 161 | -121 | 39 | 2.1% | 0.19 | 0.00 |
| - ICR (Cost/month) | 854 | | 210 | | | |
| - ICR (Cost/QALY) | Termination Dominant | | Termination Dominant | | | |
| 80%* | | | | | | |
| - Prolongation | 1116 | 425 | 1541 | 94.8% | 8.53 | 0.60 |
| - Termination | 474 | 909 | 1383 | 86.4% | 7.78 | 0.61 |
| - Difference | 642 | -484 | 158 | 8.4% | 0.75 | -0.01 |
| - ICR (Cost/month) | 854 | | 210 | | | |
| - ICR (Cost/QALY) | Termination Dominant | | Termination Dominant | | | |

*: discounting at 4%.

Table VI and Table VII show the results of the scenario analysis, which assumes that treatment with TCAs will be continued after a response. The analysis shows that the 9-month results are still similar to the base-case analysis. Prolonged treatment until 21 months (continuation treatment followed by maintenance treatment) reduces the total costs slightly more than the base case analysis (NLG 3,747 versus NLG 3,620). Table VII shows that cost savings result from a further reduction in hospitalisation costs (NLG 2,775 versus NLG 2,698). As would be expected, continuation of TCA treatment will lead to higher costs for drug therapy and associated extra consultations compared with discontinuation, but continuation will also reduce relapses and recurrences in responders to TCAs, which will lead to further reductions in hospitalisations and lost productivity and an increase in TWD. On the other hand, continuation with TCAs will not lead to a higher number of QALYs: Table II shows that the unfavourable side effect profile of TCAs is associated with lower QALYs for patients using TCAs either for acute treatment or prolongation treatment.

Table VIII and Table IX show the results of the sensitivity analyses for respectively the base-case analysis (9-month treatment) and scenario analysis (maintenance treatment), respectively. The sensitivity analyses for the base-case analysis show that varying the variables within their range does not change the outcome of our model, whereas sensitivity analyses for maintenance treatment show that the outcome of our model is not sensitive to most of the variables other than the relapse rate during prolongation of treatment. When the analysis is based on the maximum values, termination of treatment becomes more favourable than maintenance treatment.

Tables X and XI show the relationship between the expected outcomes and compliance with SSRI treatment when TCAs are discontinued after a response (base-case analysis) and compliance is varied between 20% and 80%. These results show that the lower compliance substantially reduces potential cost-savings of prolongation treatment, especially in the maintenance-treatment scenario. Although compliance does not have an impact on QALYs, lower compliance substantially reduces TWD.

## DISCUSSION

This study examined the cost-effectiveness of continuation treatment with SSRIs compared to no preventive treatment in order to assess the appropriateness of clinical guidelines for continuation treatment. In our model, the use of the NHG guideline recommending a continuation treatment for 9 months was not cost-effective, when compared to no prolongation, in responders to an initial treatment. A scenario analysis revealed that only continuation of treatment followed by maintenance treatment with SSRIs was favourable, both in terms of cost and effect, in the treatment of patients with depressive disorders in a Dutch setting. This scenario analysis showed that an extension of the continuation treatment will decrease the medical costs and the costs of lost productivity compared with no prolonged treatment, while it will increase the TWD. Sensitivity analyses showed the robustness of our conclusion. Although continuation of TCA treatment will lead to cost-savings, the use of TCAs will have a negative impact on QALYs due to their unfavourable side-effect profile of

TCAs. In addition, published data show that compliance with TCAs is low in real practice, which justifies our base-case analysis in which only continuation with SSRIs is considered.

**Table XI. Compliance with SSRIs: maintenance of 12 months.**

| Outcome | Costs (NLG) | | | Time without depression | | QALY |
|---|---|---|---|---|---|---|
| | Direct Costs | Cost of lost productivity | Total Costs | % | Months | Years |
| **20%** | | | | | | |
| - Prolongation | 3670 | 1227 | 4789 | 84.0% | 17.64 | 1.27 |
| - Termination | 4681 | 4282 | 8963 | 82.9% | 14.92 | 1.31 |
| - Difference | -1011 | -3055 | -4175 | 15.1% | 2.72 | -0.04 |
| - ICR (Cost/month) | Prolongation Dominant | | Prolongation Dominant | | | |
| - ICR (Cost/QALY) | Prolongation Dominant | | Prolongation Dominant | | | |
| **80%*** | | | | | | |
| - Prolongation | 3277 | 664 | 3506 | 91.1% | 19.13 | 1.27 |
| - Termination | 4681 | 4282 | 8963 | 82.9% | 14.92 | 1.31 |
| - Difference | -1404 | -3618 | -5457 | 23.3% | 4.21 | -0.04 |
| - ICR (Cost/month) | Prolongation Dominant | | Prolongation Dominant | | | |
| - ICR (Cost/QALY) | Prolongation Dominant | | Prolongation Dominant | | | |

*: discounting at 4%.

Although the use of the guideline increases TWD substantially, it does not increase the number of QALYs. The values for utilities (Table II) explain this outcome: the use of an antidepressant reduces the QALYs of health states in patients without depression because of side-effects. Although prolongation treatment reduces relapses and recurrences, and consequently overcomes the reduction in QALYs resulting from depression, this gain is apparently not sufficient to offset the lower utilities in treatment in responders.

The results of any modelling exercise need to be treated with some degree of caution. It is worthwhile to keep in mind that the purpose of pharmacoeconomic studies is to inform or aid in decision-making and not to usurp the relationship between patient and physician or to interfere with the physician's treatment decision based on his assessment of the patient's clinical status. This model adheres to this tenet. Decision-analytic techniques, upon which our Markov model is based, have several weaknesses. Among them is the fact that the results of our literature review do not necessarily represent real clinical practice, since much of the literature examined was based on data from clinical trials. Data from clinical trials does not necessarily have a high degree of external validity because the results are

often contingent upon protocol adherence, which is not representative of treatment modalities outside the trial setting.

We analysed lost productivity in our model. The use of lost productivity in pharmacoeconomic studies remains controversial. The Australian Guidelines did not see them as central to pharmacoeconomic analyses and recommended that they be reported only in rare instances.[69] However, more recently the Canadian guidelines seem willing to accept estimates of lost productivity from the societal perspective as long as the source of the cost data is documented and a sensitivity analysis has been performed.[70] The recent Dutch pharmaco-economic research guideline also recommends the inclusion of costs of lost productivity associated with working days lost [52], but based on the so-called friction method: the period over which the production losses are calculated is limited to the friction period, which is the period between the start of absence and the actual moment of replacement. This period is currently estimated at 3 months on the average. We used the Friction Cost Method in order to avoid any overestimation of the impact of costs of lost productivity caused by prolonged treatment. We only included indirect costs due to lost productivity. The inclusion of all indirect costs would make results less credible. There is no consensus yet, among scientists on how to deal with time spent on unpaid activities and time spent by caregiver among scientists. In addition, the healthcare decision maker tends to be only interested in medical costs, while general policy makers may also be interested in indirect costs caused by productivity losses. Therefore, the current analysis can be considered an underestimation of the true benefits to society of prolongation of antidepressant medication.

The results of this study confirm that the use of maintenance therapy with SSRIs in the treatment of depressive disorders is justified. A number of studies, which vary from modelling studies to retrospective database analysis, all have shown that the higher drug costs of a maintenance treatment with SSRIs are offset by savings in other medical costs, mainly hospitalisation. In a study by Hatziandreu, a model was constructed that followed two cohorts of 35-year-old women at high risk for recurrent depression over their lifetimes.[27] The authors conclude that a long term maintenance treatment with sertraline appears to be a clinically and economically justified choice for patients at high risk of recurrent depression, which corresponds with findings of studies by Boyer [71] and Kamlet.[29] In addition Nuijten,[58] using a Markov process analysis, showed that total costs associated with a 1-year period of maintenance therapy for depression in Germany were 33% lower with the SSRI citalopram than with TCAs, and again, relapse rates had a large impact on the model outcomes. However, most of those studies compared initial treatment with SSRI and initial treatment with TCA, assuming continuation of SSRI and discontinuation of TCA. Hence, those studies compared both type of drug and duration of treatment. In our study we compared prolongation and termination in responders to an initial treatment with SSRIs. This approach allowed us to evaluate the cost-effectiveness of a prolonged treatment alone, which is a sounder scientific design.

The increased patient tolerability of SSRIs clearly makes them ideal agents for the long-term maintenance phase of treatment.[20] Nevertheless, a study by Quik and Kleintjens [72] showed that Dutch patients in the GP setting do not receive insufficiently long antidepressant therapy, including with SSRIs. Another Dutch study by De Waal, which showed that all patients using SSRIs, received the minimal effective dosage, may indicate that the insufficient treatment period may not be a result of the unfavourable safety profile of SSRIs, but of other reasons. For example, physicians may not always communicate the importance of continuing antidepressant treatment to symptom-free patients, who may consider themselves as completely recovered. Hence, practice guidelines based on research may not be applicable to real-world practice. Important feasibility issues such as constraints faced by practitioners, the concerns of patients, and the limitations of the healthcare system need to be considered. In a supplementary analysis we showed that compliance is a main feasibility constraint from a health-economic perspective. The validity of practice guidelines ranks as the most critical attribute, even though it may be the hardest to define and measure. Conceptually, practice guidelines are valid if they lead to the health and cost outcomes projected for them while other parameters remain equal. A prospective collection of health economic data and utilities in a naturalistic setting may validate the projected health-economic outcomes of our model and identify and assess the critical real-life issues such as non-compliance.

## CONCLUSION

The results of this cost-effectiveness study show that clinical guidelines for continuation treatment are not justified from a health-economic perspective. A scenario analysis showed that extension of the continuation treatment to a maintenance treatment might substantially increase the cost-effectiveness of the treatment guideline. The increased costs for antidepressant medication are offset by the increased effectiveness and the reduction of total costs. Sensitivity analyses confirmed the robustness of these findings.

### References


1. Brown GW, Craig TKJ, Harris TO. Depression: distress or disease? Some epidemiological considerations. Br. J. Psychiatry. 1985; 147: 612-622
2. Regier DA, Myers JK, Kramer M, et al. The NIMH epidemiological catchment area program. Arch Gen Psychiatry. 1984; 41 (10): 934-941
3. Myers JK, Weissman MM, Tischler GL, Holzer III CE, Leaf PJ. Six-month prevalence of psychiatric disorders three communities. Arch. Gen. Psychiatry 1984; 41: 959-967
4. Burvill PW. Recent progress in the epidemiology of major depression. Epidemiol Rev 1995; 17: 21-31.
5. Mintz J, Mintz LI, Arrud MJ et al. Treatment of depression and the functional capacity to work. Arch Gen Psych 1992; 49: 761-768
6. Ormel J. Psychiatrische ziekten: depressie. In: Grobbee DE en Hofman A (red.). Epidemiologie van ziekten in Nederland. Utrecht:Wetenschappelijke uitgeverij Bunge, 1989
7. Souêtre E, Martin P. Clinical economics and psychiatry. In: The Use of Fluoxetine in Clinical Practice 1991, International Congress and Symposium Series, The Royal Society of Medicine Publications No. 183, 67-73

8. McGuire TG. Measuring the economic costs of schizophrenia. Schizophrenia Bull. 1991; 17 (3): 375-388
9. McCombs JS, Nichol MB, Stimmel GL. The cost of antidepressant drug therapy failure: a study of antidepressant use patterns in a medical population. J. Clin. Psychiatry. 1990; 51(6) (suppl.): 60-69
10. Greenberg PE, Stiglin LE, Findelstein SN, Berndt ER. The economic burden of depression. J. of Clin. Psychiatry. 1993; 54: 405-418
11. Stoudemire A, Frank R, Hedemark N, Blazer D, Kamlet M. The economic burden of depression. General hospital psychiatry. 1986; 8: 387-394
12. Wells KB, Stewart A, Hays RD et al. The functioning and well being of depressed patients. Results from the medical outcomes study. J Amer Med Assoc 1989; 262: 914-919
13. Souêtre E, Hardens M, Lozet H. Indirect costs of anxiety and depressive disorders: the hidden part of the iceberg. Med Interface 1995; September: 131-135
14. Broadhead WE, Blazer DG, George LK Tse CK. Depression, disability days mad days lost from work in a prospective epidemiologic survey. J Amer Med Assoc 1990; 264: 2524-2528
15. Finley, PR. Selective serotonin reuptake inhibitors. Pharmacologic profiles and potential therapeutic distinctions. Ann Pharmacother 1994; 28: 1359-1369
16. Guze BH. Selective serotonin reuptake inhibitors. Assessment for formulary inclusion. Pharmacoeconomics 1996; 9: 430-422
17. Baldessarini RJ. Current status of antidepressants: Clinical pharmacology and therapy. J Clin Psychiatry 1989; 50: 117-126
18. Thase ME. Relapse and recurrence in unipolar major depression: Short-term and long-term approaches. J Clin Psychiatry 1990; 51(6.Suppl): 51-57
19. Maj M.Veltro F, Pirozzi R et al. Pattern of recurrence of illness after recovery from an episode of major depression: A prospective study. Am J Psychiatry 1992; 149: 795-800
20. Montgomery SA. Long-term treatment of depression. Brit J Psychiat 1994; 165 (Suppl 26): 31-36
21. Agency for Health Care Policy and Research. Depression in primary care: Vol II treatment of major depression. AHCPR, Washington D.C., 1993
22. WHO Mental Health Collaborating Centres. Pharmacotherapy of depressive disorders: a consensus statement. J of Affective Disorders 1989; 17: 197-198
23. Perris C. The course of depressive psychoses. Acta Psychiatr. Scand.. 1968; 44: 238-248
24. Angst J. The epidemiology of depression. Psychopharmacology 1992; 106 Suppl.: 71-74
25. Lehmann HE, Fenton FR, Deutch M, Feldman S. An 11-year follow-up of 110 depressed patients. Acta Psychiatry. Scand. 1988; 78: 57-65
26. Carlson GA, Kotin JL, Davenport XB, et al. Follow-up of 53 bipolar manic-depressive patients. Br. J. Psychiatry. 1990;124: 134-139
27. Hatziandreu EJ, Brown RE, Revicki DA et al. Cost utility of maintenance treatment of recurrent depression with sertraline versus episodic treatment with dothiepin. Pharmacoeconomics 1994; 5: 249-264
28. National Institutes of Mental Health/National Institutes of Health. Mood disorders: Pharmacologic prevention of recurrences. Consensus development panel. Am J Psychiatry 1985; 142: 469-476
29. Kamlet MS, Paul N, Greenhouse J et al. Cost utility analysis of maintenance treatment for recurrent depression. Controlled Clinical Trials 1995; 16: 17-40
30. Frank E, Kupfer DJ, Wagner EF, et al. Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: contributing factors. Arch. Gen. Psychiatry. 1991; 48: 1053-1059
31. Klerman GL, Weissman MM. The course, morbidity and costs of depression. Arch. Gen. Psychiatry. 1992; 49: 831-834
32. Montgomery SA, Dunbar GC. Paroxetine is better than placebo in relapse in relapse prevention and the prophylaxis of recurrent depression. International Clinical Psychopharmacology. 1993; 8: 189 - 195
33. WHO Organisation Mental Health Collaborating Centers (1989, Pharmacotherapy of depressive disorders: WHO consensus statement. J Affect Disord 17: 197-198
34. Agency for Health Care Policy and Research (AHCPR). Treatment of major depression: Clinical Practice Guidelines, No 5, Vol 2. Rockville, MD: US Department of Health and Human Services, AHCPR Publication no. 93-0550
35. Paykel ES, Priest RG. Recognition and management of depression in general practise. Br Med J 1992. 205. 1992: 1198-1202
36. British Association for Psychopharmacology. Guidelines for treating depressive illness with antidepressants. J. Psychopharmacol 1993: 7; 19-23
37. American Psychiatric Association. Practice guidelines for major depressive disorders in adults. Am J Psychiatry. 1993: 150 (4, Suppl.); 1-26

38. Commission de Transparence. Fiche de Transparence Antidepresseurs. Ministere de la Sante Publique de l'environment. Brussels. 1994
39. Swedish Board of Health . Lakemedelsverket Argang. 1995. 6. nr. 5
40. Van Marwijk HWJ, Grundmeijer HGLM, Brueren MM, et al. NHG Standaard Depressie. Huisarts Wet 1994; 37(11): 482-90
41. Dunbar GC, Cohn JB, Fabre LF et al. A comparison of paroxetine, imipramine and placebo in depressed out-patients. Brit J Psychiat 1991; 159: 394-398
42. Andrews JM, Nemeroff CB. Contemporary management of depression. Amer J Med 1994; 97 (Suppl 6A): 24S-32S
43. Anderson IM, Tomenson BM. Treatment discontinuation with serotonin reuptake inhibitors compared with tricyclic antidepressants: A meta-analysis. BMJ 1995; 310: 1433-1438
44. De Waal M, Stolk J, Van Marwijk J. et al. Voorschrijven van antidepressiva in the huisartsenpraktijk. Nederlands Tijdschrift voor Geneeskunde. 1996; 140: 140-3
45. Montgomery SA, Rasmussen JGC and Tanghoi P. A 24 week study of 20 mg citalopram, 40 mg citalopram and placebo in the prevention of relapse of major depression. International Clinical Psychopharmacology. 1993; 9: 181-188
46. Montgomery SA, Dufour H, Brion S. The prophylactic effect of fluoxetine in Unipolar Depression. Br. J. Psychiatry. 1988; 153 (suppl. 3): 69 - 76
47. Woolf SH (1992). Practice Guidelines: a new reality in medicine. II. Methods of developing guidelines. Arch Intern. Med.; 152: 946-952
48. Eddy DM (1990). What do we do about costs? JAMA; 264: 1161-1165
49. Eisenberg JM (1989). Clinical Economics: a guide to the economic analyses of clinical practices. JAMA; 262: 2879-2886
50. Eddy DM (1990). Practice policies: guidelines for methods. JAMA; 263: 1839-1841
51. Eddy DM (1990). Guidelines for policy statements: the explicit approach. JAMA; 263: 2239, 2240, 2243
52. National Health Insurance Council. Report on Pharmacoeconomic Research Guidelines. Amstelveen, The Netherlands: National Health Insurance Council, 1999
53. Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-65
54. Nuijten M.J.C.. The selection of data sources for use in modelling studies. PharmacoEconomics, March 1998; 3(3): 305-316
55. Beck JR, Pauker SG. The Markov process in medical prognosis. Med Decis Making 1983; 3: 419-458
56. Nuijten MJC, Hadjadjeba L, et al. Cost-effectiveness of fluvoxamine in the treatment of recurrent depression in France. PharmacoEconomics, October 1998, (4): 433-455
57. Jonsson B, Bebbington PE. What price depression? The cost of depression and the cost-effectiveness pf pharmacological treatment. Brit J Psychiat 1994; 164: 665-673
58. Nuijten MJC, Hardens M and Souetre E.A Markov Process Analysis Comparing the Cost Effectiveness of Maintenance Therapy with Citalopram versus Standard Therapy in Major Depression. Pharmacoeconomics. 1995; 8 (2): 159-168
59. Loo H, Dufour GH, Guelfi JD et al. Long-term use of trianeptine in 38 depressed patients. Br J Psychiatry. 15 1992; 169: S61-5
60. Stein MK, Rickels K, Weise C. Maintenance therapy with amitriptyline: a controlled trial. Am J Psychiatry 1990; 137: 370-1
61. Mindham R, Howland C, Shepard M. An evaluation of continuation therapy with tricyclic antidepressants in depressive illness. Psychol Med 1973; 3: 5-17
62. Miller DK, Hofman SM. Determining transition probabilities. Confusion and suggestions. Med Dec Making 1994; 1: 52-8
63. Revicki D, Wood M. Patient-assigned health state utilities for depression-related outcomes. Journal of Affective Disorders. 48 (1998). 25-36
64. Nuijten M, Van der Kuy A. Hogere geneesmiddelenkosten ruimschoots gecompenseerd. Pharmaceutisch Weekblad. 1999. 4: 124-129
65. Geretsegger C et al. Multicenter double blind study of paroxetine and amytriptiline in elderly depressed inpatients. Psychopharmacology. 1995; 119: 277-81
66. Gershon A et al. Evaluation of tradozone in the treatment of endogenous depression: results of a multi-center double blind study. J. Clin. Psychopharmacol. 1981; 1: S39-44
67. Clerc GE et al. A double-blind comparison of venlafaxine and fluoxetine in patients hospitalised for major depression and melancholia. Int. Clin. Psychopharmacology. 1994; 9: 139-43

68. Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health care Technology: A Report on Principles. Ann Intern Med. 1995; 122: 61-70
69. Commonwealth Department of Health, Housing and Human Services. Guidelines for the pharmaceutical industry on preparation of submissions to the pharmaceutical benefits advisory committee, Canberra" Australian Government Publishing Service, 1992
70. Torrance, GW, Blaker D, Detsky A et al. Canadian guidelines for economic evaluation of pharmaceuticals. PharmacoEconomics 1996; 9:535-559
71. Boyer W, Feighner JP. The financial implications of starting treatment with a selective serotonine reuptake inhibitor or tricyclic antidepressant drug-naive depressed patients. In Health Economics of Depression (Eds B Jonsson and J Rosenbaum), 65-76, Wiley, Chichester, 1993: 65-76
72. Quik R.F.P., Kleintjens H.Th., Hylan T.R., Prescriptiepatronen van antidepressiva in de huisartspraktijk. Pharmaceutisch Weekblad. 1998; 133 (29): 1078-1085

Chapter 4

# The Selection of Data Sources for Use in Modelling Studies


## SUMMARY

Economic analysis has become increasingly important in healthcare in general, and particularly with respect to pharmaceuticals. Therefore, it is vital that the methods used in such evaluations are carefully scrutinised and refined. However, guidelines contain only a limited number of recommendations for the use of secondary data in modelling studies. In this manuscript, the selection of data sources in modelling studies will be addressed. The objectives of this manuscript are as follows: (i) to present a general strategy on how to determine the appropriateness of a data source for a model and (ii to present recommendations on a transparent reporting format for the selection of data sources.

## INTRODUCTION

Economic analysis has become increasingly important in healthcare in general, and particularly with respect to pharmaceuticals. In future, decisions over the reimbursement of pharmaceuticals are likely to require information on cost-effectiveness in addition to registration requirements. Because economic evaluation will play a fundamental role in healthcare decision-making, it is vital that the methods used in such evaluations are carefully scrutinised and refined. If economic analysis is to play an important and useful role in the allocation of scarce healthcare resources, then such analyses must be performed properly and with care.

In order to increase the scientific quality and integrity of pharmacoeconomic studies, national guidelines have been developed by various countries. The intent of the guidelines is to provide guidance to those conducting studies, and those using them, by laying out a general "state of the art" regarding methods, and by providing specific methodological advice on many matters. Australia was the first country to develop and implement guidelines for the economic evaluation of pharmaceuticals. Draft guidelines were released in 1990, and these were revised and implemented in 1992 and 1995.[1,2] Canada became the second country to release national guidelines in 1994.[3]

While these guidelines extensively describe issues like comparator, perspective and type of analysis, they contain only a limited number of recommendations for data management. The Ontario [4] and Australian guidelines are the most prescriptive on the issue of data capture. The Ontario guidelines recommend the use of databases and meta-analysis of randomised clinical trials. The Australian guidelines recommend the use of results from randomised clinical trials, supplemented by additional information. The Canadian guidelines do not specifically address this issue, but appear to favour data from clinical trials.

In practice, it is not always possible to derive information from scientifically sound prospective studies. In the absence of clinical trial data, there are no generally accepted methods describing how to collect data and validate data for local practice patterns. Because the guidelines were written mainly for studies based on primary data derived from Phase III trials, requirements for data management, especially data sources, may indeed seem redundant.

Although the revised Australian guidelines recognise the need to model an economic evaluation under certain circumstances, the principal focus is still on the quality and reporting of randomised clinical trials, without recommendations for data management in modelling studies.[2] Data management is a straightforward process, based on usual clinical trial procedures: the data on effectiveness and costs will be collected alongside a clinical trial. These guidelines may also be applied to naturalistic studies, which are also based on primary data. However, there are no recommendations on how to collect data when an analysis necessitates the use of secondary data. Even a pharmacoeconomic analysis based on a standard Phase III trial may have to rely on secondary data in order to incorporate patients with treatment failures, who in a clinical trial are usually lost to follow up, but may be the main cost drivers

from an economic viewpoint.

In most clinical trials, economic data are not collected alongside the study. Even when they are, the data may need to be projected to populations, time periods, or settings that were not observed in the clinical study. In these cases, decision-analytic models may provide some of the missing information. Decision analysis is an explicit, quantitative, prescriptive approach to medical decision-making and allows both clinical and economic consequences of medical actions and attitudes to be analysed under conditions of uncertainty.[5] The model resulting from the decision analysis must correspond, as much as possible, to the real life situation of the disease and should reflect actual treatment patterns with input values (probabilities and items of healthcare utilization) deviating as little as possible from population values. Models may take the form of simple decision-analytic trees or they may be very complex Markov models of underlying disease processes and treatments.

The subject of this manuscript is the selection of data sources, and issues relating to this subject. Those issues are:

- classification of data being used in modelling studies
- the various types of data sources
- a general strategy on how to determine the appropriateness of a data source for a model
- recommendations.

## THE SELECTION OF DATA SOURCES

Data management can be categorised into the selection of data sources and methods of data processing. The reliability of the estimates depends on the choice of the data sources (selection criteria, external validity) and data processing (calculation of mean or median, Mantel-Haenszel method, actuarial method). Hence, there may be a potential bias, both in the choice of the data sources and the methods of data processing.

### Classification of Data

The data may come from a variety of sources and are subject to varying degrees of uncertainty. Data sources for the variables being used in a model may be clinical trials, literature (e.g. meta-analysis), databases, medical records, Delphi panels and/or official tariff lists for healthcare use. These data sources will yield, for each variable, a fixed input value and a range. The standard analysis will be based on the fixed input value for all variables. The range of each variable is then used to determine the sensitivity of the outcome to the analysis, when the input value is varied within its range. The costs are based on units of healthcare utilisation and tariffs/prices. Effectiveness is based on clinical outcomes (mortality), the time without symptoms/toxicity (TWiST concept) or utilities.

Three different types of data can be distinguished in modelling studies:

1. *Transition probabilities.* These are out of the control of the physician and are derived from clinical measures. Examples of transition probabilities are response to a treatment, relapse

after a previous response, and mortality.

2. *Treatment patterns.* In addition to standard therapy being the first choice of therapy for the total cohort entering the model, the following decisions on the choice of therapy may occur in the model: (i) therapy after a response (termination of treatment or continuation of treatment) (ii) therapy after a treatment failure in response to the initial treatment; and (iii) therapy after a treatment failure, for example, a change to second-choice therapy.
3. *Costing information.* Costing involves estimating the units of healthcare utilisation used and their prices/tariffs (product of unit and price).

## Assessment of Data Sources

This section discusses the advantages and disadvantages of the different data sources.

### *Clinical Trials*

Because a clinical trial is designed for registration purposes, the objective of a clinical trial is to show a statistical difference between the study drug and the comparator, which requires a high internal validity in order reduce the impact of confounding variables. Therefore, clinical trials have strict inclusion and exclusion criteria, and treatments are protocol driven, which may lead to the overestimation of units of healthcare used. Consequently, clinical trials have a low external validity, which means both clinical and economic outcomes may not be typical and do not correspond to usual practice. Hence it should always be considered that, because of its restriction on external validity, the outcomes may not be representative of a target patient population who are going to use the drug.

In addition, a clinical trial is not designed for pharmacoeconomic purposes and the units of healthcare that are used and collected may not be complete when using these data retrospectively. Additional limitations for the collection of units of healthcare utilisation are the fact that: (i) since clinical trials are only driven to prove differences in efficacy, it is difficult to measure significant differences in healthcare resources used; (ii) the units of healthcare used are country specific and the duration of follow-up of clinical trials is too limited for pharmacoeconomic purposes. Pharmacoeconomic analyses are usually concerned with examining the discounted costs and benefits beyond the duration of patient follow-up, for example, when estimating the cost per life-year saved.

### *Meta-Analysis*

A meta-analysis may be used, in particular, in the determination of transition probabilities (based on efficacy measures for initial treatment), when the comparator in a pharmacoeconomic analysis is not the same as the comparator in a clinical trial or when clinical trials consist only of 1 arm (e.g. oncological trials). For treatment failures, a meta-analysis may be the preferred first-choice option. If we assume that clinical outcomes are not country-specific, a meta-analysis need not be limited to trials running in the study country.

The risk of meta-analyses is that they might be subject to publication and inclusion bias, particularly if criteria are chosen in order to produce the intended results. Although various tests for bias can be carried out, there is no guarantee that a meta-analysis has not been performed to give more scientific weight to what amounts to selective reporting of clinical trials. This problem is further complicated by the fact that there may be selective publication of the result of clinical trials. A meta-analysis based on a small number of trials should be interpreted with caution.[6] When there is too much heterogeneity, because of to differences in design or study population, studies may not be combined. The use of meta-analysis depends on the existence of relevant literature on treatment failure after the initial treatment.

Finally, because a meta-analysis is only a form of quantitive review of individual trials, it has the same limitations mentioned in the section on "Clinical Trials".

### *Databases*

A database may not be a suitable source for guiding decisions in healthcare since so much of the data it contains are not scientifically valid. Although databases may contain a lot of detailed information on both clinical and economic outcomes, the format of this information has to fit the structure of the model. However, the majority of the existing databases have not been developed for economic evaluations.

The 2 different databases that can be distinguished are claim databases and clinical outcomes databases. In the case of claim databases, the objective is to collect, for administrative purposes, all data on healthcare resources used; in clinical databases, the objective is to measure clinical outcomes for medical or scientific purposes.

1) Claim Databases

Claim databases have the advantages of a large number of patients and a high external validity. However, these databases usually have little or no information on clinical outcomes. Clinical tests may have been recorded for reasons of reimbursement, while the outcomes of these tests are usually not registered. The amount of healthcare resources used may be recorded, but often it is limited to the outpatient setting only, whereas hospitalisation is usually the main cost driver. In contrast, hospital management information systems only contain information limited to inpatient care.

Studies based on databases that are designed to compare different treatments, are rarely sufficient to draw definitive conclusions about the relative effectiveness of treatments because, like other non-randomised studies, the estimates are susceptible to unrecognised underlying biases that can render the conclusions invalid. Because of the retrospective nature of the review, patients may have been nonrandomly assigned to a treatment and it is possible that some differences observed resulted from differences between patient groups. A related limitation is the absence of a control group to which people were randomly assigned (i.e. those who did not receive the centre's programme). In the

absence of a control group, we cannot rule out the possibility that the observed improvements were the result of chance or statistical regression to the mean.

Another potential source of bias is the "clinic patient bias"[7], which is a variant of prevalence-incidence variance.[8] This bias occurs when only patients who had at least 2 measurement points (e.g. clinical visits) during the follow-up period are included in the analysis. This type of database abstraction will miss any cases in which, for instance, adverse drug reactions resulted in the death of the patient or discontinuation of therapy after only a single visit.

The patients in databases usually are labelled according to existing medical classification systems, such as the International Classification of Disease-9$^{th}$ revision-Clinical Modification (ICD-9-IM) or diagnosis-related groups (DRGs) for inpatient care. Because of financial incentives, patients may be labelled with a different ICD-9 code or more severe DRGs. Another problem associated with diagnosis is the lack of consensus on defining criteria for a pathology, and the overlap of symptoms. For example in psychiatric disorders patients with depression and patients with generalised anxiety disorder frequently have overlapping characteristics, which makes it difficult for a physician to clinically distinguish among these disorders. Hence, a subpopulation of patients with ICD-9-IM code for depression may include patients with generalised anxiety disorder because of inaccurate diagnosis.

The population in a claims database may not be representative of the overall population and may contain a biased sample of patients; for example, in the US, Medicare databases will only consist of the elderly, while Medicaid databases will only contain patients with a low sociodemographic status.

2) Clinical Databases

Clinical databases contain mainly clinical outcomes, but usually no, or incomplete, information on healthcare use. The external validity may be limited because of bias resulting from the site(s) of the data collection, the speciality of the physicians and the primary objective of the database implementation, which may vary between a prospective clinical study and evaluation of a healthcare programme. Prospective studies have the same limitations as clinical trials (section on "Clinical Trials"), while for a healthcare programme, the limited choice of therapy will reduce the external validity.

The clinical outcomes should have enough external validity in order to be used as an effectiveness measure or to be extrapolated to effectiveness measures Direct relevant outcomes (TWIST, QUALYs or Quality of Life) for pharmacoeconomic studies are usually not included. The sample size in a clinical database is generally lower than in a claims database. Hence, the measurement of statistical significance will be more difficult.

### *Medical Records*

There are a number of shortcomings and limitations to the use of medical records: when switching between inpatient and outpatient care, different medical records have to be used. For instance, medical records of general practitioners (GPs) can be used until a patient is hospitalised. Although after

hospitalisation, a report of the main events will be sent to the GP, medical hospital records still need to be accessed. While databases yield information that can be directly entered into statistical data processing software, information from medical records cannot be entered directly and, as a result, data entry can be time consuming. Access to medical records may be complicated in many countries because of legislation associated with privacy protection.

### *Delphi Panel Techniques*

The use of expert opinion is appropriate in situations in which there is little or no published material in a particular area, or in which the results of a thorough literature review or meta-analysis are considered unreliable,[9] conflicting,[10] or insufficient to cover the requirements of a study. Delphi panels operate in stages or rounds, in an effort to obtain a convergence of opinion in a particular area. Conventional Delphi panels must go through at least 2 rounds, and typically go through three to four stages of data collection.[11]

The second distinct feature is that in the classic Delphi there is no face-to-face contact between the respondents, although there may be face-to-face contact between the investigator and the respondents. That is, all responses by panellists are anonymous to other panellists. The rationale for this is that respondents will not be intimidated or dominated by other respondents. This is a particular concern in the healthcare profession,[12] where there is a hierarchical structure in healthcare delivery and seniority dominates most healthcare activities. Along with the questionnaire, experts may be provided with a literature review or seed algorithm in an effort to frame their responses.

There are several methodological weaknesses inherent in the process described above. As the techniques seek to quantify qualitative information, compromises have been made between scientific rigor and the need for structured information in certain areas. The main areas of weakness are as follows.

1. A problem related to the iterative process in Delphi method is the potential for experts to drop out following the first round. This behaviour may lead to a response bias if the attrition rate is substantial.
2. Expert opinion can be used as a structured way to include expert opinion about cost-effectiveness of treatments. However, there is variability in the way this is carried out and the results can be strongly influenced by unduly favourable or negative estimates being included.
3. In the pharmacoeconomic studies reviewed there were no explicit criteria for the selection of experts for participation in the studies. [13] Several of the studies did not fully justify the selection of experts or even indicate the number of panel members. In section on "Recommendations", a number of selection criteria for the members of a Delphi panel will be presented.
4. The impact of outliers is relatively high when using 10 members, which is considered as an appropriate sample size.[13] The statistical error resulting from the low sample size may lead to a

high standard deviation and the impact of an outlier on the fixed input value may be significant, leading to a large difference with the real population mean.

5. The use of a "physician-expert panel" to estimate resource use, while common, carries the risk that respondents may give inaccurate estimates or specify the resources required for ideal care, rather than that provided in practice. Physicians may adjust estimations based on other estimations, because they do not want be outliers. In addition, physicians may overestimate variables related to the success of their treatment (for example, response rates, mortality, complications, adverse events), and also underestimate variables related to healthcare use, in order be more efficient.

### *Other Data Sources*

Drug textbooks may yield information on dosages of drugs, drug prices and reimbursement percentages. Official tariff lists from health insurance companies contain information on tariffs for other medical costs (procedures, consultations and hospitalisation). Financial departments of particular institutions may yield information on costs from the provider's perspective.

## GENERAL STRATEGY

To illustrate the application of a general strategy, a Markov model is used and applied to the economic outcomes of maintenance treatment with selective serotinin reuptake inhibitors (SSRIs) as the first-choice therapy for depression over a 1-year follow-up period. The general structural details of the Markov depression model have already been described in a previous paper,[14] and the specific details of the model with respect to depression are shown in figure 1. This model defines 4 mutually exclusive states:

- no depression
- mild depression
- severe depression
- chronic depression.

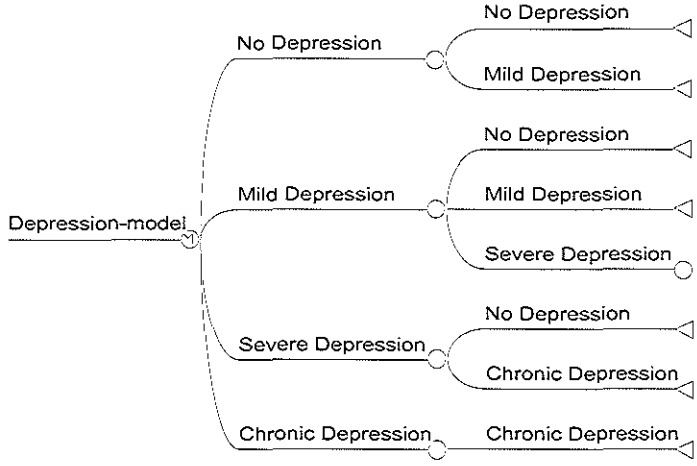
The patients' progression through these states is divided into 6 cycles of 2 months each, which closely approximates the time of the sequential therapeutic stages. Chronic depression represents an absorption state and after entering this state a patient remains there. The transition probabilities between the states are based on response rates to treatment and relapse rates after termination of treatment. This information was obtained from published literature values and current clinical practice. The cost assessment was based on the assignment of a fixed cost to each health state, which depends on the choice of therapy. Further details of the sources for the clinical and economic data, with respect to depression, are given in a previous paper.[14]

## Initial Medication

### *Transition Probabilities*

The probabilities for the initial transitions in a model usually correspond with the efficacy measures derived from a clinical trial comparing treatment with the new drug versus placebo or another drug therapy. In the model, the transition probabilities for the comparative treatment may be derived from the clinical trial if the chosen comparator for the pharmacoeconomic study corresponds to the comparator in the clinical trial of the new drug. When the comparator in the clinical trial is another drug or placebo, the transition probabilities for the pharmacoeconomic comparator need to be derived from published literature.

**Figure 1. Markov model for depression used in the present study.**




Because, in this case, the transition probabilities for the comparator are derived from a different data source, any dissimilarity in patient populations and protocol procedures that lead to a bias has to be controlled. For example, in the antidepressant model, the first option was to base the transition probabilities to the health state "no depression" on the response rates for citalopram and tricyclic antidepressant (TCA) recorded in a clinical trial. Because the clinical trial of citalopram was not based on comparison with a TCA, the response rate for a TCA was derived from a large meta-analysis.[15] This meta-analysis was a weighted average of the response rates of the 4 most widely used antidepressants in Germany, according to the proportion of their volume market shares (IMS).

In this study, we deviated from the general rule to derive the response rate of the new drug from the clinical trial. The response rate for citalopram was assumed to be equal to that of standard therapy, although in a multi-centre study by Rosenberg et al,[16] superior response rates for citalopram were found. Since this superior response rate for citalopram was based on data from a limited number of trials, and there is no evidence that the response rate for citalopram is higher than the response rate for other SSRIs, the higher response rate found in the clinical trial may be a result of statistical variance. Because SSRIs and TCAs have a similar clinical efficacy, the response rate for citalopram was assumed to be equal to the response rate for TCAs.

### *Healthcare Resource Use*

The healthcare resources used, associated with the period before the initial transitions may be derived from a clinical trial, but this information may be of limited use because of its low external validity; consequently, it may lead to an overestimation of the healthcare resources used.

Other potential data sources are databases and published pharmacoeconomic literature. These data sources may, only in a few cases yield, information of sufficient quality corresponding to the following study characteristics:

- study country
- perspective
- initial study population and different subpopulations (patients with treatment failures).

Another option is the use of expert opinion, which is often used in pharmacoeconomic studies. According to the guidelines the use of expert opinion is only allowed in situations in which no other data source is available.

Finally, dosages of drugs may be derived directly from drug textbooks, when their use in the model corresponds with indications in these books. The advantage of this approach is that it is time saving, and consequently cost savings, when compared with other methods of collecting this information.

For example, in the depression model described here, the units of healthcare use could not be derived from the clinical trial or other published country-specific literature. Hence this information was based on a Delphi panel, except for the drug dosages, which were based on standard prescription guidelines according to the official German drug textbooks (Rote Liste).[17] The Rote Liste also gave the reimbursement percentages and daily treatment costs for the antidepressant drugs, while for other medical resources, this information was based on prices and/or tariffs from official health insurance tariff lists. The associated 1993 unit costs of the medical resources used (e.g. various professional services, health services facilities and tests carried out) were derived from official German tariff lists Bewertungsmaßstab für kassenärtzliche Leistungen (BMA)[18]] and other sources [Deutsche Ort Krankenkasse (DOK 8[19]]. Since citalopram was not registered in Germany yet, we used the weighted average daily treatment cost (DTC) of SSRIs in Germany as a proxy in our calculations [2.58 Deutschmarks (DM)].

## Treatment Failure

Information on patients with a treatment failure usually cannot be derived from a clinical trial for two reasons: (i) the majority of clinical trials are not based on an intent-to-treat design and, consequently, these trials do not yield information on patients withdrawing as a result of treatment failure (e.g. adverse events, noncompliance); (ii) the treatment pattern for a patient completing the clinical trial, but showing no response, cannot be determined because of the end of the follow-up period.

While the follow-up of treatment failures may not be relevant for registration purposes, from a pharmacoeconomic perspective, patients who have failed treatment are a critical subpopulation in a model, and pharmacoeconomic studies show that treatment failures may account for the majority of costs (e.g. hospitalisation. For instance in depression the main cost driver was hospitalisation for treatment of resistant depression: a minority of the patients accounted for the majority of the costs.[14] Therefore in any pharmacoeconomic study the impact of the costs associated with treatment failure need to be assessed. If a new product is more efficacious, leading, for example, to a higher response rate (derived from a trial), the lower number of treatment failures will often lead to cost savings.

The impact of the cost-effectiveness has to be determined qualitatively by describing the disease progression of these patients in the model, for which information on treatment pattern, healthcare resource use and transition probabilities is required.

The treatment patterns after a treatment failure has to be based on the usual local practice in a study county, which may be based on local guidelines that need validation, because real practice may differ from the guidelines. Expert opinion is currently the most suitable method for this confirmation.

When there are no guidelines, databases or expert opinion have to be used to determine standard therapy for second and subsequent treatment failures. These standard therapies may consist of different treatments. For instance, treatment failure because of an adverse event will lead to switching to another therapy, while treatment failure resulting from poor efficacy of the product may lead to increase in dosage.

The pathways for all causes of treatment failure have to be incorporated into the structure of the model. In case there is no consensus on the first-choice therapy for a type of treatment failure, the subsequent usual care may be a weighted mix of existing therapies (e.g. 50% switch medication; 50% increase medication).

For the transition probabilities and for the units of healthcare use associated with treatment failures, an assessment of databases and epidemiological data in the literature has to precede the use of expert opinion. As a result, expert opinion may need to be used for only transition probabilities, healthcare resource use, or both. However, the extrapolation of data derived from literature and databases is only possible if these data sources deal with a similar study population (e.g. patients with a treatment failure after previous therapy). For subsequent treatment failures after the first treatment failure, the abovementioned strategy can be used.

Because study populations in data sources must have a similar medical history of treatment failures, the chances of finding data sources dealing with the appropriate study population, or of finding an appropriate secondary hard data source will decrease after each treatment failure. An option for these patients with multiple treatment failures is to assume that their transition probabilities and/or healthcare resource use will be similar to patients with an initial or second treatment failure. These assumptions have to be validated by expert opinion. If these assumptions cannot be justified expert opinion has to be used for the data collection.

Generally it is more difficult to find appropriate hard data sources for the transition probabilities that may be derived from published clinical trials, than for the healthcare resource use. Databases will be the most relevant hard data sources for the determination of healthcare resource use. The prices and tariffs will be based on the same data sources as discussed in previous section on "Initial Medication".

For example, in the model, the treatment pattern after an initial treatment failure, while receiving antidepressant therapy, was derived from the clinical guidelines defined in the US Clinical Practice Guidelines.[15] The choice of third-line therapy was also derived from the same source, but the response rate to this therapy, which determined the subsequent transition probabilities, could not be derived from the literature. The assumption was made that these transition probabilities were similar to second-line therapy, an assumption that was validated by an expert. The units of healthcare use could not be derived from a country-specific database, and were consequently derived from expert opinion.

### Responders

In most clinical trials, the response of patients is assessed after a short term study period. For treatment of acute diseases, such as community-acquired pneumonia, the response corresponds to a complete recovery without a higher risk for a new episode than the average population; here, there is no need to follow responders for a longer period of time. In the model, the responders will not lead to costs and remain in perfect health.

In the case of chronic diseases or increased risk of a new episode, the trial may only follow the patient to a response after a short term therapy. For instance in depression the objective of most trials is to prove a better response rate to short term treatment with the new antidepressant versus placebo and/or another antidepressant. Hence, the study duration of most clinical trials in this indication is only 6 to 8 weeks, which corresponds to the acute short term treatment period. In situations in which a long term effect is expected after treatment termination, or the continuation of a successful treatment is evaluated, patients need to be followed for a longer period of time. For instance, there have been a number of clinical trials assessing the long term effect of continuation treatment with antidepressants versus short term treatment only. In case patients are followed after response, the data collection method described for "initial treatment" may be applied. In case patients are not followed, the data collection method described for "treatment failure" may be applied.

The prices and tariffs will be based on the same data sources discussed in section on "Initial

Medication". For example, the "initial treatment" strategy was used; relapse rates were derived from the citalopram clinical trial and transformed into transition probabilities. [20] The relapse rates after successful episodic treatment with a TCA were based on the responders to an initial treatment with an SSRI, who were switched to placebo; hence, the assumption was that responders to an SSRI and a TCA will have similar relapse rates. Finally, healthcare resource consumption used was based on expert opinion.

## RECOMMENDATIONS

The following guidelines are proposed for the use of secondary data sources in pharmacoeconomic studies.

Readers should be able to understand the methods used for the selection and use of data sources in modelling studies. The sources of study data should be recommended and explained in sufficient detail. The reader should be able to evaluate the strengths, weaknesses and possible sources of bias that may be inherent in the data used in the analysis. Selection criteria for studies and databases should be described and authors should explicitly note the direction and magnitude of potential bias in the data sources used. When clinical and/or economic (in case of a "piggy back" trial) data are derived from a clinical trial, details of the design, and results of the study should be given [e.g. study population, follow-up period of patients, follow-up of patients who withdraw and treatment failures (intention-to-treat), method of randomisation and primary and secondary clinical outcomes with confidence intervals].

The general applicability of the study population is important in assessing the results of clinical trials for suitability for economic evaluations. This level of uncertainty may be dealt with using a sensitivity analysis. Factors that can limit the application of the results include differences across countries or healthcare systems, and benefits that result only from a clinical trial protocol, but that would not arise in real practice, such as unrealistically high compliance rates.

For clinical outcomes, the general rule may be to assume that data are not country-specific. Hence, data may be derived from studies performed in other countries and pooled results of international studies may also be used. For each study, this assumption has to be controlled. In some indications, clinical outcomes are country-specific. An example is the response rate to treatment with antibiotics in pneumonia, which will vary from country to country because of differences in resistance patterns. When data are derived from a synthesis of a number of articles, details should be given as to the method of synthesis, or meta-analysis of material (e.g. search strategy and criteria for inclusion of studies in the overview). For each study, the clinical outcomes should be presented, with information on the study population and the number of patients. Clinical outcome measures derived from overviews have the advantage that the confidence interval is usually narrower than that from an individual trial, and the results may be more generally applicable. Typically, the point estimate would be used in economic studies as the best case value and the confidence interval would be used as the

relevant range for sensitivity analysis. For economic measures and information on therapeutic choices, the general rule may be that country-specific data sources have to be used, while clinical measures usually do not need to be derived from country-specific literature.

At each location in the model (e.g. Markov state), the patient subpopulation has to correspond as much a possible with the population in the data source (s) being used. Both the similarities as well as the differences should be mentioned. The first step is to describe the subpopulation in the model (e.g. population after 1 failure to initial therapy). The second step is to list the transition probabilities, costing and therapeutic choices from the available data sources. The next step is to describe, for each data source: type (e.g. database, medical record), number of patients, study population (inclusion and exclusion criteria), countries, date of data collection, cost of access to database and data abstraction. A justification for the final "YES" or "NO" decision to use a data source should be presented, based on the advantages and disadvantages of the specific source.

Finally the following criteria for the selection of members for a Delphi have to be considered. First, the members of the Delphi panel have to be a representative sample of the population of physicians treating the defined patient population under investigation. Second, the type or types of physician(s) involved (GP, hospital-based specialist, office-based specialist) must be decided on. It is possible that different types of physicians need to be included; for example, patients with depression may be treated by either a GP or psychiatrist, while a patient experiencing treatment failure (e.g. resistant depression) may be referred to hospital for a treatment by a hospital-based psychiatrist. Third, the physician should have enough experience in treating the defined patients, which may be assessed by the number of treated patients who have the disorder that is under investigation. Fourth, the physician should not only have experience with the disorder, but also with the defined "usual care" as initial treatment. Finally, the selection of a physician should correspond with the perspective of the study [e.g. in the US, a physician's contract may be either with the health maintenance organisation (HMO) or another managed care organisation].

## DISCUSSION

With the growing increases use and growing importance of modelling studies for economic evaluations, a new area for research has been created. In order to obtain objective and reproducible results from those studies it is important to have standardised methods of evaluation, contained in accepted guidelines on methodology. Although various groups have published recommendations on the good practice of economic evaluations[21 22 23], those guidelines mainly focus on prospective studies that rely on primary data, and consequently, contain no recommendations for the handling of secondary data, which are closely related to the appropriate execution of a modelling study and the credibility of the results from a modelling study.

From the perspective that results from modelling studies are of interest to many parties with different responsibilities and diverging interests, it is obvious that besides guidelines for the selection of data

sources, transparent and unambiguous presentation of the data sources is required for a proper understanding.

A standardised strategy for data source selection and a transparent reporting format on data source selection has been presented in this article, which may improve the acceptance of such data and thus stimulate further research, even in the absence of formal guidelines.

## References


1 Australia Commonwealth Department of Health, Housing and Community Services. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Commonwealth Department. 1992

2 Commonwealth Department of Australia Department of Human Services and Health. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Australian Government Publishing Service, 1995

3 Canadian Coordinating Office for Health Technology Assessment (CCOHTA). Guidelines for economic evaluation of pharmaceuticals: Canada. 1st ed. Ottawa: CCOHTA. 1994

4 Ontario Guidelines for Economic Analysis of Pharmaceutical Products. Ontario: Canadian Ministry of Health, Drug Programs Branch, 1993

5 Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-65

6 Egger M, Davey-Smith G. Misleading meta-analysis. British Medical Journal. 1995: 310; 752-754

7 Detsky AS, O'Rourke KO, Corey PN, Johnston N, Fenton S. The hazards of using active clinic patients as a source of subjects for clinical studies. J Gen Intern Med. 1988; 32: 260-66

8 Sacket DL, Bias in analytical research. J Chronic Dis. 1979; 32:51-63

9 Jariath N, Weinstein J. The Delphi methodology (part two): a useful administrative approach. CJONA 1994; 7(4): 7-20

10 Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995; 311: 376-80

11 Barr J, Schumacher G. Using decision analysis to conduct pharmacoeconomic studies. pp. 1197-1214, Quality of life and pharmacoeconomics in clinical trials (2nd edition), ed by B. Spilker 1996; Philadelphia: Lippincott-Raven Publishers

12 Williams P, Webb C. The Delphi technique: a methodological discussion. Journal of Advanced Nursing 1994; 19: 180-186

13 Evans C. The use of consensus methods and expert panels in pharmacoeconomic studies: practical applications and methodological shortcomings. Pharmacoeconomics. 1997; 12 (2Pt 1): 121-9

14 A Markov Process Analysis Comparing the Cost-Effectiveness of Maintenance Therapy with Citalopram versus Standard Therapy in Major Depression. Nuijten M, Hardens M, Souctre E. Pharmacoeconomics. 1995; 8 (2): 159-168

15 Clinical practice guideline. Depression in primary care: treatment of major depression (AHCPR publication no. 93-0551). Rockville (MD): Public Health Service, Agency for Health Care Policy and Research, 1993

16 Rosenberg C, Damsbo N, Fuglum E, et al. Citalopram and imipramine in the treatment of depressive patients in general practice: a Nordic multicentre clinical study. Int Clin Psychopharmacol 1994; 9 Suppl. 1: 41-8

17 Rote Liste 1993. Arzneimittelverzeichnis des BPI. Aulendorf/Würtemberg: Edition Cantor Verlag für Medizin und Naturwissenschaften GmbH, 1993

18 BMA. Bewertungsmass-stab für kassenärtzliche Leistungen. Cologne: Deutscher Artze-Verlag GmbH, 1993

19 DOK-8: AOK-Bundesverband, 1994 Apr 15: 256

20 Montgomery SA, Rasmussen JGC, Tanghoi P. A 24-week study of 20 mg citalopram and placebo in the prevention of relapse of major depression. Int Clin Psychopharmacol 1993; 9: 181-8

21 Clemens K, Townsend R, Luscombe F, Mauskopf J, Osterhaus P, Bobula J. Methodological and conduct principles for pharmacoeconomic research. Pharmacoeconomics 1995; 8 (2): 169-74

22 Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996; 313: 275-83

23 Jefferson T, Demichelli V. Are guidelines for peer-reviewing economic evaluation necessary ? A survey of current editorial practice. Health Econ 1995; 4: 383-8

# Chapter 5

# Bridging Decision-Analytic Modelling with a Cross-Sectional Study: Application to Parkinson's disease


## SUMMARY

The ideal design for demonstrating the possible health outcomes and costs associated with a new drug would be a naturalistic prospective study. However, it is often not feasible to derive the required information from scientifically sound prospective studies. In these cases, decision-analytic models may provide some of the missing information. However, the use of a Delphi panel to gather data for these models is a major concern because of potential bias and data accuracy. Because reimbursement of pharmaceuticals is often based on economic data derived from modelling studies, it is obvious that potential bias due to the use of Delphi panels should be minimised.

In this manuscript we will present an alternative data source for a modelling study: the cross-sectional study. Data from such studies can be used to yield costs and utilities for Markov health states. The overall combined design may be considered a hybrid between a naturalistic prospective study and a modelling study by maximising the pros and minimising the cons of both types of design, including an increase of external validity. This hybrid design is based on bridging the probabilities derived from the literature and clinical trials with information on costs and utilities from a cross-sectional study. This design has also logistical advantages, namely a shorter required study duration compared with prospective naturalistic studies for chronic diseases. This combined design was illustrated using a Markov model for Parkinson's disease.

## INTRODUCTION

Most countries are, and have been during the last 2 decades, confronted with the problem of increasing costs for healthcare and how to finance these expenditures. In The Netherlands, the main response to rising costs has been the introduction of an impressive number of cost containment policies over the last decade. This resulted in the Drug Reimbursement System, which is based on the classification of drugs into groups of interchangeable drugs (Geneesmiddelen Vergoedingssysteem: GVS). In this system, there is a fixed refund price being based on the average list price for "therapeutically interchangeable drugs" belonging to the same group (cluster). Any new drug, which cannot be clustered, is not reimbursed, unless there is no treatment for the relevant pathology. Currently, the Dutch government is investigating making the submission of health economic data an official requirement when applying for reimbursement of a new innovative drug, as in Australia and Canada.[1 2]

Because there were no requirements for phamacoeconomic data when most of the current phase III clinical trials were started, pharmacoeconomics was, in general, not included in the clinical programmes of products now approaching launch. Therefore, it is expected that for most of the submissions for reimbursement in countries such as The Netherlands in which the use of pharmacoeconomic data will become an official requirement, data will be based on a modelling design.

If reimbursement of pharmaceuticals is going to be based mainly on economic data derived from modelling studies, it is especially vital to carefully scrutinise and refine this type design. A main concern about the use of modelling studies is the use of a Delphi panel to gather data which could not be derived from actual existing data sources. The use of a Delphi panel may particularly controversial in estimating quality-adjusted life-years (QALYs), which may become the primary final outcome for pharmacoeconomic studies. The existing Canadian and Australian guidelines and the draft UK guidelines strongly recommend an estimate of the incremental cost per QALY gained.[1 2 3]

The objective of this manuscript is to present a strategy to improve the quality of the data for use in pharmacoeconomic modelling studies by using an add-on cross-sectional study as an additional data source i.e. combining a modelling study with an add-on cross-sectional study.

## MODELLING STUDIES

The ideal design to demonstrate the possible health outcome and costs associated with a new drug is a naturalistic prospective study. However, in practice, it is not always possible to derive all required information from scientifically sound prospective studies. A modelling design may chosen because of practical limitations of prospective studies. This is especially the case in chronic conditions like Parkinson's disease; required study durations may vary from 5 to 10 years. In these cases, models can be used to extrapolate clinical outcomes beyond the duration of the trial. Furthermore, the forthcoming

legislation requiring pharmacoeconomic data for reimbursement decisions which need to be made within a limited time frame will expand the indication for modelling studies to other conditions (e.g. acute phases of a disease such as community-acquired pneumonia). Modelling is an attractive alternative for the generation of pharmacoeconomic data within short periods e.g. for products which are nearing launch, as is the case in The Netherlands (see Introduction).

Modelling studies are based on decision analysis, which allows both clinical and economic consequences of medical actions and attitudes to be analysed.[4] From treatment algorithms a model can be constructed which considers the timings of actions and their consequences over time. In effect, a model shows the consequences and complications of different therapeutic interventions. It should correspond as much as possible to the change to the course of the disease. Models may take the form of simple decision-analytic trees or they may be very complex Markov models. While decision tree models are appropriate for acute episodes, Markov models are the first choice for pharmacoeconomic analysis of chronic diseases.

## Data Sources

The data used in a modelling study can be categorised into transition probabilities, treatment patterns (or therapeutic choices), healthcare utilisation and utilities and prices and tariffs. The data may come from a variety of sources and are subject to varying degrees of uncertainty. Data sources for the variables used in a model may be clinical trial reports, literature (e.g. meta-analysis), claim or clinical databases, medical records, and official tariff lists. These data sources yield, for each variable, a fixed input value and a range. Standard analyses are based on the fixed input value for all variables. The range of each variable is then used to determine the sensitivity of the outcome to the analysis when the input value is varied within its range.

The limitations of the various types of data sources have been extensively discussed previously.[5] It is usually not possible to derive all data from actual existing data sources, because of lack of available data sources or unreliable data. The main drawback of all existing data sources is that they have not been developed for economic evaluations but for either administrative or medical/scientific purposes. Hence they may suffer from incompleteness (e.g. no information on either healthcare utilisation or clinical outcomes, especially QALYs), insufficient external validity and an inappropriate format of the information (i.e. one that does not fit the structure of the model.

When there are no reliable data available, the use of Delphi expert opinion is considered appropriate.[6] Delphi panels operate in rounds, in an effort to reduce variance in estimations by the experts in a particular area. The first round consists of individual interviews with the participating experts. The first round consists of individual interviews with the participating experts. The experts are provided with a description of the model and are asked to provide information based on their clinical experience and their knowledge of the literature. The interviewer asks the expert to provide information of the following items: probabilities, healthcare utilisation and therapeutic choices. All information on these

items has to match the basic characteristics of the model (time horizon, cycle time, health states). All responses by panellists are anonymous to other panellists. The rationale for this is that respondents will not be intimidated or dominated by other respondents. This is a particular concern in the healthcare profession[7] where there is a hierarchical structure in healthcare delivery and seniority dominates most care activities. In the second round the experts are asked to re-estimate the estimation provided in the first round. The second round is performed by having each expert fill in a questionnaire, which is sent by email, phone or fax.

There are a number of methodological weaknesses inherent in the Delphi process (see Nuijten[5]). Briefly, these are the potential for experts to drop out following the first round, the potential for outliers related to sample size, and the potential for adjustments of the estimations to ideal care (overestimation of variables related to the success of treatment, or underestimation of variables related to healthcare utilisation). The main methodological weakness of the Delphi panel process is that the data are based on estimations, whereas data from other data sources are based on real measurements i.e. the perceived values of the physicians may differ from real values.

### *Transition Probabilities*

For clinical outcomes, the general rule is to assume that transition probabilities are not country specific. Hence, data may be derived from studies performed in other countries and pooled results of international studies may also be used. Probabilities for clinical outcomes associated with treatment following the initial treatment (e.g. second-line therapy) may often be derived from the literature. However, clinical trial data for determining probabilities of clinical outcomes associated with subsequent treatments (e.g. third-line therapy) will be scarce. In this case, epidemiological studies may yield data on disease progression, which may be used to determine transition probabilities for those patients. The Delphi panel technique may need to be used to yield all remaining probabilities. Because a model is usually only marginally sensitive to probabilities associated with treatment failures after second-line treatment, the impact of uncertainty associated with the methodological weaknesses of Delphi panel technique in this case would be low.[5]

### *Treatment Pattern*

Data on treatment patterns or therapeutic choices may be based on local guidelines or sometimes claim or clinical databases, which are validated by expert opinion. However, for the majority of modelling studies, there will be no actual existing data and the use of expert opinion is the only option for determining treatment pathways. The first step is to obtain a quantitative description of treatment pathways by consulting 2 experts. The next step is to use a Delphi panel to determine the probabilities for each treatment option, when several treatment strategies exist. Determination of these probabilities by using a Delphi panel technique may be appropriate when the sample of Delphi panel members is representative for usual care (i.e. they treat their patients according to the most widely used treatment in the study country), and the sample size of the Delphi panel is sufficient.

### *Healthcare Utilisation and Utilities*

Data on healthcare utilisation and utilities usually needs to be completely derived from a Delphi panel.[5] Hence, a model will be especially sensitive for effectiveness (QALYs) and cost outcomes

### *Other Data Sources*

Other data sources for modelling studies may include:

- drug textbooks that contain information on dosages of drugs, drug prices and reimbursement percentages.
- official tariff lists from health insurance companies contain information on tariffs for other medical costs (procedures, consultations and hospitalisation).
- and financial departments of particular institutions, which may yield information on costs from the provider's perspective.

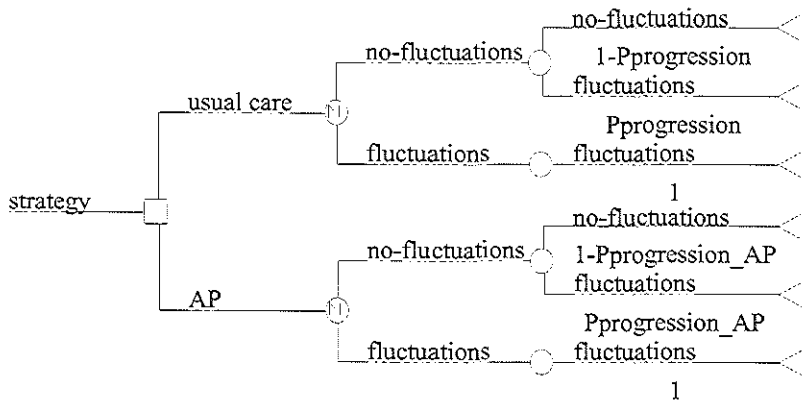
## STRATEGY

In this section, a strategy is described for substantially reducing the amount of data which need to be derived from a Delphi panel – a strategy which may substantially improve the quality of the input data in a model, and consequently, the cost and effectiveness outcomes. The strategy is illustrated by using a hypothetical Markov model for Parkinson's disease which was used to model the cost-effectiveness of a hypothetical new antiparkinsonian drug (AP) from a German health insurance (Krankenkassen) perspective in Germany. This "new drug" is modelled as add-on therapy to usual care (fig. 1). Our model is based on the following assumptions:

- the Markov health states in the model correspond with severity levels according to the existence of motor fluctuations ("no fluctuations" and "fluctuations").
- the follow-up period is 5 years; the cycle time is 1 year, which is based on the follow-up period of the clinical trials for the AP.
- there is no mortality during the study period.
- beyond the actual follow-up period of the clinical trial for the AP, the delay in disease progression due to treatment with the AP will be modelled using the same rate beyond the clinical trial period.
- disease progression corresponds with the development of fluctuations.
- a hypothetical cohort of persons with no fluctuations will receive the AP or usual care. After the first cycle, patients may have remained without fluctuations or have become "fluctuators". Patients will continue the treatment with the AP for the remainder of the study period irrespectively or whether they are fluctuators or not.
- the AP will reduce disease progression annually by 25%.

The assumptions, related to hypothetical product AP, are purely hypothetical; they are not based on any published or unpublished clinical trial data. The definition of the Markov health states according to the existence of motor fluctuations was based on clinical literature.[8]

**Figure 1. Markov model for Parkinson's disease.**




## Effectiveness

Utility is a general concept for measuring the value individuals attach to the consequences of various actions (consequences of several treatment options in our example).[9] The goal of this measurement technique is to obtain a numerical value that represents the strength of the individual's preferences for a particular outcome.[10] Utilities can be used in a cost-utility analysis, with results presented as costs per QALY for each of the treatment alternatives.[11 12 13 14]

The assessment of utility was based on the following formula:

$$\text{utility}=\sum_{s=1}^{n} t_s \times u_s$$

Where: $n$ is the total number of cycles spent in a particular health state; $s$ is the health state; $t$ is the number of cycles spent in a particular health state; and $u$ is the utility of the particular health state ("no fluctuations" or "fluctuations"). The latter is used to relate each health state to an average QALY. Various health state classification systems are available (e.g. Torrance Index[11] and EuroQol.[15]
In our model, utilities were derived from the patients by using the Torrance Index.

## Cost Assessment

The cost assessment was based on the assignment of a total fixed cost to each health state corresponding with a cycle time of 1 year. The costs of each health state was determined by the healthcare utilisation associated with a health state i.e. medication, consultations, procedures, tests and number of days of hospitalisation. Each health state is associated with a mix of healthcare utilisation;

treatment costs for all patients are included as well as costs associated with supportive care (weighted according to the proportion of patients).

In our model, costs were determined from the perspective of the Krankenkassen. The costs were derived from the following sources: drug costs - official German price lists of drugs[16 17]; hospital costs - mean cost per day of inpatient care[18]; visit costs for GPs and office-based specialists - official tariff list [19 20]; diagnostic costs - official tariff lists.[19 20]

## Clinical Trials

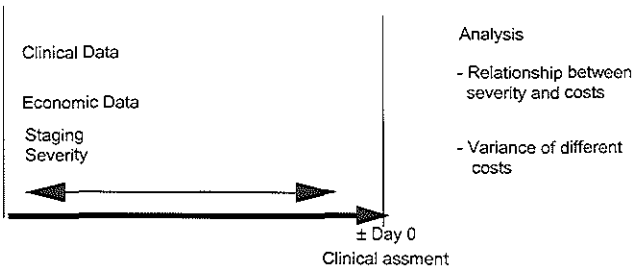
Transition probabilities between the different health states may generally be derived from clinical trials and epidemiological or observational studies. In our example, we derived the annual transition probability from "no fluctuations" to "fluctuations" for usual care from a study by Dodel et al. (Table I).[21] The average duration of disease since the onset of symptoms was determined for each patient at inclusion. The average duration of disease since onset of symptoms was 7.1 and 13.9 years for patients without fluctuations and with fluctuations, respectively. Hence, the average duration to progression to fluctuations is 6.8 years, which can be transformed into an annual probability of 0.147 by taking the reciprocal of 6.8 years. This calculation is based on a method by Beck to determine annual mortality by taking the reciprocal of life expectancy.[22] Disease progression with the AP in our model was derived from a hypothetical clinical trial showing that the AP reduces disease progression by 25%.

## Cross-Sectional Cost-of-Care Study

Our strategy includes a cost-of-care study based on a cross-sectional study design, which was performed for our model. The defining characteristic of a cross-sectional study is that the basic health state of each person in the evaluation is examined essentially once. QALYs and costs during a period of time are determined at one point in time for each patient (fig. 2). These measures of utility (QALYs) and resource utilisation can be determined for each severity level and subsequently the relationship between utility markers of clinical status and costs can be established. A cross-sectional study consists of the following steps, which have been adjusted to Parkinson's disease.

- *Selection of sites*. These sites have to be a representative sample of providers that are treating the patient population under investigation is being treated. It is possible that different types of sites need to be included. In our example, the majority of patients without fluctuations may be treated by a GP, whereas the majority of patients with fluctuations may be treated by a neurologist. In addition, the sites should have adequate experience in treating the defined patients; this may be assessed by the annual number of patients treated with the pathology under investigation.

**Figure 2. Flow chart for a retrospective cross-sectional study.**




- *Selection of patients.* In our model, a defined number of patients with Parkinson's disease was selected from the registries of the participating sites corresponding with the defined study population in the model. In this case, the primary diagnosis was idiopathic Parkinson's disease (clinically characterised by bradykinesia, resting tremor, cogwheel rigidity, and postural reflex impairment)[8]. This definition excludes all parkinsonism of known aetiology and any disorder with multiple system involvement or significant lesions of the striatum, such as progressive supranuclear palsy, olivopontocerebellar atrophy, multiple system atrophy, striatonigral degeneration.
- *Screening for eligibility criteria.* Subsequently, all selected patients with idiopathic Parkinson's disease were screened for the inclusion and exclusion criteria. Inclusion criteria may be: (i) presence of a caregiver able to give valid information; (ii) patients insured with the Krankenkassen; and (iii) patients who have consented to participate. Exclusion criteria may be: (i) patients with nonidiopathic parkinsonism; (ii) patients with severe neurological or psychiatric disorders, particularly dementia; (iii) patients with severe organic pathologies (e.g. neoplastic disease, severe cardiovascular disease); and (iv) patients currently involved in a clinical trial. The initial screening was based on the medical records. The patient was then invited to participate in the study (direct phone calls could be made). The study investigator arranged an interview with each patient and, if necessary, the caregiver. At the time of the visit, each patient was evaluated for inclusion using the study criteria and, if proven eligible, would be included. In case of inclusion, the patient and his/her related caregiver were interviewed to complete the socio-demographic, medical, and economic section of the Case Record Form (CRF). To confirm and/or add information to be recorded in the CRF's, the study investigator was allowed to contact other sources, typically the patient's GP or a hospital in the case of previous in patient care.

- *Determination of health state utilities.* At the end of the interview, health state utilities were determined (e.g. Torrance Index or EuroQol). If necessary, the patient was instructed on how to use the questionnaire.
- *Patient stratification.* The selected patient population was stratified into groups of disease severity according to existence of fluctuations i.e. "no fluctuations" and "fluctuations.
- *Data collection.* All relevant units of resource utilisation in the 3 months preceding the study visit were collected for each patient.
- *Determination of input values.* The mean and distribution of all units of costs and utility were determined for each severity level; these were incorporated into the corresponding Markov health states in the model.
- *Statistical analyses.* Statistical analyses were performed to identify any confounding variables (e.g. the type of Parkinson's disease).

The cost-care study in this example yielded relationships between severity level and costs for Germany, which are specific for Germany. The function between severity and utility may be extrapolated to other countries, if we assume that clinical outcomes are not country specific.[5] Table I shows for the relationship between costs and utility according to the existence of fluctuations.

**Table I. The fixed input values for the Markov model.**

| Stage | Costs (DM) (3 months) | Costs (DM) per year | Utility | Duration of disease since onset of symptoms (y) | Differ-ence | Transition Probabilit y |
|---|---|---|---|---|---|---|
| | Mean | | Mean | mean | | |
| No fluctuations | 1,020 | 4,080 | 0.73 | 7.1 | | |
| Fluctuations | 2,050 | 8,200 | 0.49 | 13.9 | 6.80 | P=0.147 |

## Markov Model

Subsequently, a Markov model can be used to simulate the disease progression and accrual of costs and compare the cost-effectiveness of the AP versus usual care by combining data derived from the literature and clinical trial (AP transition probabilities) with data derived from the cross-sectional study (costs, utilities). This approach has similarities with the concept of a "longitudinal prevalence analysis", which is used in epidemiology. The costs in this analysis were discounted at 5%; no discounting was applied to the utilities. The results of the analysis are shown in Table II. The results of the baseline analysis show that the use of the AP [daily treatment cost (DTC) is DM2 ($US1.5)] is more costly than usual care, with direct medical costs DM25,265 ($US18,254) versus DM22,967 ($US16,594), respectively. A secondary analysis was performed in order to assess the impact of AP on current costs associated with usual care, by excluding the drug costs of the AP. This analysis showed that the use of the AP would reduce the current costs for usual care from DM22,967 ($US 16,594) to DM21,962 ($US15,868). The AP was associated with only a 2% increase of effectiveness: 3.41

QALYs versus 3.35 QALYs. Consequently, the use of AP would lead to an incremental cost-effectiveness ratio of DM34,294 ($US24,777) per QALY.

**Table II. Results of the cost-effectiveness analysis.**

| | AP | Usual care | Incremental cost-effectiveness (ICER) $\frac{\Delta C}{\Delta E}=\frac{C_E-C_M}{E_E-E_M}$ |
|---|---|---|---|
| Costs (DM)* | 25,265 | 22,967 | 34,294 |
| QALY | 3.41 | 3.35 | |

*: Discounting at 5%.

The results of a sensitivity analysis performed for percentage reduction in disease progression by the AP, are shown in Table III. Results show that the model is rather sensitive to a change in the percentage reduction of disease progression; at least 50% reduction in disease progression may justify the use of AP in patients with Parkinson's disease.

**Table III. Results of the sensitivity to reduction of disease progression.**

| | Usual care | AP | | | | |
|---|---|---|---|---|---|---|
| **Reduction** | | **0 %** | **25 %** | **50 %** | **75 %** | **100 %** |
| QALY | 3.35 | 3.35 | 3.41 | 3.48 | 3.57 | 3.65 |
| Costs (DM)* | 22967 | 26270 | 25265 | 24182 | 23016 | 21762 |
| ICER | | usual care dominant | 38300 | 9346 | 223 | AP dominant |

*: Discounting at 5%.

## DISCUSSION

Because study drugs are usually not approved for registration at the time economic studies are conducted, prospective naturalistic trials and the submission of data on effectiveness and expected costs at the time of reimbursement are usually not feasible. The execution of a prospective naturalistic study would substantially delay the product launch, shortening the period of useful patent life and reducing the return on research and development investment.

The use of a hybrid design (combining a modelling study with an add-on cross-sectional study) has logistical advantages compared with a prospective naturalistic study. The model development and the execution of the cross-sectional study can be performed within 6 months, which compares favourable with the optimal duration of prospective naturalistic studies for chronic diseases (which varies between 1 year and a life-time follow-up). Aside from this practical convenience of data availability,

observational data derived from a cross-sectional study may well have higher external validity than the data derived from randomised prospective trials.

A hybrid design may also have other methodological advantages compared with the use of data from a naturalistic study. Instead of data collection being retrospective, data from a hybrid design may be collected prospectively during, for example, a 3-month period following inclusion. Both the prospective and retrospective approach may have pros and cons. The risk of the retrospective approach is that data collection on resource use may be incomplete. For instance, in the case of switching between in- and outpatient care, different medical records may need to be accessed. The risk of the prospective approach is the introduction of a potential bias on both part of the patient and the physician. For example, physicians may adjust their treatment patterns to those of ideal care, which may reduce the external validity of the prospectively collected data (e.g. resource use may be underestimated). Hence, an advantage of the retrospective approach to data collection is the high external validity of the data. An advantage of the prospective approach is that the data collection form (CRF) will be primarily developed for the cross-sectional study. Consequently, the type of data (clinical and economic) collected will be pre-defined and will correspond with the characteristics of the model.

A further methodological advantage of a hybrid design is that economic measures will be determined over a fixed period of time (3 months in our model for Parkinson's disease). A requirement for the cross-sectional approach is that patients do not switch between health states over the study period; the costs associated with each health state need to be determined. In our model, we assumed that patients with Parkinson's disease will not have disease progression during a 3-month study period. Consequently, another advantage of the prospective approach is that clinical assessments may be repeated at the end of the study period in order to confirm the stability of the disease over the chosen study period. The risk of disease progression may be reduced by minimising the follow-up period, which will increase the statistical error. Consequently, a larger sample size may be required.

In our model, resource utilisation, excluding the AP, and utilities, were derived from a cross-sectional study based on usual care practises i.e. without the use of the new hypothetical product. The resource utilisation and utility of a health state are functions of both the underlying disease and adverse events associated with treatment. In our model, we valued the health states with the use of the AP according to the resource utilisation and utilities of health states with usual care (without the AP). Hence, the assumption in our model was that the AP was not associated with any adverse events which affected the resource utilisation and utilities of the health states.

In our Markov model, the transition probabilities only depended on the impact of the AP and usual care on moving from the health state "no fluctuations" to "fluctuations". However, disease progression in Parkinson's disease may also depend on co-variables such as past clinical history and co-morbidity. Therefore, the extension of a Markov model to a so-called semi-Markov model, would have allowed a more realistic simulation of the disease, incorporating all relevant variables in real life.

## CONCLUSION

The ideal design to demonstrate the possible health outcomes and costs associated with a new drug would be a naturalistic prospective study. However, a prospective naturalistic study cannot always be performed for logistical and budgetary reasons. In addition, study drug are usually not approved for registration at the time of the economic studies are performed, which limits the use of a prospective naturalistic trial. "Piggy-back trials" have limited external validity, because they have strict inclusion and exclusion criteria and treatments are protocol driven, leading to overestimation of units of healthcare utilisation; on the other hand, the units of utilisation collected may not be complete. In addition, the units of healthcare utilisation are country specific and the time to follow-up of clinical trials is too limited for pharmacoeconomic purposes.

In these cases, decision-analytic models may provide some of the missing information. However, a main concern with modelling studies is the use of a panel of experts to gather data which can not be derived from the literature. There is no guarantee that their assessment of resource utilisation and utilities is an accurate reflection of reality. From the perspective that results from modelling studies are of interest to many parties with different responsibilities and diverging interests, it is obvious that potential bias due to the use of a Delphi panel should be minimised.

In this article, we presented the cross-sectional study as an alternative data source for a modelling study; data from such a study can be used to determine costs and utilities for Markov health states. The overall design may be considered a hybrid between a naturalistic prospective study and a modelling study by maximising the pros and minimising the cons of both types of design. This hybrid design is based on bridging probabilities derived from literature and clinical trials with information on costs and utilities from a cross-sectional study. This design was illustrated using a Markov model for Parkinson's disease.

### References


[1] Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Commonwealth Department of Human Services 1995

[2] Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. 2st ed. Ottawa: CCOHTA. 1997

[3] UK draft guidelines. 1998

[4] Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-265

[5] Nuijten M.J.C.. The selection of data sources for use in modelling studies. PharmacoEconomics, March 1998; 3(3): 305-316

[6] Fink A, Kosecoff J, Chassin M, et al. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984; 9: 979-83

[7] Williams P, Webb C. The Delphi technique; a methodological discussion. Journal of Advanced Nursing 1994; 19: 180-186

[8] Fearnley J, Lees A. Pathology of Parkinson's disease. In: Calne DB, editor. Neurogenerative diseases. Philadelphia: WB Saunders, 1994; 545-54

[9] Lane DA. Utility, Decision and Quality of Life. J. Chron. Dis, 1987; 40; 6; 585-591

[10] Torrance GW, Feeny D. Utilities and Quality Adjusted Life Years. Intl. J. of Technology Assessment in Health Care, 1989; 5; 559-575

[11] Torrance GW. Measurement of health state utilities for economic appraisal. J Health Economics. 1986; 5: 1-30
[12] Torrance GW. Utility approach to measuring health related Quality of Life. J. Chron. Dis. 1987; 40: 593-600
[13] Gafni A. The quality of QALYs: do QALYs measure what they at least intend to measure? Health Policy. 1989; 13: 81-83
[14] Mehrez A, Gafni A. Quality Adjusted Life Years, Utility theory, and Healthy Years Equivalents. Med. Decis. Making 1989; 142-149
[15] The EuroQol Group. EuroQol – A new facility for the measurement of health related quality of life. Health Policy. 1990: 199-208
[16] International Medical Publications (IMP) Kommunikation. Gelbe liste pharmaindex. München: IMP, 1995
[17] Bundesverband der pharmazeutischen industrie. Rote Liste 1995. Frankfurt: Cantor-Verlag, 1995
[18] Statistisches Bundesamt. Statistisches Jarhrbuch. Wiesbaden: Metzler-Poeschel-Verlag, 1995
[19] Mundenbruch R. Gebuhrenordnung fur Artze, GOÄ, BG-GOÄ. Dachau: Zauner Druck und Verlags GmbH, 1995
[20] Mundenbruch R. BMÄ, E-GO, EBM. Dachau: Zauner Druck und Verlags GmbH, 1995
[21] Dodel R, Singer M et al. The Economic Impact of Parkinsons's Disease: estimation based on a 3-month prospective analysis.. PharmacoEconomics, September 1998; 14 (3): 299-312
[22] Beck J.R., Pauker G.P. et al. A convenient approximation of life expectancy (The Deale). The American Journal of Medicine. 1982. 73: 889-897

Chapter 6

# Cost-Effectiveness Analysis of Interferon Beta in Multiple Sclerosis: a Markov Process Analysis


## SUMMARY

OBJECTIVE: The objective of this study was to examine the cost-effectiveness of preventive treatment with interferon beta (IFNB) versus no preventive treatment in patients with multiple sclerosis. METHODS: The setting for this study was the United Kingdom. A lifetime Markov process model was constructed to model the average quality-adjusted life years (QALYs) and the costs of both treatment strategies. Data for the construction of the model came from published literature, including large multicenter randomised clinical trials in relapsing-remitting and secondary progressive multiple sclerosis. Costs were obtained from published sources. RESULTS: The results of the baseline analysis from the National Health Service (NHS) perspective showed that the use of interferon beta as preventive treatment for MS increased the total average discounted cost from £51,214 to £221,436 per patient. The undiscounted effectiveness increased from 24.9 QALYs to 28.2 QALYs, resulting in an incremental cost-effectiveness ratio of £51,582 per QALY. Sensitivity analyses showed the robustness of this model for other interferons. CONCLUSION: The study showed that preventive treatment with interferon beta in patients with multiple sclerosis may not be fully justified from a health-economic perspective, although interferon beta is associated with an improved effectiveness compared with no preventive treatment.

# INTRODUCTION

## Epidemiology

Multiple sclerosis (MS) is the second most common cause of neurological disability in young and middle-aged adults.[1 2] Females are more susceptible by a factor that approaches 2:1 in population studies, but this varies among surveys.[1] Males are more likely to have progressive disease from onset.[3 4] The mean age of onset of MS is during the third and fourth decades of life with a peak incidence during the late 20's - early 30's.[5 6]

Most patients initially have the relapsing-remitting form (RRMS) with a variable frequency of exacerbations (mean 1-2 per year). At onset, 65% of patients fall into the RRMS category.[3 6] Eventually most RRMS patients will develop the secondary progressive form (SP). In this stage, fewer exacerbations occur, and recovery from them is always incomplete. Eventually no recovery is possible. In addition, a chronic and slow increase in neurological deficits takes place. Thus, the patient suffers from increasingly severe disability.[7-10]

## Quality of Life

It is generally recognized that MS can dramatically affect the quality of life (QoL) experienced by an afflicted patient and/or his/her family. Many MS patients have a normal life span and have to live with some degree of disability over a prolonged time period. Family life, economic status and social interaction may be affected by somatic symptoms of the disease.[11] Cognitive dysfunction affects 43% to 65% of MS patients and also has a major negative influence on QoL in MS patients. Cognitively impaired patients are less likely to be professionally active, are more dependent, report more sexual dysfunction and tend to be less socially engaged than cognitively intact MS patients.[12]

## Economic Impact of Multiple Sclerosis

The full economic cost of MS to society and to the individuals concerned is uncertain, but is most likely to be substantial, bearing in mind that: (i) MS patients experience a major perturbation in their daily activities; (ii) MS affects mainly young productive people between the ages of 25 and 40 [1] who are obliged to interrupt their professional activities either temporarily or permanently.[13]

A positive correlation between total health-care costs and the Expanded Disability Status Scale (EDSS) has been demonstrated among MS patients. Above an EDSS scores of 5.5 (needs aid to walk), costs increased particularly quickly.[13]

Few economic studies of MS have been published. A prevalence-based cost of illness study in the UK (1995) estimated the annual burden of MS at £1.199 million.[14] The largest share was carried by the state, and the total cost included the following: state benefits, £287 million (23.9%); NHS costs (hospitalization, other treatment and support), £153 million-12.8%): lost tax revenue £148 million (12.3%). Annual NHS costs range from £336 to at least £4,275 per patient depending on their level of mobility which influences the number of hospital in-patients visits. The costs for drugs were

insignificant (£23/patient, (<2%) interferon not included). Lost earnings (33%), private expenses (11.7%) and employers' expenses (6.3%) were higher. The authors suspected that this total cost value was an underestimation in that diagnostic costs were not completely assessed and the use of average (general population) values for consultation time may not have been accurate for MS patients. In an earlier study conducted in England and Wales, the annual costs of the disease were estimated at £125.4 million (1986/87 prices). In this study, the largest component (20%) was the lost earnings of MS sufferer, although the cost burden of the family was not considered.[15] A more recent cross-sectional cost of care study was performed to assess the economic burden of multiple sclerosis (MS) in the UK, France and Germany.[16] Patients were stratified into three severity groups according to the EDSS: stages I, II and III, corresponding to mild (EDSS 1.0-3.5), moderate (EDSS 4.0-6.0) and severe (EDSS 6.5-8.0) MS respectively. From the societal perspective, the total cost of MS for three months was estimated at £ 3500, £ 4612 and £ 9989 per patient for stages I, II and III patients respectively in the UK. From the health-insurance perspective, the cost for three months was estimated at £ 535, £ 616 and £ 2020 in the UK per patient with stage I, II and III MS respectively.

Indirect costs tend to be the largest contributors to the overall cost burden. Taking into account employment history, medical insurance, amount and source of family income and disease progression, MS can cost an individual 40% of their lifetime earnings. [13 17 18] Recent studies have indicated that fifty to eighty percent of MS patients are unemployed within 10 years of disease onset. [19 20 21] As mentioned for the UK study, indirect costs related to lost earnings represented £395 million per annum or 33.0% of the total burden. Of this, 26.3% was attributed to lost earnings of non-professional caregivers. It has also been reported that 39% of men and 19% of women with MS retire early due to disability, with one of the most frequently cited reasons being MS-related fatigue.[22] Cognitive impairment, spasticity, perturbation of co-ordination and disturbances of bladder and bowel functions, non-remittent disease course, heavy physical work and age over thirty may be factors contributing to early retirement or unemployment.[20 23 24]

A previous cost-effectiveness study was performed for preventive treatment with interferon beta in patients with initial RRMS.[25] This study was based on a modelling design following patients over a period of five and ten years. The study showed that the additional cost of INFB of £ 43,400 is offset by gains in QALYs of 0.13, indicating a cost-utility ratio of £ 328,300 per QALY gained. The 10-year analysis produced a slightly more favorable cost-effectiveness ratio of £ 228,300. This study could not take into account the potential reduction in disease progression associated with use of interferon beta in patients who had already progressed to SP because clinical data documenting the favorable effect of interferons in SP were not yet available.

## Treatment of Multiple Sclerosis

We will focus on the prevention of disease activity by interferons. The rationale for such therapy is that MS is generally thought of as an autoimmune disease. Interferons have, by definition, antiviral

properties, but they also have immunomodulating and antiproliferative properties. The classification of interferons as IFNα, IFNβ, and IFNγ, is based on their antigenicity and the similarities of their nucleotide sequences. Two types of recombinant IFNβ exist, IFNβ-1a and IFNβ-1b. IFNα and IFNβ are used in the treatment of MS because of their immunomodulatory capacity to suppress the activated immune system. A major effect is thought to occur through the blocking of IFNγ effects, which activate the immune system.[26 29] A major breakthrough in the treatment of RRMS patients is evident in the recent results of clinical trials in which recombinant IFN beta (IFNβ-1b or IFNβ-1a) was shown to significantly reduce the relapse rate.[20 30 31] IFNβ-1b was the first of the interferons tested in a large-scale clinical trial in RRMS and SP.

The IFNβ-1b results in RRMS were based on pooled results from three-year, multicenter, randomized, double blind, placebo-controlled clinical trials each with three parallel treatment groups (placebo vs. 1.6 MIU vs. 8 MIU IFNβ-1b).[30] All patients in the study were given the option of continuing treatment in a double-blind fashion, extending the total treatment period to 5.5 years for a proportion of patients. Subsequently, two studies with IFNβ-1a were performed in RRMS patients.[20 32]

1. The results of all interferon beta trials are comparable.[30 32] All of the IFNβ's reduce the relapses in MS.
2. The IFNβ's have a positive effect on delaying progression in disability over the two to three year trial period.
3. These effects of IFNβ's are dose related. Debate over dosage exists, and although the approved dosage should be given; the highest tolerable dose is preferred.

The IFNβ-1b results in SP were based on a three-year, multicenter, randomized, double blind, placebo controlled clinical trial with patients receiving either 8 MIU interferon or placebo.[31] This was the first clinical study that showed that interferons might delay disease progression in patients with SP.

In the event that the favorable effect of interferons on the frequency of acute exacerbations and disease activity may have an impact on cost-effectiveness, the present study uses a modelling technique to compare the cost-effectiveness of interferon beta with no preventive treatment in patients with MS in the UK. The hypothesis is that interferons will lead to a long-term reduction in disease activity, including acute exacerbations and disease progression (change in disability over time defined as change in EDSS), and a decrease in morbidity leading to an increased Quality of Life, which will offset the expected increased drug costs.

## METHODS

To estimate the costs and effectiveness of preventive treatment with interferon beta in patients with MS versus usual care in the health-care setting of the UK, a lifetime model was constructed using decision analysis techniques.[33 34] The main analytical plan for the study was an incremental cost-effectiveness analysis. This approach consists of combining cumulative measures of costs over time

with a cumulative measure of effectiveness, resulting in incremental costs per clinical benefit gained (e.g., cost per life year gained). The analysis was performed for a hypothetical cohort of 30-year old female patients with initial RRMS, i.e., patients whose clinical picture was characterized by periods of exacerbation (relapse) from which recovery was complete or partial with subsequent mild disability. For a more specific measure of impairment, the Expanded Disability Status Scale (EDSS) was used.[35] In general, the inclusion criteria used in the clinical trials of interferon beta were used. Effectiveness was expressed as Quality Adjusted Life Years (QALYs). The model also included all relevant economic measures such as resource utilization patterns associated with outpatient and inpatient care for the treatment of MS, indirect costs (working days lost) and direct, non-medical costs.

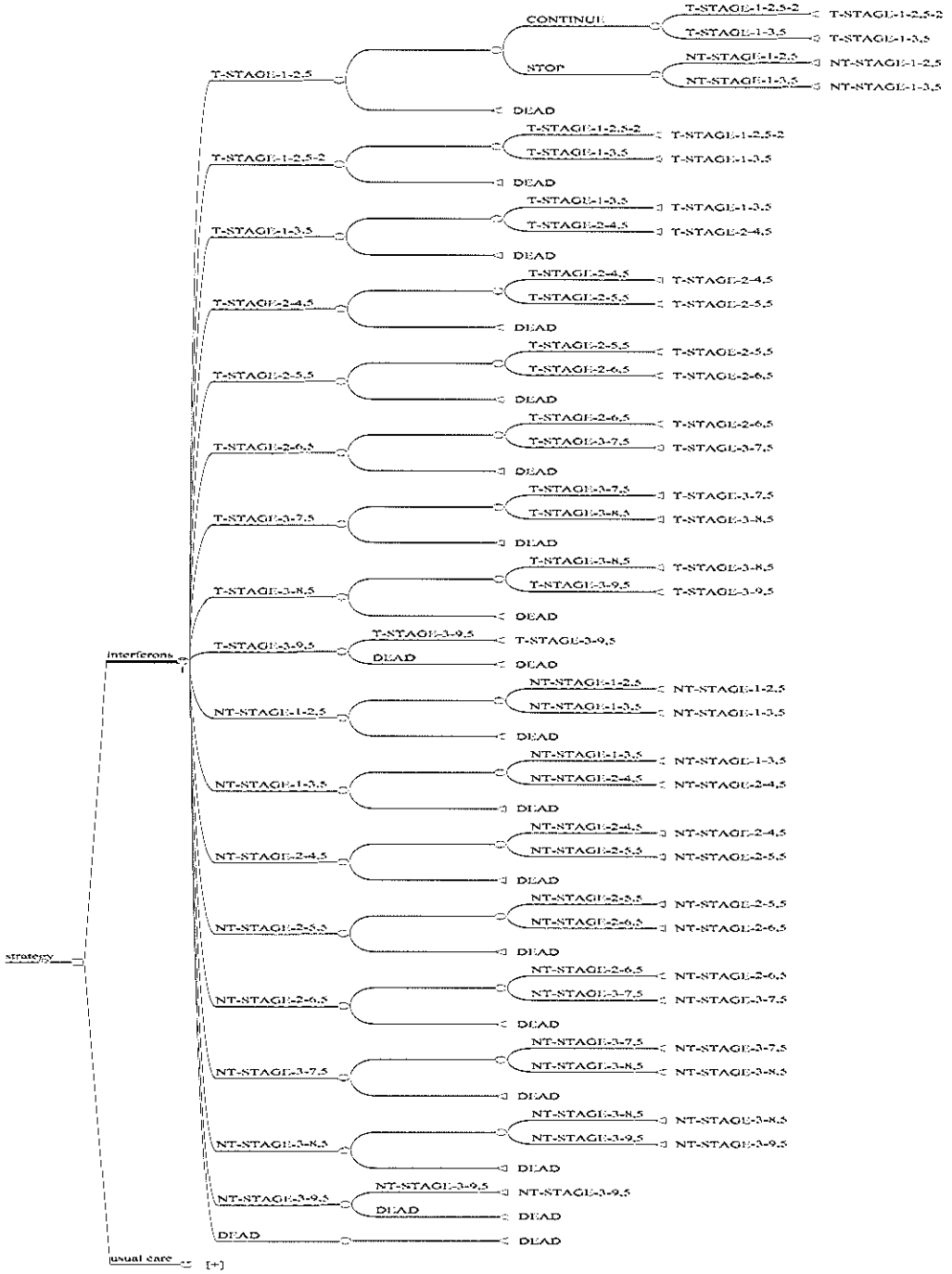
The setting of the study was that of the UK health-care system of 1998. The primary perspective of the study was that of the NHS according to the NICE guidelines [36], while the secondary perspective was that of the society in 1998. Costs used in the model were expressed in 1998 sterling pounds (£), and study costs were adjusted to 1998 costs from 1996 costs by using a 4.3% inflation correction.[37] The cost valuation was based on the direct health-care costs, direct nonmedical costs and indirect costs. The costs were discounted by 6% from the second year onwards, and no discounting was applied in the first year. The data sources were literature, including clinical trials of interferon beta and other interferons, and official price and tariff lists. An external opinion leader validated the methodology (model structure and assumptions).

## Description of the Model

Markov-process analysis techniques were used to model the clinical and economic outcomes accrued over a lifetime with or without preventive treatment. An advantage of the Markov process model is that it allows modelling of MS disease progression beyond the follow-up of the clinical trials.

Figure 1 shows the structure of the Markov model for MS. The model structure for the usual-care treatment arm is identical. The first branch point on a tree is called a decision node because it corresponds to a choice of treatment — interferon beta or usual care. A decision node is represented as a small square (□). Subsequent to the decision node, the structure of the semi-Markov process model is shown, and is identical for both treatment options. The other branch points indicate allowed transitions. The cycle time chosen for the model was three years, which closely approximates the follow-up period for the interferon beta clinical trials in RRMS and in secondary progressive MS.

**Figure 1. Simplified version of the semi-Markov model for the base case scenario. Only the sub-tree after interferon beta is shown for feasibility purposes.**



The follow-up time (analytical horizon) used in the model was the remainder of the lifetime after diagnosis of RRMS. For this analysis we defined mutually exclusive Markov states. Initially we defined Markov health states corresponding with severity levels of MS:

- Stage I: EDSS ≤ 3.5
- Stage II: EDSS 4 to 6.5
- Stage III: EDSS 7 to 9.5
- DEAD: EDSS = 10

Because the primary outcome variable in most clinical trials was defined as deterioration from baseline by at least 1.0 point on the EDSS for at least three months, the health states indicated above have been further categorized to allow for the incorporation of clinical trial results in the model: Stage I-2.5 corresponds with an average EDSS of 2.5; Stage I-3.5 with an average EDSS = 3.5, etc. Finally DEAD corresponds with an EDSS of 10.

In the model, all patients start in the health state "Stage I-2.5", which corresponds to the average EDSS of patients in the interferon beta clinical trial at inclusion. After one cycle of three years there are two possibilities: a patient will stay in "Stage I-2.5" if the EDSS score has not changed by more than 1; a patient will move to "Stage I-3.5"after a worsening of the EDSS score of more than 1.

In addition, there will be a number of patients on preventive treatment who drop out during the first cycle period due to adverse events, which means that they will receive no treatment for the remainder of the follow-up period in the model. Stage I-2.5 has been divided in Stage I-2.5-1 and Stage I-2.5-2 to reflect non-compliance during the first cycle (Stage I-2.5-1) and full compliance during subsequent cycles (Stage I-2.5-2).

Therefore the health states were further categorized into Treatment (T) or No Treatment (NT), as shown in figure 1.

## Key Model Assumptions

There are a number of key assumptions upon which the model is based. A three-year cycle time was used, because this interval closely approximates the follow-up period of the interferon beta clinical trials for preventive treatment in RRMS and secondary progressive MS.

## Follow-Up

A study by Runmarker et al.,[38] showed that only 20% of the patients progressed to SP after 25 years. Therefore, we decided to construct a lifetime model, which allowed us to simulate the reduction in disease progression by interferon beta in real life more realistically.

In the model, all patients start in the health state Stage I-2,5 with an average EDSS of 2.5, which corresponds with the average EDSS of the patients in the beta interferon clinical trial in RRMS (IFNβ-1b 2.4, placebo 2.3). Hence, the model is based on a hypothetical patient with an EDSS-score of 2.5.

The model was based on an analysis of high-dose interferon beta, as prospective clinical studies have been performed in RRMS as well as in SP.

Follow-up for estimating the effect of preventive treatment with interferon beta in RRMS and secondary progressive MS are available for up to three years. Follow-up data for the no preventive treatment arm came from follow-up data for the placebo arms of interferon clinical trials. Disease progression is based on the definitions which are used in the clinical trials: worse = change in EDSS>1; stable = no change in EDSS.

### Disease Progression

- Stage I, EDSS < 4.5. Disease progression was based on the IFNβ-1b clinical trial in RRMS (mean EDSS at inclusion = 2.4);
- Stage II, EDSS 4.5 to 7.5. Disease progression was based on the IFNβ-1b clinical trial in secondary progressive MS (mean EDDS at inclusion = 5.2);
- Stage III, EDSS 7.5 to 9.5. We assumed that beyond the actual follow-up period for the trial of IFNβ-1b in secondary progressive MS, all treatment strategies have the same rate of disease progression and use the disease progression for placebo from clinical trials as a state transition matrix in the Markov process. Hence, although the disease progression can differ during the actual follow-up period of the clinical trials as a result of the treatment, progression beyond this period is assumed to be the same whether patients received preventive treatment or not, although this may be an underestimation of the effectiveness of IFNβ-1b.
- EDSS 9.5 until death. We assumed that there was no difference in mortality between patients with or without preventive treatment.

Noncompliance was derived from the IFNβ-1b trial in RRMS only. We assumed that noncompliance occurred only during the first cycle of our model and was due only to adverse events. Consequently, noncompliance data from the interferon beta clinical trial in SP was not used as patients in the IFNβ-1b clinical trial in SP started treatment with IFNβ-1b which does not correspond with patients in our model who had already proven to be compliant.

### Acute Exacerbations

- Stage I = EDSS < 4.5. The incidence of acute exacerbations was based on the IFNβ-1b clinical trial in RRMS.
- Stage II = EDSS 4.5 to 7.5. The incidence of acute exacerbations was based on the IFNβ-1b clinical trial in SP.
- Stage III = EDSS 7.5 to 9.5. We assumed that SP patients in stage III did not experience exacerbations anymore, and consequently assumed no benefit in terms of preventing acute exacerbations with IFNβ-1b beyond the actual follow-up period of the interferon beta clinical trials.

The initial cohort in the model consisted of 30–year-old women, based on literature data. The mean age of onset for MS is during the late 20's and early 30's, and females are more susceptible by a factor 2:1.[1]

Age-specific population mortality rates for women were used, assuming that MS did not alter life expectancy. Literature did not yield evidence for an increased mortality risk due to MS.

### Clinical and Economic Outcomes

#### *Effectiveness*

The effectiveness measurement was based on the concept of utility, which measures the QALY (Quality Adjusted Life Years).[39 40]

#### *Cost Assessment*

The cost assessment was based on the assignment of costs to the health states associated with a cycle time of three years. The costs of each health state were determined by the resource utilization associated with each health state: medical resource utilization (e.g., medication, consultations) and non-medical resource utilization. Non-medical resource utilization included workdays lost, transportation, community assistance and home modifications.

## DATA SOURCES

Three different types of data can be distinguished in modelling studies:

- Probabilities of clinical events, which generally are out of the control of the physician, e.g., probability of an acute exacerbation, disease progression;
- Utilities of different Markov health states;
- Costing information derived from estimates of the units of resource utilization and their prices/tariffs (product of unit and price).

The probabilities of clinical events were based on interferon beta clinical trial data, while the utilities and costs were derived from two cross-sectional studies. The prices and tariffs were derived from official lists.

**Table 1. Clinical data and sources.**

| RRMS | Probability per cycle | Reference |
|---|---|---|
| From EDSS = 2.5 to EDSS= 4.5 | | |
| Continuation after first cycle INFB 8 MIU | 0.920 | P. Duqette et al. Neurology 1993 43: 655-661 [29] |
| Progression placebo | 0.28 | Idem |
| Progression INFB 8 MIU | 0.20 | Idem |
| Progression interferons | | |
| minimum* | 0.193 | PRIMS Study Group. Lancet 1998:352:1498-504 [20] |
| maximum* | 0.219 | Idem |
| **SP** | **Probability per cycle** | |
| From EDSS = 4.5 to EDSS = 7.5 | | |
| Progression placebo | 0.50 | European Study Group (mean 5.2) [30] |
| Progression interferon INFB 8MIU | 0.39 | Idem |
| From EDSS = 8.5 to death | 0.31 | Weinshenker BG et al. Neurology. 1996:46;1613-19 [3]. |
| **EXACERBATIONS** | **Annual frequency** | |
| From EDDS = 2.5 to EDSS = 4.5 | | |
| Placebo | 1.21 | P. Duqette et al. Neurology 1993;43:655-61 [29] |
| Interferon | 0.84 | Idem |
| Difference | 0.37 | |
| From EDDS = 4.5 to EDSS = 7.5 | | European Study Group (NB (mean 5.2) [30] |
| Placebo | 0.64 | |
| Interferon | 0.44 | |
| Difference | 0.20 | |
| From EDDS = 8.5 to Death | 0 | Assumption |

## Probabilities

Probabilities were derived from published literature (Table I).

- Stage I: EDSS = 2.5 to 4.5: probabilities of disease progression and incidence of acute exacerbations derived from the interferon beta clinical trial in RRMS.
- Stage II: EDSS = 4.5 to 7.5: probabilities of disease progression and incidence of acute exacerbations derived from the interferon beta clinical trial in secondary progressive MS.

- Stage III: EDSS 8.5 to death: probabilities of disease progression derived from a meta-analysis of the placebo-treated groups in clinical trials in progressive MS.[17]
- Age-specific population mortality rates for women derived from national statistics data.[22]

## Utilities

The QALYs for the different health states were derived from a cross-sectional study, and are shown in Table II.[25] The classification of disease severity in this study corresponds with semi-Markov health states in our model. We also incorporated a temporary utility loss per relapse, which was derived from the same study — a reduction of 0.5 utility over a period of one month.

**Table II. Utilities for health states. [Health Technology Assessment 1998 (25)]**

| **Health state** | | **Utility** |
|---|---|---|
| EDDS utilities | Stage | Fixed input value |
| EDDS-3 | I | 0.71 |
| EDDS-4 | II | 0.66 |
| EDDS-5 | II | 0.52 |
| EDDS-6 | II | 0.49 |
| EDDS-7 | III | 0.35 |
| EDDS-8* | III | 0.17 |
| EDDS-9* | III | 0.08 |
| DEAD | | 0.00 |
| | | Utility loss |
| Relapse | | 0.5 |

## Costs

Data on costs were derived from another cross-sectional study (Table III).[16] This study used a retrospective approach, in which resource utilization and clinical data were collected at a single time-point and covered the three months period prior to the dates of inclusion.

The cost of care of MS was calculated for patients in current clinical practice in the UK in 1996, which did not include costs of preventive treatments like interferons. Therefore the cost of interferon beta was added to health states corresponding to preventive treatment ("T-health states").

The costs due to relapses were not separately presented in the cost of care study. Therefore we subtracted the costs of avoided exacerbations from the costs of health states that included preventive treatment with interferon beta. The direct medical costs due to an exacerbation were derived from the cost-effectiveness study in initial RRMS.[25]

**Table III. Costs for health states, excluding cost of interferon. [Murphy et al., 1998 (15)]**

| Health state | Costs (£) | | |
|---|---|---|---|
| Perspective | mean | median | Reference |
| Societal | | | Murphy 1998 |
| - Stage I | 3,500 | 2,829 | |
| - Stage II | 4,612 | 3,266 | |
| - Stage III | 9,989 | 6,906 | |
| Health insurance | | | |
| - Stage I | 535 | 361 | |
| - Stage II | 616 | 348 | |
| - Stage III | 2020 | 682 | |
| | | | Health Technology Assessment 1998 |
| Cost per relapse | 2,115 | | |
| Annual costs INFB | 10,500 | | |

The friction cost method was applied to the evaluation of workdays lost by active patients. According to this method, the value of productivity loss was assumed to be 80% of the average value of a worker's productivity during the "friction period". Thereafter, it was assumed that sick employees could be replaced. Time lost by inactive patients was considered as leisure time lost and was valued at 40% of the average wage in the UK.[41] The replacement salary method (i.e. estimation of the wage that would be paid to a home-helper to perform the work that the patient was unfit to do) was applied to the valuation of care-giving time.[42] We assumed that the number of working days lost associated with a relapse corresponded with relapse length, which was derived from the same study.

## ANALYSES

The base-case analysis represents the expected average effectiveness and costs per patient discounted at 6%, while undiscounted costs were also included. An incremental cost-effectiveness analysis representing the additional cost and effectiveness obtained when a preventive treatment with interferon beta is compared to no treatment was also performed.

Univariate sensitivity analyses were based on the modification of the basic clinical and economic assumptions in the clinical outcome model to test the stability of the conclusions of the analysis over a range of assumptions, probability estimates and value judgments. Sensitivity analyses were performed only for the primary perspective of this study, that of the NHS.

The first sensitivity analysis was performed to assess the impact of relapses on the analysis by excluding the clinical and economic consequences of relapses. We performed a second analysis to assess the sensitivity of the analysis to the annual costs of interferon beta. The bounds for this analysis were based on annual costs of other interferon treatments. A third sensitivity analysis was performed for the cost of a relapse. The costs per relapse in the study by Health Technology Assessment included both ambulatory treatment and hospitalization. As this input variable had a wide range of values, the model was potentially highly sensitive to it.[25] A fourth sensitivity analysis was performed on the costs of the health stages derived from the study by Murphy [16] (Table II) by using median costs of all three health stages instead of the mean costs. The fifth sensitivity analysis involved constructing an interval around the rate of disease progression. The bounds for this analysis were based on the rates of disease progression in RRMS, which have been published for other interferons. The rates were adjusted to transition probabilities corresponding to a three-year cycle time.

The three previous sensitivity analyses were performed from the NHS perspective only. We also performed a sensitivity analysis on a 3% discount rate from the societal perspective according to VH's guidelines.

## RESULTS

Based on the probabilities, utilities and the costs of the direct health-care utilization described in Tables I, II and III a base case-analysis was performed to obtain an expected value for the average total costs and the average QALY. Tables IV and V show the economic and effectiveness outcomes of the base case analysis for preventive treatment with interferon beta and no preventive treatment in MS from the third-party payer and societal perspectives respectively. The total average discounted cost per patient for the preventive treatment of an MS patient starting from RRMS was £221,436 for the interferon beta group versus £51,214 for the no treatment group (discounted at an annual rate of 6%). A break down of the costs shows that the higher costs of the interferon group are attributed to the cost of interferons (£179,367), while the other costs were £51,214 and £42,069 for the no treatment group and interferon group respectively. Based on the model, the average QALY of the interferon group was 28.2 years vs. 24.9 years for the no treatment group (Table IV), a gain in QALYs of 3.3 years. The differences in cost and effectiveness resulted in an incremental cost-effectiveness for interferon beta of £51,582 per QALY gained.

**Table IV. Results of cost-effectiveness: health insurance.**

| | $INFBCE_F$ | Usual care $CE_M$ | Difference |
|---|---|---|---|
| Undiscounted Costs (£) | | | |
| *INFB* | 498,505 | 0 | |
| *Other* | 161,475 | 210,824 | |
| *Total* | 659,980 | 210,824 | 449,156 |
| Discounted Costs (£)* | | | |
| *INFB* | 179,367 | 0 | |
| *Other* | 42,069 | 51,214 | |
| *Total* | 221,436 | 51,214 | 170,222 |
| Effectiveness | | | |
| QALYs | 28.2 | 24.9 | 3.3 |
| Cost per QALY gained: | | | |
| Discounted costs and QALYs | | | 51,582 |

Table V shows the results of an analysis from the societal perspective. The total average discounted cost per patient for preventive treatment of an MS patient was higher in the interferon beta group at £473,115 versus £322,499 for the no treatment group, discounted at 6% per annum. The use of interferons substantially reduced the other costs from £322,499 to £293,748.

**Table V. Results of cost-effectiveness analysis from societal perspective.**

| | $INFBCE_F$ | Usual care $CE_M$ | Difference |
|---|---|---|---|
| Undiscounted Costs (£)* | | | |
| *INFB* | 498,505 | 0 | |
| *Other* | 1,032,212 | 1,209,409 | |
| *Total* | 1,530,717 | 1,209,409 | 321,308 |
| Discounted Costs (£)* | | | |
| *INFB* | 179,367 | 0 | |
| *Other* | 293,748 | 322,499 | |
| *Total* | 473,115 | 322,499 | 150,616 |
| Effectiveness | | | |
| QALYs | 28.2 | 24.9 | 3.3 |

## Sensitivity Analysis

Sensitivity analyses were performed on the main probabilities and cost assumptions to test the robustness of the cost-effectiveness results. Table VI shows the results of the sensitivity analyses. The first two sensitivity analyses were performed to assess the impact of discounting. They show that the

incremental cost-effectiveness ratio is quite sensitive to discounting of both economic and clinical outcomes. A sensitivity analysis was also performed to assess the impact of relapses on the incremental cost-effectiveness ratio. When the cost of relapse and the utility loss due to relapse were excluded, the incremental cost-effectiveness of interferon beta was £38,222 compared to the base case result of £51,582, which shows that the model is moderately sensitive to the inclusion of relapses. A sensitivity analysis was also performed on the annual cost of IFNβ-1b. When the annual cost is varied from £ 6000 to £ 12,000, the incremental cost-effectiveness ratio changes substantially from £28,280 to £59,348. When the medical cost of a relapse is varied between its lower and higher range (£1,000 to £3,000), the incremental cost-effectiveness ratio changes only from £47,028 to £50,275. We also performed a sensitivity analysis on the disease progression with interferon beta. When this was varied between minimum and maximum values, the incremental cost-effectiveness ratio increased from £43,520 to £56,756. Finally a sensitivity analysis was performed on a 3% discount rate for both costs and effectiveness, which shows that the use of interferon beta increased the costs from £510,940 to £701,410, while the QALYs increased from 24,98 to 28,12.

**Table VI. Results of sensitivity analysis.**

| Variable | Range of values | Cost per QALY gained |
|---|---|---|
| Health insurance perspective | | Discounted |
| Basis | | 51,582 |
| No discounting of costs | | 136,108 |
| Discounting of effectiveness (2%) | | 113,108 |
| Excluding relapses | | 38,222 |
| Annual costs INFB (£) | 6,000-12,000 | 28,280–59,348 |
| Medical relapse cost | 1,000-3,000 | 47,028–50,275 |
| Disease progression | RMMS: 0.193 – 0.219SP: 0.351 – 0.429 | 43,520–56,756 |
| Society perspective | Extra costs (£) | QALYs gained |
| Discounting of costs and effectiveness at 3% | 199,470 | 3,14 |

## DISCUSSION

This study examined the cost-effectiveness of preventive treatment with interferon beta compared to no preventive treatment of patients with MS in a British setting. The analysis revealed that interferons were much more cost-effective compared with the results of a previous cost-effectiveness study, which was performed in patients with RRMS only.[25] The incremental cost-effectiveness ratio in our study was £51,582 per QALY, while the outcomes in the other study were £328,300 and £228,300 per QALY over a period of 5 and 10 years respectively. This substantial reduction in the cost-effectiveness ratio of interferon beta may be due to a longer follow-up period in our study, including

continuation of treatment with interferon beta in SP. Sensitivity analysis showed that our model was especially sensitive to discounting, the cost of interferon beta and disease progression. The results of this study showed that cost-effectiveness of a preventive treatment with interferon beta in MS substantially improves, when the use in SP is incorporated in the health economic analysis.

The results of any modelling exercise need to be treated with some degree of caution. We have used various data sources for our model, which all have their pros and cons from a health economic perspective.[21] Among them is the fact that our literature review does not necessarily represent real clinical practice, as much of the literature examined was based on data from clinical trials. Data from clinical trials do not necessarily have a high degree of external validity, as the results are often contingent upon protocol adherence, a situation that may not be easily reproduced outside the trial setting. The utilities in our model were derived from a cross-sectional study based on clinical treatment practice without the use of preventive treatment with interferon beta. The utility of a health state is a function of both underlying disease (e.g. severity) and adverse events associated with treatment. In our model we valued the health states only according to severity levels corresponding to the EDSS severity scale regardless of the use of preventive treatment with beta interferon. Hence, the assumption in our model was that treatment with interferons did not affect the utility of the health states. Consequently the utilities did not take into account a potential lower utility of health states associated with interferon treatment due to adverse events. Finally we did not include the costs associated with treatment of adverse events resulting from treatment with interferon beta, which may have underestimated the total costs for preventive treatment.

Modelling of disease progression was constrained by published clinical trial data. Those data were presented as the percentage of patients staying at the same EDSS level or moving to a worse EDSS level, which may have involved more than one transition. However the published data did not show more detailed data, which may be due to statistical constraints: interferon may indeed reduce the number of patients progressing more than one EDSS level, but the statistical significance of this effect may be more difficult to prove.

Finally the study focused mainly on the cost-effectiveness of IFNβ-1b 8 MIU only, demonstrating robustness of the outcomes of the model when using clinical trial data from other interferons. A meta-analysis based on pooled results of the various interferon trials may have lead to more reliable input data for the model. However this approach was not feasible for a number of reasons: there was only one trial in SP that showed that interferon beta reduces disease progression; dose relationships did not allow us to cluster the different interferons and determine average rates; clinical trials did not have similar follow-up. The IFNβ-1b trial had a follow-up period of three years, while other trials had a follow-up of only two years; populations in the trials in RRMS were not similar, which is shown by different placebo outcomes as well as initial EDSS scores.

Comparing the modeled disease progression for no preventive treatment with real life data from a study by Runmarker et al.,[38] quantitatively validated our model, which did not include preventive treatment with interferons. After 24 years, approximately 18% of the patients in our model progressed to SP. This outcome corresponds with findings in the study by Runmarker et al.,[38], where 20% progressed to SP after 25 years. A limitation in the study by Runmarker may be the small number of patients and the inclusion of patients with progressive MS, which may explain the slower disease progression in our model.

## CONCLUSION

In conclusion, the study showed that preventive treatment with interferon beta in patients with multiple sclerosis may not be fully justified from a health-economic perspective, although interferon beta is associated with an improved effectiveness compared with no preventive treatment. Development of clinical guidelines may be used to optimize the cost-effectiveness outcomes and budgetary constraints. A suitable approach is the prospective collection of health economic data and utilities in a naturalistic setting to validate the preliminary findings of this model and generate guidelines based on real life data.

**References**


1. Hauser SL. Multiple sclerosis and other demyelinating diseases. In: Harrisons's Principals of Internal Medicine, 1994. Isselbacher KJ, Braunwald E, Wilson JD, et al., eds. New York: McGraw-Hill
2. Noseworthy JH, Lucchinetti C, Rodriguez M, Weinshenker BG.. Medical Progress: MultipleSclerosis. New Eng J Med 2000. 343: 938-52
3. Weinschenker B, Bass B, Rice GP, Ebers GC. The natural history of multiple sclerosis: A geographically-based study. I. Clinical course and disability & II. The predictive value of the early clinical course. Brain 1989;112:133-46
4. Weinschenker B, Bass B, Rice GP, Ebers GC. The natural history of multiple sclerosis: A geographically-based study. II. The predictive value of the early clinical course. Brain 1989;112:1419-28
5. Rieumont MJ, Deluca SA. Neuroimaging evaluation in multiple sclerosis. Am Fam Physician 1993;48:273-6
6. Hibberd PL. The use and misuse of statistics for epidemiological studies of multiple sclerosis. Ann Neurol 1994;36:S218-30
7. Clark W (1996). Therapeutic advances: beta-interferon for multiple sclerosis. J Clin Pharm Ther. 1996;21:195-9
8. Goodkin D, Hertsgaard D, Rudick R. Exacerbation rates and adherence to disease type in a prospectively followed-up population with multiple sclerosis. Arch Neurol 1989;46:1107-12
9. Weatherall DJ. Oxford textbook of medicine, 2nd edn., Oxford: Oxford Medical Publications, 1987
10. Weinshenker BG (1995). The natural history of multiple sclerosis. Neurol Clin;13:119-46
11. Rudick RA, Miller D, Clough JD, et al. Quality of life in multiple sclerosis. Arch Neurol 1992;49:1237-42.
12. Rao SM, Leo GJ, Ellington L, et al. Cognitive dysfunction in multiple sclerosis. II Impact on employment and social functioning. Neurology 1991;41:692-6
13. Catanzaro M, Weinert C. Economic status of families living with multiple sclerosis. Int J Rehabil Res 1992;15:209-18
14. Holmes J, Madwick T, Bates D. The cost of multiple sclerosis. British Journal of Medical Economics 1995;8:181-93
15. O'Brien B. Multiple sclerosis. London: Office of health economics, 1987
16. Murphy N, Confavreux C, Haas J. Economic Evaluation Of Multiple Sclerosis in The UK, Germany and France. Pharmacoeconomics 1998;13:607-622
17. Weinfeld FD, Baum HM. The national multiple sclerosis survey: background and economic impact. Ann N Y Acad Sci 1984;436:469-71

18. Weinshenker BG, Issa M, Baskerville J. Meta-analysis of the placebo-treated groups in clinical trials of progressive MS. Neurology 1996; 46:1613-9
19. LaRocca N, Kalb R, Kendall P, Scheinberg L: The role of disease and demographic factors in the employment of patients with multiple sclerosis. Arch Neurol 1982;39:256
20. PRIMS Study Group. Randomised double-blind placebo-controlled study of interferon beta-12a in relapsing/remitting multiple sclerosis. Lancet 1998;352:1498-504
21. Nuijten MJC. Data management in modelling studies: the selection of data sources. Pharmacoeconomics 1998;3:305-16
22. Office for National Statistics, Earnings and Employement Division; New earnings survey, 1999
23. Bauer HJ, Firnhaber W, Winkler W. Prognostic criteria in multiple sclerosis. Ann. N Y Acad Sci 1965;122:542-51
24. Grönning M, Hannisdal E, Mellgren SI: Multivariate analyses of factors associated with unemployment in people with multiple sclerosis. J Neurol Neurosurg Psychiatry 1990;53:388-90
25. A cost-utility analysis of interferon beta for multiple sclerosis Health Technology Asssesment 1998;2: 54-64
26. Cirelli R, Herne K, Tyring, SK. Interferons. An overview of their pharmacology. Clinical Immunotherapeutics 1996;5:S22-30
27. Reder AT. Clinical pharmacology of interferon beta-1b and its potential role in multiple sclerosis. Clinical Immunotherapeutics 1996;5:S31-40
28. Arnason BGW, Dayal A, Qu ZX, et al. Mechanisms of action of interferon-β in multiple sclerosis. Springer Semin Immunopathol 1996;18:125-48
29. Johnson KP. The historical development of interferons as multiple sclerosis therapies. J Mol Med 1997;75: 89-91
30. Duquette P, The IFNB Multiple Sclerosis Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. Neurology 1993;43:655-61
31. European study Group. Placebo-controlled multicenter randomised trial of interferon beta in treatment of secondary progressive multiple sclerosis. The Lancet 1998;352:1491-7
32. Jacobs LD, Cookfair DL, Rudick RA et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Ann Neurology 1996;39:285-93
33. Weinstein M, Fineberg H. Clinical decision analysis. London: W.B. Saunders Company, 1980
34. Holloway CA. Decision making under uncertainty models and choices. Englewood Cliffs, NJ: Prentice Hall Inc,1979
35. Kurtzke JF. Clinical definition for multiple sclerosis treatment trials. Ann Neurol 1994;36:S73-9
36. Nice guidelines. NHS National Institute for Clinical Excellence (NICE). Guidance for Manufacturers and Sponsors. http://www.nice.org.uk. First accessed June 2001
37. UK Hospital and Community Health Services Inflation Index (HCHS), 1998/1988-1998/1999 & CPI 1999/2000
38. Runmarker B, Andersson C, Oden A, Anderson O. Prediction of outcome in multiple sclerosis based on multivariate models. J Neurol 1994;241:597-604
39. Lane DA. Utility, Decision and Quality of Life. J Chronic Dis 1987;40:585-91
40. Torrance GW, Feeny D. Utilities and Quality Adjusted Life Years. Int J Technol Assess Health Care 1989;5:559-75
41. Office for National Statistics, Earnings and Employement Division; New earnings survey, 1995
42. Cook J, Richardson J, Street A. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. Health Econ 1994;3:157-68

Chapter 7

# The Incorporation of Potential Confounding Variables in Markov Models


## SUMMARY

The data in modelling studies may come from a variety of sources and are subject to varying degrees of uncertainty. A main concern is the use of a panel of experts to gather data, which may be replaced by information on costs and utilities from a cross-sectional study for the various health states in a Markov model. However the costs and utilities may not be only a function of the defined health states, but also of other explanatory variables, which may act as confounding variables when they are not taken into account. Hence the external validity of Markov models may be limited, and consequently the results of the model are not an accurate reflection of reality. In this manuscript we presented a strategy to improve the quality of the methods used in Markov modelling studies by increasing the external validity. This strategy consists of an incorporation of an extra explanatory variable in the Markov health states by means of health state specific relationships between this explanatory variable and costs as well as time-dependent values of the extra explanatory variable. In addition we determined the relevance of the incorporation of an extra explanatory variable by means of various sensitivity analyses. The concepts were illustrated using a hypothetical Markov model for Parkinson's disease. The results showed that the outcomes of a health economic model may be severely biased, when a confounding effect of an extra explanatory variable is not taken into account, which proves the need for the incorporation of other explanatory variables in a health economic model.

## INTRODUCTION

A recent paper showed that there is growing evidence that health economic data is beginning to be used more widely by decision makers in the decision- making process for reimbursement of drugs.[1] Because in practice it is not always possible to derive information from scientifically sound prospective studies, health economic models are increasingly used to provide the necessary cost-effectiveness information.[2] This especially applies to chronic conditions such as Parkinson's disease, which otherwise might require prospective studies over periods varying from 5 to 10 years. In these cases models can be used to extrapolate clinical outcomes beyond the duration of the trial.[3]

If economic modelling is to play a fundamental role in healthcare decision-making, it is vital that the methods and data sources used in modelling studies are carefully scrutinised and refined. The data may come from a variety of sources and are subject to varying degrees of uncertainty. Data sources for the variables being used in a model may be clinical trials, literature (e.g. meta-analysis), databases, medical records, Delphi panels and official tariff lists (for healthcare utilisation). In a previous paper, we extensively discussed the advantages and disadvantages of the different kinds of data sources.[2] A main concern was the use of a panel of experts to gather data, which cannot be derived from the literature. There is no guarantee that such an assessment of resource utilisation and utilities accurately reflects reality. Considering that results from modelling studies are of interest to many parties with different responsibilities and diverging interests, it is obvious that potential bias due to the use of Delphi panels should be minimised. In a subsequent paper we described a strategy for substantially reducing the amount of data, which needs to be derived from a Delphi panel.[4] This may substantially improve the quality of the input data in a model, and consequently the outcomes on costs and effectiveness. An alternative data source for a modelling study was presented, namely a cross-sectional study, which can be used to yield costs and utilities for Markov health states. The overall combined design may be considered a hybrid between a naturalistic prospective study and a modelling study by maximising the pros and minimising the cons of both types of design. This hybrid design is based on bridging the probabilities derived from literature and clinical trials with information on costs and utilities from a cross-sectional study.

In our previous study, the health states in our Markov model for Parkinson's disease corresponded with severity levels according to the existence of motor fluctuations: "no fluctuations" and "fluctuations". However costs and utilities may not be only a function of the defined Markov health states, but also of other variables. Studies by Dodel and Lepen showed that costs are also a function of severity according the Hoehn&Yahr scale, a different clinical scale that combines clinical signs and functional disability.[5] [6] The study by Dodel also showed that utilities of the health states also depend on the Hoehn&Yahr scale. Hence costs and utilities may not be only a function of the defined Markov health states, but also of other explanatory variables, which may act as confounding variables when they are not taken into account.

The objective of this manuscript is to present a strategy to incorporate another explanatory variable in a Markov model in order to increase the external validity and to determine the relevance of the incorporation of another explanatory variable.

# METHODOLOGY

## Types of Confounding Effect by Other Explanatory Variables

The confounding effect of other explanatory variables may affect the costs and utilities of health states as well as the transition probabilities.

### *Health States*

We may distinguish three types of contradictions with external validity, when another explanatory variable is not taken into account.

The first limitation is that assigning a fixed cost to each health state over the follow-up period assumes a static distribution of explanatory variables for each health state, while a dynamic distribution should reflect the reality. Although some explanatory variables may not change over time (e.g. gender), there may be other explanatory variables, which are time dependent (e.g. disease progression).

A second limitation is that in our original model, the distribution of the patients over the various health states was based on the distribution of the sample of the cross-sectional study, which may not correspond with the population distribution in the clinical trial. A related limitation is that the trial population may not be representative of the target patient population that is going to use the drug due to its restriction on external validity. Consequently the average health-specific costs/utilities derived from the sample may differ from reality, when the distribution of an explanatory variable is not taken into account.

### *Transition Probabilities*

The model structure and stratification of Markov health states are based on the primary clinical outcome of the clinical trial and the initial transitions in a model usually are derived from those efficacy measures. In this example the efficacy measure of the trial was the reduction of the development of fluctuations, and the health states were "no fluctuations" and "fluctuations". However, disease progression in PD may be also measured by the Hoehn&Yahr scale, which takes into account a much broader range of factors.[7] It is conceivable that progression according to Hoehn&Yahr may depend on existence of fluctuations and consequently may be different for patients with and without fluctuations. This type of covariance should be taken into account by balancing the treatment arms for all potential confounding variables: in this example the distribution according to Hoehn&Yahr stage should be similar between the treatment arms. Since clinical trials are only powered to prove differences in efficacy between the study drug and the comparator, it is difficult to identify other explanatory variables and determine a statistically significant relationship on the basis of the clinical

trial data, for example a regression equation where efficacy is a function of the study drug and the other explanatory variable.

Finally covariance may also exist between the efficacy of study drug and the explanatory variable. In this example the new study drug may not only reduce the development of fluctuations, but may also reduce disease progression according to Hoehn&Yahr stage.

**Figure 1. Markov model for Parkinson's disease**

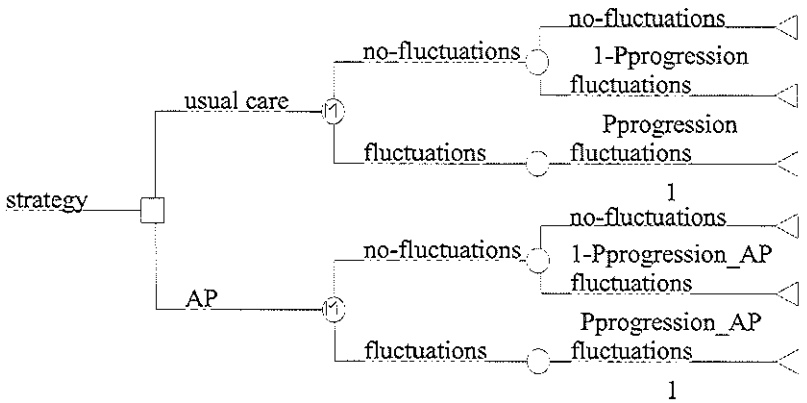



**Table I. The fixed input values for the Markov model.**

| Stage | Costs (€) (3 months) | Utility | Disease progression (3 months) |
|---|---|---|---|
| | Mean | Mean | |
| No fluctuations | 972 | 0.73 | 0.0252 |
| Fluctuations | 2,138 | 0.49 | |

As in the previous paper, concepts are illustrated for a hypothetical Markov model for Parkinson's disease, which will be used to model the cost-effectiveness of a hypothetical new anti-Parkinson drug (AP) in Germany.[4] Published real data were used, whenever possible, but we had to rely on extrapolation methods in absence of real data in order to illustrate the relevant issues. This new drug is supposed to be used as add-on therapy to usual care. The perspective is that of the health insurance. Figure 1 shows the overall structure of this model and the clinical and economic data used to construct the model summarised in Table I. The stratification of the Markov health states is based on the primary clinical efficacy measure of the phase III clinical trial, which was the existence or absence of motor fluctuations. Our model is based on the following assumptions:

- The Markov health states in this model correspond with severity levels according to the existence of motor fluctuations: "no fluctuations" and "fluctuations".
- Disease progression corresponds with the development of fluctuations. We assume that AP will reduce the probability of development of fluctuations annually by 10%. The probability of fluctuations for usual care was based on a study by Rinne (Table I).[8]
- We assume that beyond the actual follow-up period of the trial (1 year) the delay in disease progression due to treatment with AP will continue at the same rate beyond the clinical trial period.
- The severity according to Hoehn&Yahr scale is the only other explanatory variable.
- The model has a follow-up period of 5 years; the cycle time is 3 months, which is based on the follow-up period of the cross-sectional study. We assume that there is no mortality during this study period.
- In our model a hypothetical cohort of persons with no fluctuations will receive AP or usual care. After the first cycle patients may have remained without fluctuations or have become fluctuators. Patients will continue the treatment with AP for the remainder of the study period, also when they have become fluctuators.
- Costs and utilities are discounted at 5%.

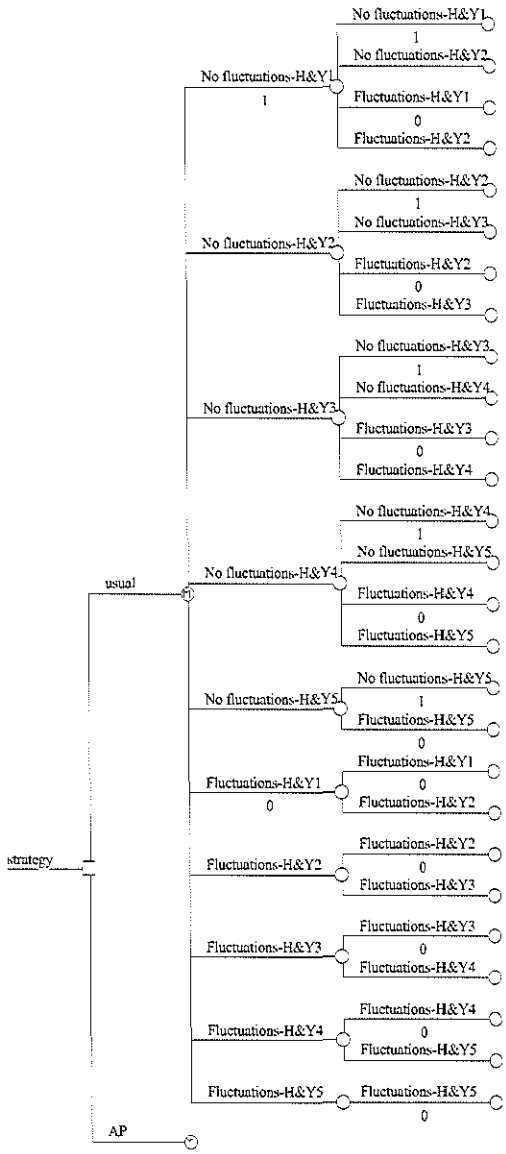
## Incorporation of Other Explanatory Variables

An option is a further stratification of the Markov states by incorporating severity according to the Hoehn&Yahr scale. Figure 2 shows the Markov model after incorporating the severity according to the Hoehn&Yahr scale. This approach has several methodological consequences. Figure 2 shows that the incorporation of only one explanatory variable substantially increases the complexity of the model, which will decrease the transparency of the model. There is general consensus to limit the complexity of models for purposes of transparency. Buxton offered a number of recommendations for good practice in modelling. The first recommendation is that the model should be kept as simple as possible to aid understanding by decision makers.[9] The incorporation of more than one extra explanatory variable would further increase the complexity of the model.

The stratification of the model is also constrained by the availability of data sources to be used in the model yielding data on: transition probabilities, units of healthcare utilisation and utilities.

- Transition probabilities: The probabilities for the initial transitions in a model usually correspond with the efficacy measures derived from a clinical trial. However the clinical trial for a new drug is only powered to show a statistical significant difference between the study drug and comparator for the broad indication. Therefore the trial will not have enough power to determine statistically significant transition probabilities for subpopulations based on the stratification according to other explanatory variables.

**Figure 2. Markov model for Hoehn&Yahr stage.**




- Healthcare utilisation and utilities: When the cross-sectional study is used to determine costs and utilities for Markov health states, the stratification of health states according to the other explanatory variables will result in a substantial increase of subpopulations and consequently a larger total sample size is required, which may not be feasible, especially when more explanatory

variables are included.[2] In the initial model for Parkinson's disease the data collection was performed only for two health states: "no fluctuations" and "fluctuations".

The incorporation of other explanatory variables in a Markov model consists of the following steps:

*Step1: Development of health state-specific linear models*

This strategy consists of an extension of the hybrid design by combining Markov modelling and regression models using a cross-sectional study design. We described in the previous paper in detail the execution of a cross-sectional study, and consequently will focus here only on relevant issues related with the incorporation of explanatory variables. The mean values of all units of costs and utility were initially determined for each severity level ("no fluctuations" and "fluctuations"), which were incorporated in the corresponding health states in the Markov model. An extension of the previous strategy is to determine relationships between the costs/utilities and the explanatory variable within each health state. We will focus on costs only, but the approach for utilities may be considered to be largely similar.

In the existing Markov model the costs per health state are only a constant function of the defined health state: "no fluctuations" and "fluctuations".

C-NF= f(state-NF) (Eq. 1)

C-F= f(state-F)

Where C-NF represents the mean costs of the health state "no fluctuations" (NF) and C-F the mean costs of the health state "fluctuations" (F) in Euro (€).

The first step consists of the development of a relationship between the health states and the other explanatory variable (Hoehn&Yahr score). We develop two separate health-specific linear models for "no fluctuations" and "no fluctuations", because covariance may exist between the Hoehn&Yahr stage and the existence of motor fluctuations. Consequently we defined the following relations:

C-NF= $C1_{NF}$, when Hoehn&Yahr is 1: (Eq. 2)

$C2_{NF}$, when Hoehn&Yahr is 2:

$C3_{NF}$, when Hoehn&Yahr is 3:

$C4_{NF}$, when Hoehn&Yahr is 4:

$C5_{NF}$, when Hoehn&Yahr is 5:

C-F= $C1_{F}$, when Hoehn&Yahr is 1:

$C2_{F}$, when Hoehn&Yahr is 2:

$C3_{F}$, when Hoehn&Yahr is 3:

$C4_{F}$, when Hoehn&Yahr is 4:

$C5_{F}$, when Hoehn&Yahr is 5:

The weighted costs of each health state at cycle *t* are based on the distribution of the five Hoehn&Yahr at cycle *t*:

$C\text{-}NF_t = F1_{NFt} \cdot C1_{NF} + F2_{NFt} \cdot C2_{NF} + F3_{NFt} \cdot C3_{NF} + F4_{NFt} \cdot C4_{NF} + F5_{NFt} \cdot C5_{NF}$

$C\text{-}F_t = F1_{Ft} \cdot C1_F + F2_{Ft} \cdot C2_F + F3_{Ft} \cdot C3_F + F4_{Ft} \cdot C4_F + F5_{Ft} \cdot C5_F$

Where $F1_{NFt}$, $F2_{NFt}$, $F3_{NFt}$, $F4_{NFt}$, $F5_{NFt}$ are frequency distributions of the five Hoehn&Yahr stages in the health state "no fluctuations" at cycle *t*; and $F1_{Ft}$, $F2_{Ft}$, $F3_{Ft}$, $F4_{Ft}$, $F5_{Ft}$ are frequency distributions of the five Hoehn&Yahr stages in the health state "fluctuations" at cycle t.

*Step 2: Incorporation of health state-specific linear models in a Markov model: determination of time-dependent values of an explanatory variable*

In the model, all patients start in the Markov health state corresponding with the indication of the drug (Hoehn&Yahr 1) and subsequently progress to higher severity levels. The values F ($F1_{NFt}$, $F2_{NFt}$, ...$F5_{NFt}$ and $F1_{Ft}$, $F2_{Ft}$, ...$F5_{Ft}$) in the above-mentioned equations reflect the disease progression from Hoehn&Yahr 1 to Hoehn&Yahr 5: the distribution will be skewed to the left in early PD, but skewed to the right in advanced PD. Hence they are time-dependent variables, because they are a function of the time spent in the model.

## Analysis

We performed various analyses in order to assess the impact of the above- mentioned contradictions with external validity, when an explanatory variable is not taken into account.

*Initial Distribution*

The distribution of the patients in the health state NF in the initial model at t=0 was based on the sample of the cross-sectional study, which did not correspond with the clinical trial population consisting only of patients in Hoehn&Yahr stage I ($F1_{NF1}=1$). Consequently the average health-specific costs derived from the sample may differ from the reality, when the cost of the health state NF at t=0 is only based on patients in Hoehn&Yahr I.

The new anti-Parkinson drug AP is indicated for patients with initial symptoms of PD: state=NF and $F1_{NF}=1$ corresponding with the clinical trial population. However the trial population may not be representative of the patient population who is going to use the drug in daily practise. We used in our model the frequency distribution of the AP clinical trial and performed a sensitivity analysis on the initial frequency distribution in order to test the impact of this type of error.

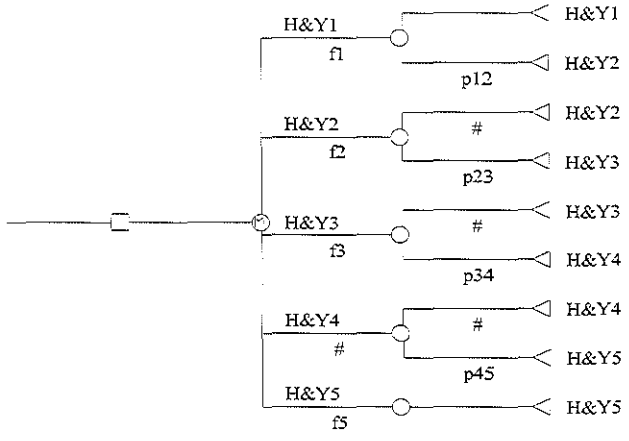
**Table II. The fixed input values for the Markov model.**

| Stage | Transition Probability (3 months) | Costs (€) | |
|---|---|---|---|
| | | NF | F |
| H&Y1 | 0.195 | 121 | 633 |
| H&Y2 | 0.095 | 848 | 1.036 |
| H&Y3 | 0.090 | 917 | 1.181 |
| H&Y4 | 0.055 | 1.223 | 1.601 |
| H&Y5 | 1.000 | 2.137 | 5.956 |

### *Subsequent Distributions*

The study by Davey provided data on the time-dependent relationship for disease progression according to Hoehn&Yahr (Table II).[10] A Markov model was constructed, which consisted of five Markov health states corresponding with five Hoehn&Yahr stages (Figure 3); the cycle time and follow-up corresponded with original model; 3 months and 5 years respectively. The Markov model was used to determine a frequency distribution for the Hoehn&Yahr stages at every cycle. In the base case analysis we assumed that disease progression according to the Hoehn&Yahr scale is independent of the existence of fluctuations and that AP will not influence the disease progression according to the Hoehn&Yahr scale. Therefore extra sensitivity analyses were performed to determined on: 1) The existence of covariance between efficacy of study drug and explanatory variable and 2) The existence of covariance between health states and the explanatory variable.

**Figure 3. Markov model for Hoehn&Yahr stage.**




## RESULTS

The base case analysis is based on a 10% improvement for AP. The adjustment for the explanatory variable is based on 1) a correction for the distribution of the patient population, 2) the stratified costs for each health state according to the Hoehn&Yahr s and 3) the incorporation of the disease

progression for the Hoehn&Yahr. The first column of Table III shows the outcomes, when the analysis is based on fixed values for the health states without incorporating the explanatory variable Hoehn&Yahr stage. The other column shows the results of analysis, when adjusting for the Hoehn&Yahr stage.

**Table III. Results: base case comparison and sensitivity on improvement AP.**

| | **Base-case comparison (AP: 10% improvement)** | | **AP: 20% improvement** | |
|---|---|---|---|---|
| | **Fixed Value** | **Adjusted** | **Fixed value** | **Adjusted** |
| **Costs (€)** | | | | |
| Usual care | 22,007 | 9,358 | 22,007 | 9,358 |
| AP | 25,000 | 12,620 | 24,619 | 12,514 |
| Difference | 2,993 | 3,262 | 2,612 | 3,156 |
| Change | | 269 (9.0%) | | 544 (20.8 %) |
| **Utilities** | | | | |
| Usual care | 3.12 | 3.12 | 3.12 | 3.12 |
| AP | 3.14 | 3.14 | 3.16 | 3.16 |
| Difference | 0.02 | 0.02 | 0.04 | 0.04 |
| **ICR** | 149,650 | 163,100 | 65,300 | 78,900 |
| Change | | 13,450 (9.0%) | | 13,600 (20.8%) |

The absolute costs for both usual care and AP decrease substantially from respectively €22,007 to €9,358 and from €25,000 to €12,620. On the other hand, the increase in costs by using AP increases from €2,993 to €3,262, which is a 9.0% increase. The utilities do not change, because we did not adjust utilities for the explanatory variable. The incremental cost-effectiveness ratio (ICR) increases from 149,650 to 163,100 €/QALY, which is also an increase of 9.0%, because the utility difference does not change. This table also shows that a 20% improvement for AP further decreases the costs and ICR for AP in analyses with and without adjustment for the explanatory variable. However the difference increases from 9.0% to 20.8%. Hence the incorporation of an explanatory variable substantially changed the outcomes, which is sensitive to the extra clinical benefit of the study drug.

Table IV shows the results of a sensitivity analysis for AP, when the distribution of the cost of care patient population is used, which differs from the clinical trial population (external validity). H&Y 1: 20%, H&Y 2: 30%, H&Y 3: 20%, H&Y 4: 20%, H&Y 5: 10%. The results show that the costs for usual care and AP are very sensitive to the distribution of the initial population: an increase from

€9,358 to €22,738 and from €12,620 to €25,811 for respectively usual care and AP. However the ICR only changes from €163,100 to €153,650.

**Table IV. Results: patient distribution based on the cost-of-care study.**

| | Base-case | New outcomes |
|---|---|---|
| **Costs (€)** | | |
| Usual care | 9,358 | 22,738 |
| AP | 12,620 | 25,811 |
| Difference | 3,262 | 3,073 |
| **ICR** | 163,100 | 153,650 |

Table V shows results of a sensitivity analysis for AP, when the clinical trial population is not representative of the patient population who is going to use the drug in daily practise. We assumed that not only non-fluctuating patients in H&Y stage I are going to use the drug, we assumed following distribution H&Y 1: 80%, H&Y 2: 20%, H&Y 3: 0%, H&Y 4: 0%, H&Y 5: 0%. The results show that the model is not very sensitive to the discrepancy between clinical trial population and the "real" patient population.

**Table V. Results: external validity (clinical trial versus real population).**

| | Base-case | New outcomes |
|---|---|---|
| **Costs (€)** | | |
| Usual care | 9,358 | 10,867 |
| AP | 12,620 | 14,134 |
| Difference | 3,262 | 3,267 |
| **ICR** | 16,300 | 163,350 |

Table VI shows the outcomes, when we assume covariance between efficacy of the study drug (AP) and the explanatory variable (H&Y): AP reduces disease progression according to H&Y by 10%. The costs for AP decrease from €12,620 to €12,190, while the costs difference with usual care decreases by €430. The incremental cost-effectiveness ratio (ICR) changes from 163100 to 1,416,000 €/QALY. Hence the existence of covariance between efficacy of study drug (AP) and explanatory variable (H&Y) may also influence the outcomes of the model.

**Table VI. Results: covariance AP and H&Y progression.**

| | Base-case | New outcomes |
|---|---|---|
| **Costs (€)** | | |
| Usual care | 9,358 | 9,358 |
| AP | 12,620 | 12,190 |
| Difference | 3,262 | 2,832 |
| **ICR** | 163,100 | 141,600 |

Table VII shows the outcomes, when we assume covariance between health state and explanatory variable (H&Y): disease progression according to H&Y is 10% lower in patients without fluctuations and 10% higher in patients with fluctuations. This analysis also affects the outcomes for usual care. The costs for usual care decrease from €9,358 to €7,338; the costs for AP decrease from €12,620 to €10,066. The costs difference with usual care changes from €3,262 to €2,728 and the incremental cost-effectiveness ratio (ICR) changes from €163,100 to 136,400 €/QALY. Hence the existence of covariance between health state and the explanatory variable (H&Y) may substantially influence the outcomes, and in this example especially the ICR.

**Table VII. Results: covariance fluctuations and H&Y (H&Y progression depends on fluctuations).**

| AP | Base-case | New outcomes |
|---|---|---|
| **Costs (€)** | | |
| Usual care | 9,358 | 7,338 |
| AP | 12,620 | 10,066 |
| Difference | 3,262 | 2,728 |
| **ICR** | 163,100 | 136,400 |

Table VIII shows the results of a probabilistic sensitivity analysis, which was based on the distributions of the costs. This analysis shows that the total uncertainty associated with outcomes of the model substantially decreases after the adjustment for the Hoehn&Yahr classification: The standard deviation after incorporation of the Hoehn&Yahr state is only approximately 10% of the standard deviation of the initial model for both treatment arms.

**Table VII. Results of probabilistic sensitivity analysis.**

| SD (€) | Base-case | New outcomes |
|---|---|---|
| Usual care | 14,614 | 1,947 |
| AP | 14,862 | 1,960 |

## DISCUSSION

In this manuscript we presented a strategy to improve the quality of the methods used in Markov modelling studies by increasing the external validity. This strategy consisted of an incorporation of an explanatory variable in the Markov health states by means of health state specific relationships between the explanatory variable and costs as well as time-dependent values of the explanatory variable. In addition we further explored the consequences of incorporating an explanatory variable by means of various sensitivity analyses. The strategy was applied to a hypothetical Markov model for a new product AP in Parkinson's disease. The results showed that the outcomes of a health economic model can be severely biased, when other explanatory variables are not taken into account.

In addition we showed that this discrepancy is due to three different types of contradictions with external validity: 1) fixed costs for each health state over time versus time-dependence of costs of each health state over time due to other time-dependent explanatory variables; 2) initial model population distribution corresponding with the cost of care study versus initial model population distribution adjusted to the clinical trial population and 3) the discrepancy between the trial population and the actual population. In addition we showed the sensitivity of the outcomes of the model to: 1) covariance for time-relationship between health state and the other explanatory variable and 2) covariance between efficacy of study drug and the other explanatory variable. The sensitivity analyses showed that the bias may vary for each type of error and may differ for the various outcomes (costs, ICR) of the model. For example, in this study the error due to inappropriate patient distribution mainly changed the economic outcomes, while the cost-effectiveness remained similar.

There are a number of limitations in this study. The current strategy is based on a number of assumptions, which may often not hold, and may be in conflict with reality. An assumption was that the Hoehn&Yahr scale was the only confounding variable. Literature showed also the existence of other confounding variables, for example age and ADL.[6] Incorporation of a higher number of confounding variables may increase the external validity of the Markov model, but the feasibility depends on the statistical power of underlying cost-of-care data and may consequently substantially increase the necessary sample size for a cross-sectional study. An alternative is the use of longitudinal observational databases, which are designed to measure the clinical and economic impact of a particular disease in a large number of patients without any intervention with real practice. Another

limitation was that this study only considered the impact of other explanatory variables on costs, while ignoring the consequences on utility outcomes.

Although the incorporation of other explanatory variables in the model introduces the extra uncertainty associated with those variables, the sensitivity analysis for this model in PD showed that this extra uncertainty is more than offset by the reduction of the overall uncertainty of the model. The Markov states "F" and "NF" are heterogeneous populations, when taken into account the Hoehn&Yahr stage and consequently the costs show a large variance. The additional stratification according to the Hoehn&Yahr stage within each Markov state yields more homogenous subpopulations, which consequently will show less variance in costs. This reduction in variance may be counterbalanced by lack of statistical power, when the sample sizes for the subpopulations become too small.

From the perspective that results from modelling studies are of interest to many parties with different responsibilities and diverging interests, it is obvious that recognising the relevance of explanatory variables and appropriately dealing with explanatory variables is required for obtaining unbiased results from health economic modelling studies, especially when those data are being used for reimbursement decisions. Although this study clearly showed the relevance of the confounding variable using a specific example, we recommend development of more general methodologies, which can be applied for dealing with confounding variables taking into account all types of error, which were identified in this study. For example, the impact of an explanatory variable on utilities should also be incorporated in the model, and subsequently the modelling of the interaction of costs/utilities and measures of disease progression needs to be considered.

## References


[1] Drummond MF, et al. Current Trends in the Use of Pharmacoeconomics and Outcomes Research in Europe. Value in Health 2(5):323-332

[2] Nuijten MJC. Data management in modelling studies: The selection of data sources. PharmacoEconomics, March 1998; 3 (3): 305-316

[3] Nuijten MJC. Applications of Modelling Studies. PharmacoEconomics, March 1998; 3 (3): 289-291

[4] Nuijten MJC. Bridging decision analytic modelling with a cross-sectional approach. PharmacoEconomics. 2000, March: 17 (3): 227-236

[5] Dodel RC, Singer M, Kohne-Volland R, Szucs T, Rathay B, Scholtz E, Oertel W. The economic impact of Parkinson's disease. Pharmacoeconomics 1998; 14 (3):299-312

[6] LePen C, Wait S, Moutard-Martin, and Ziegler M. Cost-of-illness and disease severity in a cohort of French patients with Parkinson's disease. Pharmacoeconomics. 1999; 16(1): 59-69

[7] Fearnley J, Lees A. Pathology of Parkinson's disease. In: Calne DB, editor. Neurogenerative diseases. Philadelphia: WB Saunders, 1994: 545-54

[8] Rinne et al. Early treatment of Parkinson's disease with cabergoline delays the onset of motor complications. Drugs, 1998; Vol. 55, Suppl. 1,23-30

[9] Buxton MJ, Drummond MF, van Hout B. Modelling in economic evaluation: an unavoidable fact of life. Health Econ. 1997: 6 (3): 217-28

[10] Davey P, Rajan N, Lees M, Aristides M. Cost-effectiveness of pergolide compared to bromocriptine in the treatment of Parkinson's disease: a decision-analytic model. Value Health 2001 Jul-Aug;4(4):308-15

Chapter 8

# Measuring Sensitivity in Pharmacoeconomic Studies: An Integration of Point-sensitivity and Range-sensitivity


## SUMMARY

The level of uncertainty with regard to the outcomes of pharmacoeconomic studies cannot be completely covered by the statistical methods routinely employed to handle uncertainty in clinical research. Sensitivity analysis is the most common methodology to deal with the extra uncertainty associated with pharmacoeconomics, and has also been incorporated in recent guidelines on healthcare evaluation. However, the execution of a sensitivity analysis and the interpretation of its results have not yet been standardised, which may lead to subjectivity and consequently weaken the value of economic evaluations. This article presents a method of dealing more systematically with uncertainty and eliminating potential bias in sensitivity analysis, with regard to the measurement of sensitivity and the comparison of the degree of sensitivity between variables. An assessment of the disadvantages of using slope as a measure of sensitivity leads to 2 types of sensitivity analyses (point-sensitivity and range-sensitivity), which are integrated into one method for the measurement of sensitivity.

## INTRODUCTION

Escalating costs have become a major concern for healthcare professionals, decision-makers and the public, prompting the increasing use of economic evaluations of healthcare alternatives. Health care economic research identifies measures and compares the costs and clinical outcomes of different treatment strategies. In the process, inputs (in the form of costs of resources consumed) are compared with outputs (such as the benefits to the patient resulting from a specific treatment). Healthcare economic evaluation is a science using epidemiological, economic and clinical methodologies that studies the appropriate use of medical interventions and is a tool that can give guidance to efficient resource allocation in health care.

Economic evaluation is beginning to be used widely to support decision makers in health services for allocation of scarce health care resources, and guidelines for evaluations have been developed in Australia and Canada.[1,2] Because economic evaluation is used for pricing and reimbursement issues by authorities or third-party payers, it is vital that these analysis are performed according to generally accepted and standardized methods. This increases the transparency of pharmacoeconomic studies, allowing a more rational interpretation of results and a better comparison of different studies.

The importance of dealing systematically and comprehensively with uncertainty is often overlooked by many analysts. Udverhelayi et al note that although authors frequently mentioned the limitations in their underlying assumptions, only 30% of the studies used sensitivity analysis to explore the effects of changes in those assumptions.[3]

Sensitivity analysis is currently the most widely applied method of dealing with uncertainty in economic evaluations. A sensitivity analysis is based on the modification of the basic clinical and economic estimates of parameters over a plausible range of values, in order to judge the effect on study results of alternative assumptions for the range of potential values for uncertain parameters.[4]

Most recent pharmacoeconomic publications contain sensitivity analyses on several parameters. However, a sensitivity analysis is often performed on only a limited number of variables, without justification of the choice of selected variables or the chosen range of each variable. In addition, the interpretation of the results of a sensitivity analysis remains subjective as there is no scale for measurement of sensitivity and consequently there is no threshold value for discriminating between "sensitive" and "not sensitive".

A recent publication of by Briggs recommends standardization of the methodology of sensitivity analysis in order to increase the reliability of economic evaluations.[5] Although the current guidelines recommend sensitivity analyses, they do not contain any particular requirements regarding how to perform such an analysis.[6] [7] The Australian guidelines, for example, suggest determining the sensitivity of the overall analysis by substitution of the upper and lower confidence limits of the variable parameters. In addition to this one-way sensitivity analysis, the Canadian guidelines encourage more sophisticated approaches, such as Monte Carlo simulation, which permits the analyst

to assign ranges and distributions to uncertain variables that are being modelled using decision-analytical techniques.[8]

The present study describes a mathematical approach to measuring sensitivity, and defines 2 types of sensitivity (point-sensitivity and range-sensitivity); a general strategy is described incorporating both types of analysis. The concept is illustrated by application of the process to an economic evaluation based on decision analytic techniques, since such analyses are widely used in pharmacoeconomic studies and the results are becoming generally accepted by regulatory authorities.[1,2,6] In addition, since sensitivity analysis is already being applied in decision analysis provides a useful tool for illustration of the refined concept described here.[9]

# METHODS

## Model Description

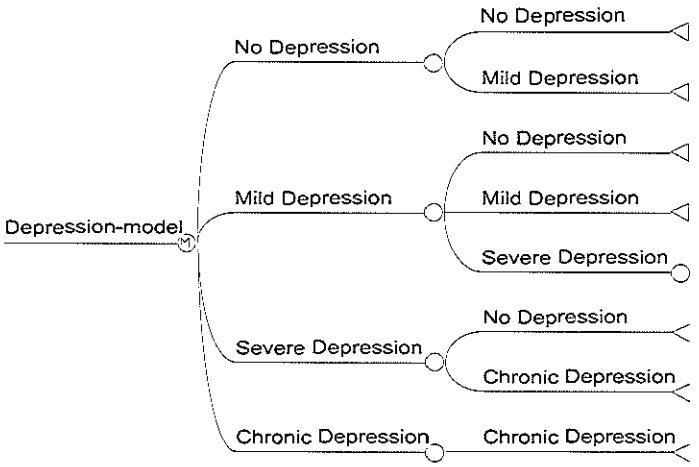
Decision analysis a well-recognized method for analysing the consequences of decisions that are made under uncertainty.[9] It is an explicit, quantitative, prescriptive approach to medical decision making and allows both clinical and economic consequences of medical actions and attitudes to be analysed under conditions of uncertainty.

From treatment algorithms, a decision tree model can be constructed, which considers the timings of actions and their consequences over time. A decision tree consists of a series of branches, each representing different options (decisions or events), which arise at different points, referred to as nodes (e.g. decision node and chance nodes). In effect, the decision tree provides a therapeutic model, showing the consequences and complications of different therapeutic interventions.

The model resulting from the decision analysis must correspond, as much as possible, to the real life situation of the disease and should reflect actual treatment patterns, with input values (probabilities and items of health care utilization) deviating as little as possible from population values. Data sources for the variables being used in a model may be meta-analysis, databases, clinical trials and/or Delphi panels, and for each variable a fixed input value and a range can be derived.

The analysis, being based on a fixed input value for all variables, will yield the basic average outcome per patient (e.g. average number life years, average costs per patient). The fixed input value can be the mean in the case of a large sample size (for example a database), whereas the median is recommended for smaller sample sizes. The range of each variable is then used to determine the sensitivity of the outcome to the analysis, when the input value is varied within its range.

**Figure 1. The Markov state model for depression.**




## Application

To illustrate the application of the sensitivity analysis technique described here, a Markov model is used and applied to the economic outcomes of maintenance treatment with selective serotinin reuptake inhibitors (SSRIs) as the first choice therapy for depression over a 1-year follow up period. The general structural details of the Markov depression model have already been described in a previous paper and the specific details of the model with respect to depression are shown in figure 1.[10]

The model in figure 1 defines 4 mutually exclusive states: no depression, mild depression, severe depression and chronic depression and the patient's progression through these states is divided into 6 cycles of 2 months each, which closely approximates the time of the sequential therapeutic stages. Chronic depression represents an absorption state, and after entering this state, a patient remains there. The transition probabilities between the states are based on response rates to treatment and relapse rates after termination of treatment; Table 1 shows the probabilities for each transition in the model. This information was obtained from published literature values and current clinical practice.[10] The cost assessment was based on the assignment of a fixed cost to each health state, which depends on the choice of therapy; Table 1 lists these costs, together with the source material for the valuation. Further details of the sources for the clinical and economic data, with respect to depression, have been given in a previous article.[10]

In the present study, the depression model is being used only to illustrate the concept of sensitivity. The range of variables has been chosen to show the relevant issues involved in sensitivity analysis

without considering their external clinical relevance. The input values have been varied to illustrate the concepts of point sensitivity as applied to the model for depression.

**Table I. Point-sensitivity analysis at fixed input values for variables in the depression model.**

| Variable | Fixed Input values | Point Sensitivity |
|---|---|---|
| **Clinical probabilities** | | |
| Response rate to first-line treatment with an SSRI (p1) | 0.60 | -1.23 |
| Response rate to TCA after SSRI treatment failure (p2) | 0.30 | -0.39 |
| Response rate to another TCA after earlier TCA treatment failure (p3) | 0.30 | 0.40 |
| Response rate to hospitalisation (p4) | 0.50 | 0.45 |
| Relapse rate during maintenance treatment with SSRI (r1) | 0.04 | 0.06 |
| Relapse rate during no treatment after response to a TCA (r2) | 0.10 | 0.01 |
| **Economic ($US)**** | | |
| Daily treatment cost of SSRI (DTC-SSRI) | 2 | 0.03 |
| Daily treatment cost of TCA (DTC-TCA) | 0.19 | 0.00 |
| Consultation tariff (CT) | 50 | 0.02 |
| Per diem tariff (pd)* | 500 | 0.94 |

Abbreviations: SSRI = selective serotonine reuptake inhibitor; TCA = tricyclic antidepressant

* From Nuijten[10]

** From TLG Costing Database, 1995

# RESULTS

## Measurement of Sensitivity

The most common graphical result of a sensitivity analysis for a decision-analytic model is a linear relationship between an input variable and an outcome. Figure 2 shows the relationship between the *per diem* tariff and the average total direct medical costs per patient over a 1-year period, when the former is varied from $US200 to $US800 within the depression model. The basic average cost per patient is $13,813 for a fixed input value for the *per diem* of $US500.

The slope of the line in figure 2 provides a measure of the sensitivity of the outcome to a variation in the *per diem* tariff and is directly proportional to the outcome. However, the slope of a sensitivity line is a limited measure of responsiveness of the outcome to a change in a variable because it does not take into account the fixed input value and the range associated with that variable. Instead the slope depends only on the structure of the model and the fixed input variables of the other variables.

## Point-Sensitivity

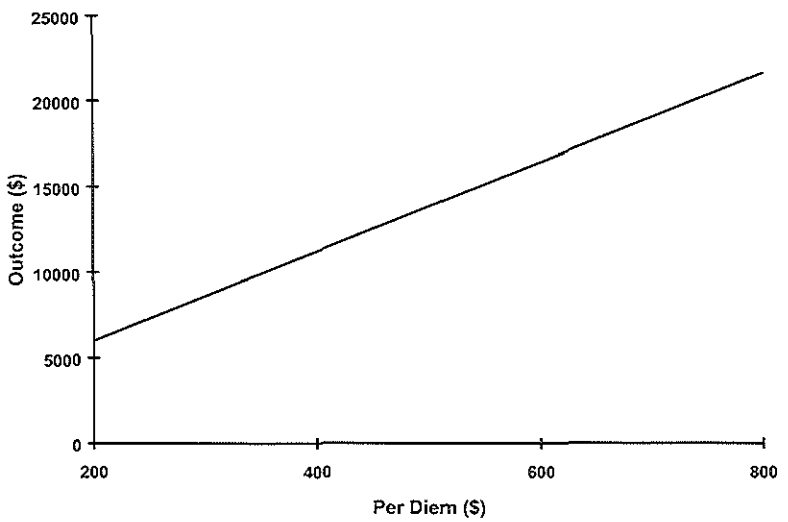
In the sensitivity analysis for the *per diem*, which is shown in figure 2, the slope of the line is identical for every range of input values. However, an increase in the *per diem* of, say, $US200 is a large

increase for a fixed input value of $US200 and a less significant increase for a fixed input value for the *per diem* of $US800. In addition, by an analogous argument, knowing the difference in outcome is not very revealing, unless compared with the basic average outcome. An increase of $US100 is quite a significant reaction to the value of the *per diem*, if the foldback outcome (average total direct medical costs) is $US200, but it is less significant if the foldback outcome is SUS1,000.

To take account of the fixed input value, sensitivity is best calculated by comparing the percentage change in input value to the percentage change in outcome value.

In many respects, measuring the responsiveness of the demand for a quantity of a particular commodity to changes in its price is similar to a sensitivity analysis in the present context, in which the extent of responsiveness of an outcome to changes in an input variable over a range are quantified.

**Figure 2. Sensitivity analysis for the *per diem* tariff. Using the model for depression, the outcome (average total direct medical cost over a 1-year period) was determined for different input values for the *per diem* tariff, ranging from $US200 to $US800.**




In economics, the price elasticity of demand ($\eta$) is a measure of the responsiveness of quantity of a commodity to a change in market price and is defined as: the percentage change in quantity demanded divided by the percentage change in price.[11] The changes in price, and the resulting changes in quantity, are expressed as a percentage change, with respect to the average price or quantity. Therefore, this equation provides a measure of the responsiveness, or elasticity, of one item with respect to the other. The more responsive the quantity demanded to changes in price, the greater the elasticity and the higher the absolute value of $\eta$. The direction of the relationship between the price and quantity determines the sign of $\eta$. Measuring the extent of responsiveness of quantities to changes in price is similar to the concept of sensitivity analysis, in which the extent of responsiveness of an

outcome to changes in an input variable over a range is determined. Hence, sensitivity may be defined according to the formula of elasticity, as:

Sensitivity = (Δ outcome/Δ input) x (average input/average outcome). (Eq. 1)

where "Δ outcome" represents the change in outcome resulting from a change in input ("Δ input"). This definition for sensitivity ($\eta_s$), may be written symbolically in the following form, in which Y and X represent outcome and input values respectively:

$\eta_s = (\Delta Y/\Delta X)$ x (average X/average Y) (Eq. 2)

where the averages over the arc of the sensitivity curve are being considered (fig 2). The responsiveness of outcome to an input value at a particular point on a sensitivity curve can also be determined, a concept referred to as point-elasticity. The precise definition of point-elasticity uses the concept of a derivative, and is given by:

$\eta_s = dY/dX \times X/Y$ (Eq. 3)

where $\eta_s$ in this case refers to point-elasticity, and $X$ and $Y$ represent the fixed input value and the fold back outcome, respectively.

Point-elasticity in sensitivity analysis measures the responsiveness of the outcome to the input value at a fixed point and does not include any associated range of the input. In the case of a linear sensitivity relationship, the formula for point-elasticity can be simplified in he following manner. If Y=outcome and X=input then the equation of the sensitivity line is:

$Y=(a \times X) + b$ (Eq. 4)

The slope $a$ of the line may be given by:

$a=dY/dX$ (Eq. 5)

Hence, by substitution in equation 3:

$\eta_s = a \times X/Y$ (Eq. 6)

or:

$\eta_s = (a \times X)/[(a \times X) + b)]$ (Eq. 7)

This equation for point-elasticity permits the determination of the sensitivity of outcome to input, an analysis that in the present study, is referred to as point-sensitivity. From equation 7, it is clear that the determination of point-sensitivity is dependent on the slope of the line and the fixed input value, but is not related to the range of a variable.

For example, if the point-sensitivities are determined for *per diem* tariffs of $US200 and $US800, using the equation derived above, the respective values are 0.87 (relatively inelastic) and 0.96 (nearly

elastic) for a=26 and b=795. These values show that the outcome in the depression model is more point-sensitive to a *per diem* of $US800 than for of $US200. The point-sensitivity shows how critical a variable is for the model and it allows a comparison between all variables, which is not confounded by the range of each variable. The overall relationship between a range of *per diem* values and point-sensitivity is shown in figure 3, and similar calculations can be made for all the variables in the depression model. In Table I, point-sensitivity values, together with the respective sign of the value, have been calculated for all the variables in the model. According to the terminology of economic elasticity, the higher the absolute value for point-sensitivity, the more sensitive the model is to this particular variable.[11] The most critical parameter for the model is the response rate to first-line treatment with an SSRI (*p1*); the outcome of the model is perfectly point-sensitive to this variable, in that $\eta_s$ (*p1*) is larger than one, whereas the *per diem* tariff (*pd*) yields a moderate point-sensitivity, with a $\eta_s$ (*pd*) value of 0.94. The negative sign of the point sensitivity of *p1* means that an increase in response rate to first-line treatment leads to a reduction in direct medical costs. In contrast, the outcome of the model is not point-sensitive (i.e. inelastic) to the other variables since $\eta_s$ is smaller than one. Nevertheless, there is a substantial difference in point-sensitivity between the response rate to hospitalisation (0.45) and the daily treatment cost of an SSRI (0.03).

## Range-Sensitivity

The slope of a sensitivity analysis (fig. 2) does not take into account the responsiveness of outcome to the range of an input value. For example, using the data in figure 2, the sensitivity, as determined by the slope, for a fixed input value of $US500 (slope = 26) is identical for a small range ($US400 to $US600) and a large range ($US200 to $US800). However, a sensitivity analysis based on an input value with a large range clearly leads to a larger difference in outcome than an input value with a small range and, hence the response of the outcome is very much dependent on the range of the input value. Therefore, the difference in outcome between a fixed input value of $US500 with a small range ($US400 to $US600) and a large range ($US200 to $US800) is $US5,207 and $US15,621, respectively (fig. 2).

The most common currently used method of measuring sensitivity is based on determination of the difference in outcome when an input value is varied within its range. This type of sensitivity analysis, defined for the remainder in this article as range-sensitivity, does not depend on the fixed input value, but instead depends only on the range of the variable. Therefore, the range-sensitivity for a *per diem* ranging from $US200 to $US300 would be identical to that of a *per diem* varying from $US400 to $US500. As range-sensitivity depends only on the range of a variable, the change in outcome with respect to a variable input can be expressed as:

$\Delta$Outcome = slope x $\Delta$Input

or:

$\Delta Y = a \times \Delta X$

where the slope *a* is derived from the appropriate sensitivity curve (fig 2). Table II shows a range-sensitivity analysis for all the variables in the depression model, in which range-sensitivity is expressed as the absolute difference in outcome between the minimum and maximum input values. It is clear that the *per diem* tariff is the most critical variable when this value ranges from $US200 to $US800, leading to an absolute difference in outcome of $US15,621. However, if the data source used provides a *per diem* with a much smaller range in values, the absolute difference in outcome would clearly be considerably less.

**Table II. Range-sensitivity analysis. Range-sensitivity is expressed as the absolute difference in outcome between the minimum and maximum input values.**

| Variable | Input values min. – max. | Absolute change in outcome ($) | Relative range-sensitivity |
|---|---|---|---|
| **Clinical probabilities*** | | | |
| Response rate to first-line treatment with an SSRI (p1) | 0.50-0.70 | 5,650 | 0.41 |
| Response rate to TCA after SSRI treatment failure (p2) | 0.15-0.20 | 891 | 0.06 |
| Response rate to another TCA after earlier TCA treatment failure (p3) | 0.10-0.50 | 7,436 | 0.54 |
| Response rate to hospitalisation (p4) | 0.10-0.90 | 9,850 | 0.71 |
| Relapse rate during maintenance treatment with SSRI (r1) | 0.02-0.10 | 1,661 | 0.12 |
| Relapse rate during no treatment after response to a TCA (r2) | 0-0.50 | 932 | 0.07 |
| **Economic ($US)**** | | | |
| Daily treatment cost of SSRI (DTC-SSRI) | Fixed | 0 | 0 |
| Daily treatment cost of TCA (DTC-TCA) | Fixed | 0 | 0 |
| Consultation tariff (CT) | 10-90 | 526 | 0.04 |
| Per diem tariff (pd)* | 200-800 | 15,621 | 1.13 |

Abbreviations: SSRI = selective serotonine reuptake inhibitor; TCA = tricyclic antidepressant

* From Nuijten[10]

** From TLG Costing Database, 1995

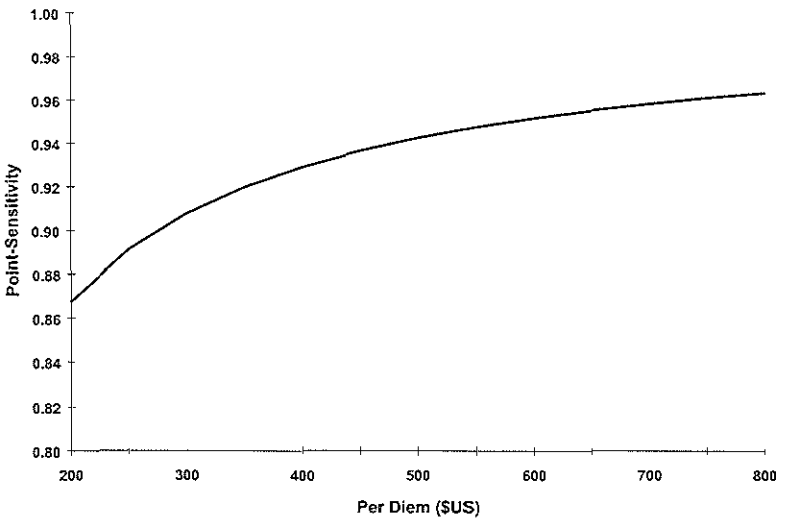
Another disadvantage of the absolute difference in outcome as a measure of range-sensitivity is that it does not take into account the basic average outcome: the model for depression discussed in the present study with an outcome of $US13,813 is less sensitive to an absolute difference of $US15,621 than a model with a 10-fold lower outcome of $US1,381. The relative-range-sensitivity, determined by dividing the absolute difference in outcome by the basic average outcome, permits direct comparison of different pharmacoeconomic studies, provided they have similar outcomes.

## Overall Sensitivity

In the examples shown in Table II, the *per diem* tariff is the most range-sensitive variable. However, as discussed above, the degree of range sensitivity of the per diem tariff depends on the range of the input and the same principle applies to any variable used to assess sensitivity. In order to determine

which variable has the most impact on a model, point-sensitivity and range-sensitivity can be integrated into an overall measure of sensitivity. The combination is achieved using the concept of an integral of point-sensitivity between the lower and upper limits of the range of an input variable. For the *per diem* tariff, this would correspond to the area under the point-sensitivity/*per diem* curve (AUC) [fig. 3] between the upper and lower limits of the range of the variable.

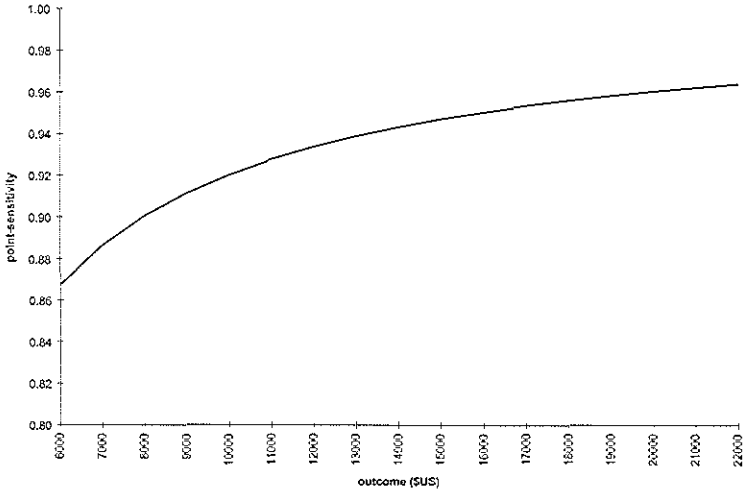
**Figure 3. The relationship between point-sensitivity and *per diem* tariff for the depression model. Point-sensitivity was determined for different *per diem* values (US$200 to US$800) using the methods described in the text.**




However, a severe limitation of this approach is that the area under the curve for one variable cannot be compared to another, as each measurement has its own units, based on the units of the input variable. For example, for the *per diem* tariff the unit of the overall measurement is $US, while the measurement for the response to hospitalisation is a value without a unit of measurement.

On the other hand, a relationship between point-sensitivity (determined according to the range of an input variable) and the corresponding outcome allows the determination of an overall measure of sensitivity, based on the units of the outcome (in this case, $US). Consequently, the measurement of sensitivity has the same units for all variables in the model. In figure 4, point-sensitivity has been determined for a range of the *per diem* tariff and plotted against the outcome for each *per diem* tariff.

**Figure 4. The relationship between point-sensitivity and outcome (direct medical cost in $US).**




The AUC in this case gives a measure of overall sensitivity, which can also be calculated by integration.

From equation 3:

$\eta_s = (dY/dX) \times (X/Y)$

$\eta_s = (a \times X)/[(a \times X)+b]$

$Y=(a \times X)+b$

$X=(Y-b)/a$

Therefore, by substitution:

$\eta_s = (Y-b)/Y=1-b/Y$

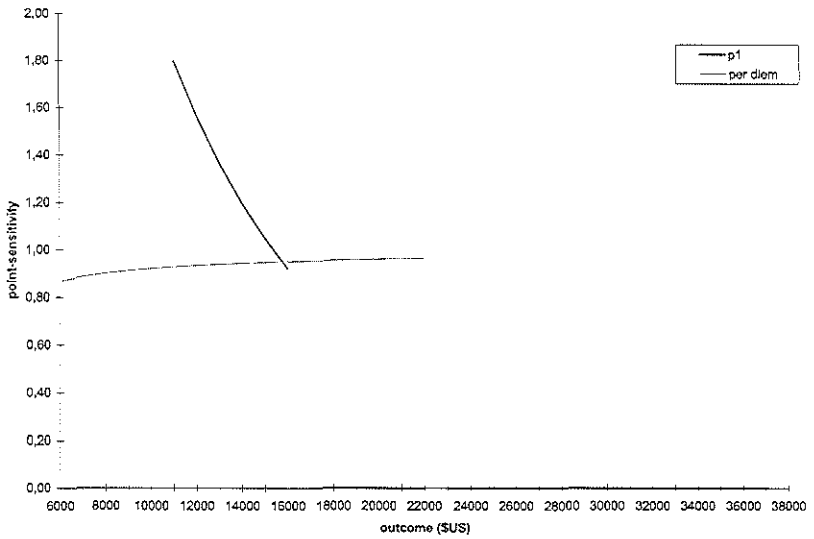
By integration of this last equation, the following expressions for the measurement of overall sensitivity can be obtained:

$$\text{overall sensitivity} = [Y-b \times \text{Ln}(Y)]_{\text{lower limit}}^{\text{upper limit}}$$

The advantages of this approach are demonstrated in figure 5, in which the sensitivity of the depression model has been determined for *per diem* input (*pd*) and response rate to first-line treatment with a SSRI (*p1*). The variable inputs can be directly compared, in terms of their effect on outcome, even though they themselves have different units, in this case US$ (*per diem* tariff) and a probability of response rate. This graph shows that the overall sensitivity of the *per diem* (*pd*) is largely

determined by its range-sensitivity, while the overall sensitivity of the response rate (*p1*) mainly depends on the point-sensitivity.

**Figure 5. Overall sensitivity analyses for 2 variables of the depression model in. Point-sensitivity and outcome (average total direct medical cost per patient over a 1-year period) were determined for a range of input values for the *per-diem* tariff (pd) and the response rate to first-line treatment with an SSRI (p1), respectively.**




The advantage of the concept of overall sensitivity is that all variables can be ranked, incorporating both types of sensitivity as shown in Table III. In addition to the absolute value, a relative overall sensitivity can be determined. This measure is determined by dividing the value for overall sensitivity by the basic average outcome of the model. Relative overall sensitivity allows a comparison of sensitivity of variables between different studies with similar end-points (e.g. costs).

Perhaps the greatest advantage of the measurement of overall sensitivity is the ability to directly compare the impact of all the variables on the model for depression. Table III shows that the *per diem* tariff is the most sensitive parameter for this model while other parameters, such as the relapse rate during maintenance treatment with an SSRI, have little or no effect.

**Table III. Overall sensitivity for all variables in the model for depression.**

| Variable | Overall absolute sensitivity ($US) | Overall relative sensitivity |
|---|---|---|
| **Clinical probabilities*** | | |
| Response rate to first-line treatment with an SSRI (p1) | 7,114 | 0.52 |
| Response rate to TCA after SSRI treatment failure (p2) | 173 | 0.01 |
| Response rate to another TCA after earlier TCA treatment failure (p3) | 3,266 | 0.24 |
| Response rate to hospitalisation (p4) | 5,045 | 0.37 |
| Relapse rate during maintenance treatment with SSRI (r1) | 144 | 0.01 |
| Relapse rate during no treatment after response to TCA (r2) | 30 | 0 |
| **Economic ($US)**** | | |
| Daily treatment cost of SSRI (DTC-SSRI) | 0 | 0 |
| Daily treatment cost of TCA (DTC-TCA) | 0 | 0 |
| Consultation tariff (CT) | 12 | 0 |
| Per diem tariff (pd)* | 14,602 | 1.06 |

Abbreviations: SSRI = selective serotonine reuptake inhibitor; TCA = tricyclic antidepressant
* From Nuijten[10]
** From TLG Costing Database, 1995

A measure of the total level of uncertainty of the model associated with input values may be determined by measurement of the cumulative relative overall sensitivity, being a summation of all relative overall sensitivity values for all variables, which are shown in Table III. In the current model, the total relative sensitivity was 2.20. This measure of total relative sensitivity may also be used to compare different studies.

## DISCUSSION

A method of dealing with uncertainty is presented in this paper. The concept has been illustrated for a modelling study, but may also be applied to other designs of economic evaluations. In most pharmacoeconomic studies, however, the uncertainties created by sampling error are minuscule compared to those created by assumptions and therefore sensitivity analyses are indispensable for modelling studies, like the one presented.

The overall sensitivity measurement is based on combining point-sensitivity and range-sensitivity. This method allows an objective judgment to be made of the sensitivity of all variables of the model, which subsequently may be ranked according to degree of sensitivity. This approach avoids the subjective selection of variables for the sensitivity analysis and the potential bias in judging the degree of sensitivity in most current economic studies. Additionally, the relative overall sensitivity enables an objective comparison between the impact of a variable in 2 different studies. Finally, the total level of uncertainty of a model may be determined by summation of the relative overall sensitivity of all variables.

## References


1. Australia Commonwealth Department of Health, Housing and Community Services. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Commonwealth Department. 1992
2. Detsky, A.S. Guidelines for economic analysis of pharmaceutical products: a draft document for Ontario and Canada. Pharmacoeconomics. 1993; 3: 345-61
3. Udvarhelyi S, Colditz GA, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? Annals of Internal Medicine 1992; 116: 238-244
4. Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health care Technology: A Report on Principles. Ann Intern Med. 1995; 122: 61-70
5. Briggs A. Handling uncertainty in the results of economic evaluation. OHE briefing No 23. 1995. London
6. Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. 1st ed. Ottawa: CCOHTA. 1994
7. Torrance et al. Canadian Guidelines for Economic Evaluation of Pharmaceuticals. PharmacoEconomics. 1996; 9(6): 535-559
8. Critchfield G. C., Willard K. E., Connelly D.P. Probablistic sensitivity analysis methods for general desicion models. Computers and Biomedical Research. 1986; 19: 254-65
9. Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-65
10. A Markov Process Analysis Comparing the Cost Effectiveness of Maintenance Therapy with Citalopram versus Standard Therapy in Major Depression. Nuijten M, Hardens M, Souetre E. Pharmacoeconomics. 1995; 8 (2): 159-168
11. Lipsey R, Steiner P, Purvis D, et al. Economics. 9th ed. New York. Harper and Row Publishers, 1990

Chapter 9

# Measuring Sensitivity in Pharmacoeconomic Studies: Refining Point-Sensitivity and Range-Sensitivity by Incorporating Probability Distributions


## SUMMARY

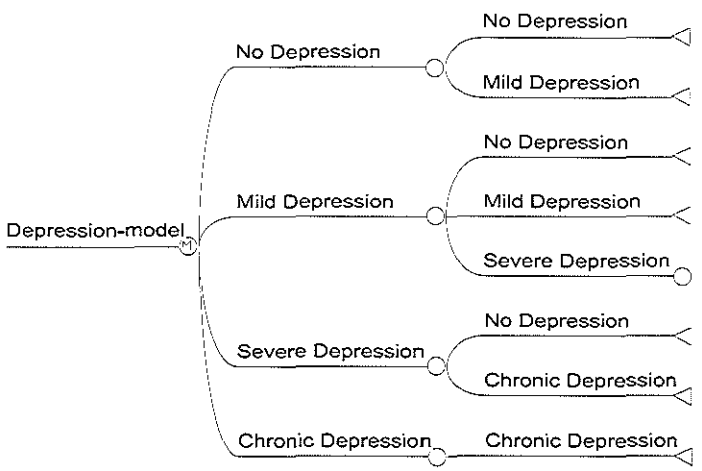
OBJECTIVE: The aim of the present study is to describe a refinement of a previously presented method, based on the concept of point-sensitivity, to deal with uncertainty in economic studies. DESIGN: The original method was refined by the incorporation of probability distributions which allow a more accurate assessment of the level of uncertainty in the model. In addition, a bootstrap method was used to create a probability distribution for a fixed input variable based on a limited number of data points. The original method was limited in that the sensitivity measurement was based on a uniform distribution of the variables and that the overall sensitivity measure was based on a subjectively chosen range which excludes the impact of values outside the range on the overall sensitivity. PATIENTS AND PARTICIPANTS: The concepts of the refined method were illustrated using a Markov model of depression. MAIN OUTCOME MEASURES AND RESULTS: The application of the refined method substantially changed the ranking of the most sensitive variables compared with the original method. The response rate became the most sensitive variable instead of the 'per diem' for hospitalisation. CONCLUSIONS: The refinement of the original method yields sensitivity outcomes, which greater reflect the real uncertainty in economic studies.

## INTRODUCTION

Sensitivity analysis is currently the most widely used method to deal with uncertainty in economic evaluations. A sensitivity analysis is based on modification of the basic clinical and economic estimates of parameters to judge the effect on study results of alternative assumptions for the range of potential values for uncertain parameters.[1] Most recent pharmacoeconomic publications contain sensitivity analyses on only a limited number of variables, without justification of the choice of selected variables or the chosen range of each variable. In addition, the interpretation of the results of a sensitivity analysis remains subjective, as there is no scale for measurement of sensitivity and, consequently, there is no threshold value for discriminating between "sensitive" and "non sensitive" variables.

In a previous study, we described an objective method we developed for use in sensitivity analyses which minimizes the amount of potential subjectivity.[2] This new procedure, employing point-sensitivity and range-sensitivity, allows an objective judgment of the sensitivity of all variables of a model, permitting the variables to be ranked according to the degree of sensitivity. This approach avoids the subjective selection of variables for the sensitivity analysis and the potential bias in judging the degree of sensitivity in most current economic studies. Additionally, the determination of relative overall sensitivity enables an objective comparison between the impact of a variable in 2 different studies. Finally, the total level of uncertainty of a model may be determined by summation of the relative overall sensitivity of all variables.

**Figure 1. The Markov state model for depression.**



The aim of the present study is to describe a refinement of the strategy by the incorporation of probability distributions in the original study. As in the previous study, concepts are illustrated using a Markov model which determines the economic outcomes of a maintenance treatment with selective serotinin reuptake inhibitors (SSRIs) as first choice short term treatment, then continued as a maintenance treatment in responders over a 1-year follow up period. Figure 1 shows the overall structure of this model, and the clinical and economic data used to construct the model are summarised in Table I. Further details on the structure of the model and the sources of the clinical and economic data as well as the fixed input values for the model, are given in a previous paper.[2] However, in the present study the distribution of variables are not based on real data, but rather have been chosen to illustrate the relevant issues associated with the incorporation of probability distributions.

## MEASUREMENT OF SENSITIVITY

### Summary of Concepts

Point-sensitivity analysis measures the responsiveness of the outcome to the input value at a fixed point and does not include any associated range of the input. In the original study, the definition of point-sensitivity was based on the concept of point-elasticity (equation 1) where $\eta_s$ refers to point-elasticity and $X$ and $Y$ represent the fixed input value and the outcome, respectively.

$$\eta_s = dY/dX \cdot X/Y \qquad \text{(Eq. 1)}$$

In the case of a linear sensitivity relationship (Eq. 2) a more simple formula (Eq. 3) can be used:

$$Y = a \cdot X + b \qquad \text{(Eq. 2)}$$

$$\eta_s = a \cdot X/(a \cdot X + b) \qquad \text{(Eq. 3)}$$

where $a$ and $b$ are constants

Range-sensitivity analysis measures the difference in outcome when an input value is varied within its range; this difference depends on the range of a variable and the slope. This can be expressed as:

$$\Delta Y = a \cdot \Delta X \qquad \text{(Eq. 4)}$$

where the slope $a$ is derived from the appropriate sensitivity curve. In this equation $\Delta X$ represent the range of the input value.

Point-sensitivity and range-sensitivity were integrated in the previous study into an overall measure of sensitivity by using the concept of an integral of point-sensitivity between the lower and upper limit of the range of the output. This can be expressed as:

$$AOS = \left[ Y - b \times Ln(Y) \right]_{\text{lower limit}}^{\text{upper limit}} \qquad \text{(Eq. 5)}$$

where $AOS$ is the absolute overall sensitivity of an input variable and $Ln$ is the natural logarithm.

Table I. Clinical and economical data: input values and range.

| Variable* | Input values | | |
|---|---|---|---|
| | fixed | lower range | upper range |
| **Clinical probabilities** | | | |
| Response rate to first-line treatment with an SSRI (p1) | 0.60 | 0.49 | 0.70 |
| Response rate to TCA after SSRI treatment failure (p2) | 0.30 | 0.25 | 0.35 |
| Response rate to another TCA after earlier TCA treatment failure (p3) | 0.30 | 0.10 | 0.50 |
| Response rate to hospitalisation (p4) | 0.50 | 0.10 | 0.90 |
| Relapse rate during maintenance treatment with SSRI (r1) | 0.04 | 0.02 | 0.10 |
| Relapse rate during no treatment after response to a TCA (r2) | 0.10 | 0.00 | 0.50 |
| **Economic (SUS)** | | | |
| Daily treatment cost of SSRI (DTC-SSRI) | 2 | N.A.** | N.A. |
| Daily treatment cost of TCA (DTC-TCA) | 0.19 | N.A. | N.A. |
| Consultation tariff (CT) | 50 | 10 | 90 |
| Per diem tariff (pd) | 500 | 200 | 800 |

Abbreviations: SSRI = selective serotonine reuptake inhibitor; TCA = tricyclic antidepresssant

* Further detail on data sources is given in a previous article (Nuijten 1997)[2].

**: N.A. : not applicable.

## Rationale for Adjustment of the Original Method

The formula (Eq. 5) for the determination of the *AOS* shows that this measure of sensitivity assumes a uniform distribution for the outcome (*Y*). This relationship holds if the input value (*X*) of a variable has a uniform distribution according to the relationship defined in the Eq. 2:

$$Y=a\cdot X+b \qquad \text{(Eq. 2)}$$

However the real distribution of the input variable may correspond with a variety of probability distributions. Table II shows the real distribution and the rational behind the determination of the fixed input value and the range for each variable: these were the values used for the determination of overall sensitivity in the previous study.[2] The impact of the assumption of uniform distributions of the outcome (*Y*) on the overall sensitivity is shown in figure 2, where outcome is total direct medical costs associated with treatment of depression over a period of 1 year following initial medication SSRIs. The distribution of the outcome (*Y*) corresponding to a real distribution of *p1* (the response rate to a first choice treatment with an SSRI) and that corresponding to a uniform distribution and range of *p1* (on which the measurement of the overall sensitivity was previously based) are depicted. On this basis, the assumption of a uniform distribution will lead to an underestimation of the point-sensitivity for values near the mean (*p1*=0.6; *Y*=13,813), whereas this assumption will lead to an overestimation of the point-sensitivity near the lower (*p1*=0.5; *Y*=16,637) and upper value (*p1*=0.7; *Y*=10,987) of the range. Hence the original method for the determination of the overall sensitivity should be adjusted to the real probability distribution by weighting the different intervals.

The second limitation of *AOS* (Eq. 5) was that the overall sensitivity measure was based on a subjectively chosen range, which excludes the impact of values outside the range on the *AOS*. In the original analysis, the range for *p1* was based on a 90% confidence interval (CI), where the lower limit was >5% and upper limit was <95%, while the range for the *per diem* hospitalisation values (*pd*) was based on the absolute minimum and maximum values of a sample of *per diem* values (n=200). In the original analysis, *pd* was a more sensitive variable than *p1*: the *AOS* of *pd* and *p1* was 14,602 and 7,114, respectively. However, an extension of the range of *p1* from a 90% to a 98% CI more than doubles the *AOS* from 7,114 to 17,073, and consequently *p1* becomes a more sensitive variable than *pd* (7,114>14,602). Hence, in the case of a continuous distribution (e.g. normal distribution) without a real minimum and maximum, the *AOS* of a variable depends on the choice of the CI. In order to solve the impact of the subjectively chosen range, a correction for the weight of each interval should be applied. This can also be used to determine an appropriate CI (e.g. 99%, 99,9%), which approximates the hypothetical 100% interval.

**Table II. Clinical and economical input data for the model: determination of fixed input values and range, data source and distribution.**

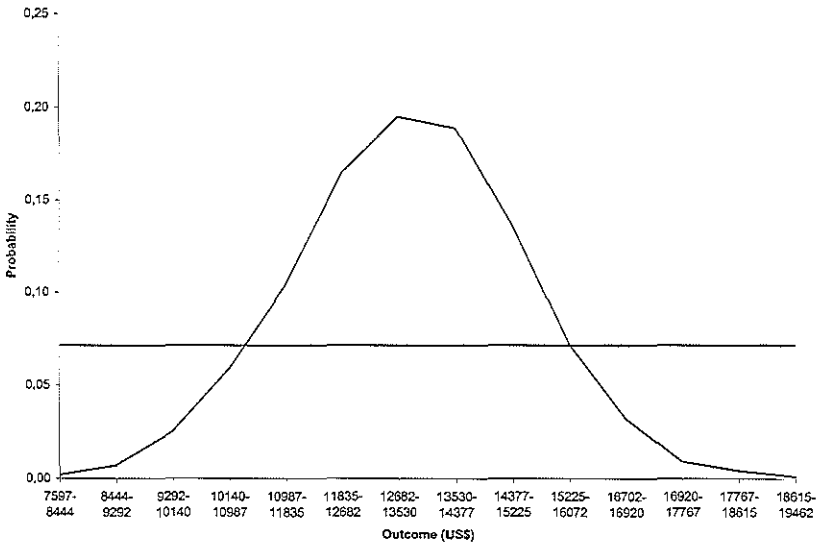
| Variable | Input values | | Data source | Distribution |
|---|---|---|---|---|
| | Fixed | range | | |
| **Clinical probabilities** | | | | |
| p1 | mean | 5-95% C.I.* | meta-analysis | Normal |
| p2 | mean | 5-95% C.I. | meta-analysis | Normal |
| p3* | mean | Q1-Q3** | meta-analysis | not known |
| p4 | median | Q1-Q3 | Delphi | not known |
| r1 | mean | Minimum-maximum | clinical trial | Discrete |
| r2 | mean | Minimum-maximum | clinical trial | Discrete |
| **Economic ($US)** | | | TLG costing database**** | |
| DTC-SSRI | price | N.A.*** | price list (n=1) | N.A. |
| DTC-TCA | price | N.A. | price list (n-1) | N.A. |
| CT | mean | Minimum - maximum | tariff list (n=10 tariffs) | Discrete |
| pd | mean | Minimum - maximum | per diem list (n=200 hospitals) | Discrete |

* **C.I.: confidence interval.**
** **Q1-Q3: first and third quartile**
*** **N.A.: not applicable.**
**** **TLG Costing Database.**

## METHODS

The details of the refinement (incorporation of probability distributions) to the previously described method of reducing uncertainty in economic studies based on the concept of point-sensitivity (see Measurement of Sensitivity section) are described in this section. The concepts are illustrated by using the depression Markov model (figure 1).

**Figure 2. The extrapolated distribution of the outcome (US$) for *p1*.**




## General Methodology for the Adjustment

This section presents a mathematical method to correct for the real distribution of an input variable. A correction consists of the following steps: (i) the probability distribution of outcome Y is extrapolated from the real probability distribution for input value *X*, (ii) the total area under the point-sensitivity/outcome curve is divided into a number of intervals; (iii) the *AOS* is determined for each interval; (iv) a total weighted *AOS* is determined by summation of the weighted *AOSs* of all intervals and; (v) the total weighted *AOS* is divided by the width of the intervals. This correction is performed for each input variable, and subsequently the input variables can be ranked according to the adjusted absolute overall sensitivity. The remainder of this section describes the mathematical approach to determine the adjusted *AOS* based on the previous steps.

## Refined Mathematical Method for Determining the Adjusted Absolute Overall Sensitivity

The formula for overall sensitivity (Eq. 5) was described above sensitivity (see Measurement of Sensitivity section). In order to divide the total area under the curve into a number of intervals, Eq. 5 is divided into the following parts:

$$AOS=\left[Y-b\cdot Ln(Y)\right]_{Y0}^{Y0+d}+\left[Y-b\cdot Ln(Y)\right]_{Y0+d}^{Y0+2d}+\ldots.\left[Y-b\cdot Ln(Y)\right]_{Y0+(n-1)d}^{Y0+nd} \quad \text{(Eq. 6)}$$

Where *Y0* and *Y0+nd* correspond with respectively the lower and upper limit, *d* is width of each interval and *n is* number of intervals. Within each interval, we assume that there is a uniform distribution.

In Eq. 6, the contribution of every part to the total *AOS* is similar, because of the assumption that the distribution of *Y* is uniform within its range. In Eq. 7 the *AOS* of each interval is weighted according to its relative frequency.

$$AOS=\frac{f1}{fu}\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0}^{Y0+d}+\frac{f2}{fu}\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0+2\cdot d}^{Y0+n\cdot d}+\ldots\ldots\ldots\frac{f}{fu}\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0r}^{Y0+n\cdot d}+ \quad \text{(Eq. 7)}$$

Where *fu* is frequency of each interval according to a uniform distribution, which equals $1/n$ and where *n* is the number of intervals and *fo, f1....fn* correspond to the frequency of each interval derived by extrapolation from the real distribution of input *X* by using the formula, Y=a.X+ b. In the case of a uniform distribution all intervals have the same frequency, fu=1/n and in this case Eq. 7 can be simplified to Eq. 6, and subsequently to Eq. 5.

Eq. 7. shows that the *AOS* depends on the range of the distribution, which determines the number of intervals *n* and consequently, fu=1/n. A correction consists in dividing *AOS* by the range, which equals product of number of intervals and with of each interval ($n \cdot d$), which is shown in Eq.8.

$$AAOS=\left\{f1\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0}^{Y0+d}+f2\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0+2\cdot d}^{Y0+n\cdot d}+\ldots f\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0r}^{Y0+n\cdot d}\right\}\frac{n}{nd} \quad \text{(Eq. 8)}$$

Subsequently Eq. 8 can be simplified to Eq.9.

$$AOS=\left\{f1\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0}^{Y0+d}+f2\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0+2\cdot d}^{Y0+n\cdot d}+\ldots\ldots f\cdot\left[Y-b\cdot Ln(Y)\right]_{Y0r}^{Y0+n\cdot d}\right\}/d \quad \text{(Eq. 9)}$$

Where *AAOS* is the adjusted *AOS* according to the refined method. In addition, an adjusted relative overall sensitivity *(AROS)* can be determined by dividing the *AAOS* by the basic average outcome of the model. The advantage is that this *AROS* allows comparison between different studies (as long as they have similar outcomes). Finally, an overall level of uncertainty of a model may be determined by summation of the *AROS* values for all variables.

In the original method[2] the upper and lower limit of each input variable were subjectively chosen, excluding the impact of values outside the range on the overall sensitivity. An adjustment to the real probability distribution will lead to an extension of the range for each input variable. In the case of a distribution with a real minimum and maximum, the upper and lower limit will be based on those values. In the case of a continuous distribution (e.g. normal distribution) without a real minimum and maximum, the upper and lower limit will be extended to areas of the probability distribution, which can be assumed not to contribute to the *AAOS*. In this refined method, the initial range was based on a 98% CI, where the lower limit was >1% and the upper limit was < 99%. An additional sensitivity analysis was performed by extending the range in order to test the appropriateness of the chosen CI. When an extension of the range only marginally increases the *AAOS*, we may assume that the calculated *AAOS* approximates the *AAOS* of the complete distribution.

**Table III. The outcomes of the refined method for determination of the *AAOS* for p1. (Y=a X+b; Y, where a= -28250 and b=30762, d=848).**

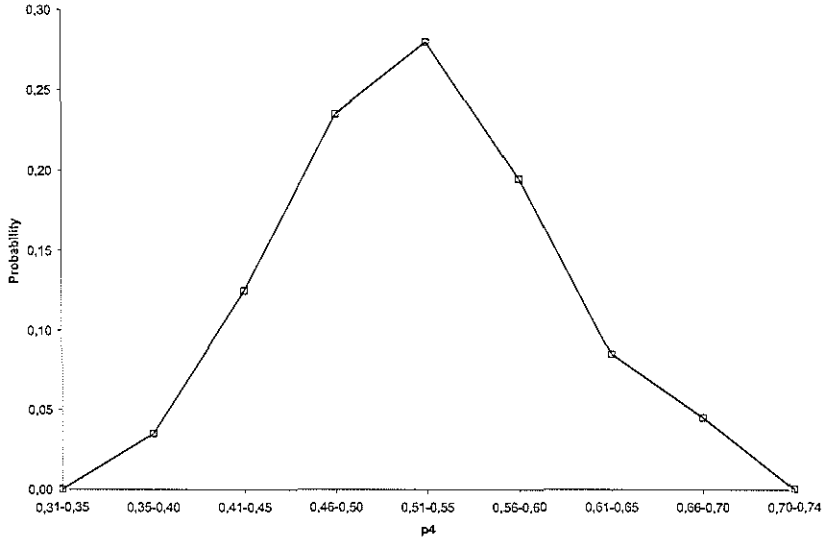
| Interval X | Interval Y | Frequency | AOS | AAOS |
|---|---|---|---|---|
| 0.40-0.43 | 18615- 19462 | 0.002 | 522 | 1 |
| 0.43-0.46 | 17767- 18615 | 0.007 | 586 | 4 |
| 0.46-0.49 | 16920- 17767 | 0.024 | 656 | 16 |
| 0.49-0.52 | 16072- 16920 | 0.059 | 733 | 43 |
| 0.52-0.55 | 15225- 16072 | 0.104 | 819 | 85 |
| 0.55-0.58 | 14377- 15225 | 0.161 | 914 | 147 |
| 0.58-0.61 | 13530- 14377 | 0.191 | 1022 | 195 |
| 0.61-0.64 | 12682- 13530 | 0.184 | 1142 | 210 |
| 0.64-0.67 | 11835- 12682 | 0.131 | 1280 | 168 |
| 0.67-0.70 | 10987- 11835 | 0.072 | 1438 | 104 |
| 0.70-0.73 | 10140- 10987 | 0.027 | 1622 | 44 |
| 0.73-0.76 | 9292- 10140 | 0.009 | 1838 | 17 |
| 0.76-0.79: | 8444- 9292 | 0.004 | 2095 | 8 |
| 0.79-0.82 | 7597- 8444 | 0.002 | 2406 | 5 |
| Sum | | 0.990 | 17073 | 1046.6 |

## Bootstrap Method

The above mentioned method enables the assessment of the *AOS* for variables with a probability distribution. However, there may be variables in the model for which no probability distribution can be determined because of a limited amount of data. Table II contains a description of the real probability distribution of each input variable, except for the response rate to hospitalisation (p4), which was derived from a 20-member Delphi panel. Delphi panels generally consist of a limited number of members, varying from 5 to 20, the determination of a probability distribution and a CI using traditional statistical methods is not possible. The most common approach is to use the median as an input variable and use a range corresponding to the first and third quartile for the standard sensitivity analysis (range sensitivity). However, this method does not allow the creation of a probability distribution, which is required for the measurement of the adjusted overall sensitivity, based on the method described above.

The so-called bootstrap method allows the creation of an artificial probability distribution, which subsequently can be used to assess the adjusted overall sensitivity. This can be compared with the adjusted overall sensitivity of the other input variables. The bootstrap approach uses the original sample data to generate an empirical distribution. A sample is drawn from the original sample with replacement, where each observation has an equal probability of being drawn on any occasion. The size of the sample drawn is equal to the size of the original sample and may contain some duplicated observations, while other observations may be omitted altogether. This new bootstrapped sample is then used to calculate a mean. Another bootstrap sample is drawn from the original sample and another mean is calculated. Generally it is accepted that 50 to 200 samples should be sufficient for estimating the mean and standard deviation.[3]

**Figure 3. The probability distribution for the input variable response rate to hospitalisation (*p4*) determined using the bootstrap method.**




# RESULTS

## Specific Results

Table III shows the application of the refined method for input variable *p1*, the response rate to first choice treatment with an SSRI. The probability distribution of *p1*, was extrapolated to Y using the formula, Y=a·X+b, where a=-28,250 and b=30,762. The width of each interval for *X* and extrapolated *Y* are 0.003 and 847, respectively. The CI of the chosen range equals 98, which was based on a lower limit of > 1% and an upper limit of < 99%. The *AAOS* was determined for each interval by assuming that a uniform distribution exists within each interval due to its small width. Consequently Eq.9 was used to determine the *AAOS:*

AAOS={0,002·522+0,007·586+0,024·656+0,059·733+0,104·819+0,161·914+0,191·1022+0,184·1142 +0,131·1280+0,072·1438+0,027·1622+0,009·1838+0,004·2095+0,002·2406}/847=1046,6

The *AROS* was determined by dividing this value by the average outcome (*Y*), which leads to a value of 0.075772. Subsequently, a sensitivity analysis was performed on the CI by an extension of the CI from 98% to 99%, which showed that an adequate CI was chosen in the standard analysis: the *AAOS* only increased 3,2% from 1046.6 to 1079.8. Hence, the *AAOS* may be considered an approximation of the real *AAOS* of a hypothetical 100% CI.

The bootstrap method was applied to input variable *p4*, which was based on Delphi panel estimations and consequently does not have a probability distribution. Figure 3 shows the probability distribution of *p4*, determined using a bootstrapping procedure, based on a sample of 20 estimations and 200 resamples.

**Table IV. The overall absolute sensitivity for all variables in the model for depression.**

| Variable | Point-sensitivity | Original Range | AOS | New Range | AOS | AAOS | AROS | Ranking* |
|---|---|---|---|---|---|---|---|---|
| p1 | 1.23 | C.I.:5-95% | 7348 | C.I.: 1-99% | 17073 | 1,2342 | 0.00008935 | 1-2 |
| p4 | 0.45 | Q1-Q2 | 5045 | Q1-Q2 | 29063 | 0.9551 | 0.00006914 | 2-3 |
| pdicm | 0.94 | Min-max | 14602 | min-max | 14602 | 0,9363 | 0.00006778 | 3-1 |
| p3 | 0.40 | Q1-Q3 | 3266 | C.I.: 1-99% | 6183 | 0,4358 | 0.00003155 | 4-4 |
| p2 | 0.39 | C.I.:5-95% | 173 | C.I.: 1-99% | 1583 | 0,4084 | 0.00002957 | 5-5 |
| r1 | 0.06 | min-max | 144 | min-max | 150 | 0.0600 | 0.00000434 | 6-6 |
| CT | 0.02 | min-max | 12 | min-max | 12 | 0.0227 | 0.00000165 | 7-8 |
| r2 | 0.01 | min-max | 30 | min-max | 30 | 0.0134 | 0.00000097 | 8-7 |
| DTC-SSRI | 0.03 | N.A. | 0 | N.A. | 0 | 0 | 0 | 9-9 |
| DTC-TCA | 0.00 | N.A. | 0 | N.A. | 0 | 0 | 0 | 10-10 |
| Sum | | N.A. | 30620 | | 68696 | 4,0658 | 0.00029436 | |

* **Ranking according to refined and original method.**
** **Actual range corresponds with the chosen range.**

### Overall Results

After application of the refined method for all input variables, the variables can be ranked according to their *AAOS* and/or *AROS* (Table IV). The variables *DTC-SSRI* and *DTC-TCA* were ranked according to their point-sensitivity, because the overall sensitivity for fixed drug prices is zero. The total adjusted relative sensitivity of the model was 0.00029436, which can be considered a measure of the overall uncertainty of the model. Table IV shows that the use of the refined method considerably changed the ranking of the variables. The most sensitive variable according to the original method, *pd*, became less sensitive than *p4* and *p1*, which became the most sensitive variable of the model (*p1*). In addition, consultation tariff (CT) became more sensitive than the relapse rate during no treatment after response to a TCA (*r2*).

## DISCUSSION

In a previous publication by me,[2] an alternative method was presented to deal with uncertainty in economic studies, based on the concept of point-sensitivity. The concept was illustrated using a

modelling study, but may also be applied to other types of economic evaluations. The method allows a more objective judgment of robustness of a model by avoiding the subjective selection of variables for the sensitivity analysis and the potential bias in judging the degree of sensitivity in most current economic studies. However, the first limitation of this method was that the sensitivity measurement was based on a uniform distribution of the variables, which in reality may have different distributions. The second limitation was that the overall sensitivity measure was based on a subjectively chosen range, which excludes the impact of values outside the range on the overall sensitivity.

The present study describes a refinement of the original method by the incorporation of probability distributions, which allow a more accurate assessment of the level of uncertainty in the model. In addition, the bootstrap method creates probability distributions for fixed input variables that are based on a limited quantity of data. A technical limitation is that determination of overall sensitivity for a variable is based on fixed input variables for all other input variables. Hence, the method may underestimate the real uncertainty. Finally, one needs to consider that there are currently no sensitivity data available, which are based on the discussed method in this paper. Hence, there is not yet a scale for grading the level of uncertainty. Hence, the interpretation of a total overall sensitivity outcome for a model or sensitivity outcomes for individuals variables may be difficult at the moment. Currently, the method can already be used to compare uncertainty between different variables in one study and total uncertainty between studies.

## CONCLUSION

The application of this refinement substantially changed the ranking of the most sensitive variables, as illustrated using a depression model. Hence, the changes observed in the depression model show that the refinement of the procedure for estimating sensitivity may have important implications for the interpretation and analysis of health economic assessments

### References

1. Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health care Technology: A Report on Principles. Ann Intern Med. 1995; 122: 61-70
2. Nuijten MJC, Hardens M. Measuring sensitivity in pharmacoeconomic studies: an integration of point-sensitivity and range-sensitivity. Pharmacoeconomics. 1997 Nov; 12(5): 555-564
3. Mooney CZ, Duval RD, 1993, Bootstrapping: A nonparametric approach to statistical inference (Sage University Paper Series on Quantitative Applications in the social sciences, series no. 07-095) Newbury Park (CA): Sage University, 1993

Chapter 10

# Inter-Variable Uncertainty in Decision-Analytic Modelling: The Concept of Second-Order Sensitivity


## SUMMARY

The constraint of the current methods of sensitivity analysis in decision-analytic models is that those methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taken into account the existing relationships between those variables. The present study presents various methods, which considers this type of inter-variable uncertainty for economic evaluations based on modeling techniques. We presented initially a method assuming only uniform distributions, and subsequently various methods incorporating the real distributions. The results show that this second-order sensitivity of a dependent input variable depends very much on the distribution of the independent input variables. Overall our analysis showed that the most sensitive variables for the outcome of the model, were also the most sensitive for the 2-order sensitivity of the other input variables. Although a more general mathematical prove is required to support those conclusions, we ended with general recommendations for execution of a sensitivity analysis, including the results of a 2-order sensitivity analysis.

## INTRODUCTION

There is growing evidence that health economic data is beginning to be used more widely to support decision makers in the health service for allocation of scare healthcare resources.[1] The most evident impact of health economic studies is expected for reimbursement audiences because of the elaboration of economic reporting requirements. If economic evaluation is to be used for pricing and reimbursement issues by authorities or third party payers, it becomes important for the different stake holders (industry, government, health insurance) that these analyses are being performed according to generally accepted and standardised methods. The handling with uncertainty in economic evaluation is an important area that remains relatively underdeveloped. Sensitivity analysis is currently the most widely used method to deal with uncertainty in economic evaluations. A sensitivity analysis is based on modification of the basic clinical and economic estimates of parameters to judge the effect on study results of alternative assumptions for the range of potential values for uncertain parameters.[2] The methods used, the choice of parameters and the range of these parameters must be stated and substantiated. If this procedure is followed for all the estimates in turn, then we refer to it as a 'univariate sensitivity analysis'. In a multivariate sensitivity analysis, the effect of simultaneous changes in different variables is examined on the outcomes of the study. Probabilistic sensitivity analysis is the most advanced method, which tries to obtain random distributions around each parameter and -- in conformity with the distributions -- then tries to arrive at a new estimate for each parameter. For each combination of estimates a new estimate of the costs, effects and cost-effectiveness ratio is therefore obtained. By repeating this procedure many times a random distribution can then be presented based on the results of the study.

**Figure 1. Markov states and possible transitions used in the depression model.**

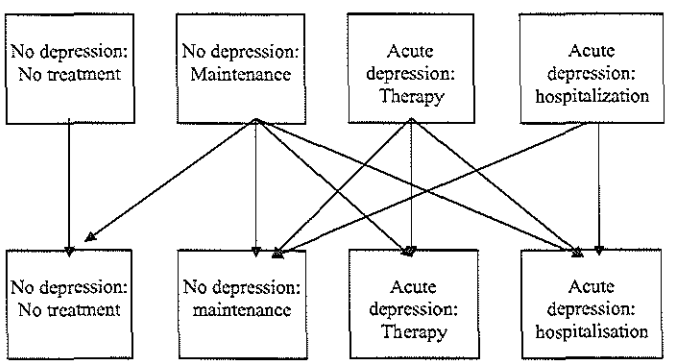


**Table I. Clinical and economical data: input values and range.**

| Variable* | Input values | | |
|---|---|---|---|
| | **Fixed (SD)** | **Range** | **Distribution** |
| **Clinical probabilities** | | | |
| Response rate to first-line treatment with an SSRI (p1) | 0.60 (0.05) | 0.50-0.70 | Normal |
| Response rate to TCA after SSRI treatment failure (p2) | 0.30 (0.03) | 0.25-0.35 | Normal |
| Response rate to another TCA after TCA treatment failure (p3) | 0.30 (0.03) | 0.25-0.35 | Normal |
| Response rate to hospitalisation (p4) | 0.50 (0.04) | 0.41-0.58 | Normal |
| Relapse rate during maintenance treatment with SSRI (r1) | 0.04 (0.015) | 0.02-0.06 | Normal |
| Relapse rate during no treatment after response to a TCA (r2) | 0.10 (0.033) | 0.07-0.16 | Normal |
| **Economic ($US)** | | | |
| Daily treatment cost of SSRI (DTC-SSRI) | 2 (0) | N.A.** | N.A. |
| Daily treatment cost of TCA (DTC-TCA) | 0.19 (0) | N.A. | N.A. |
| Consultation tariff (consult) | 50 (23.4) | 10-90 | Uniform |
| Per diem tariff (diem) | 500 (173) | 200-800 | Uniform |

Abbreviations: SSRI = selective serotonine reuptake inhibitor; TCA = tricyclic antidepresssant

* Further detail on data sources is given in a previous article (Nuijten 1997). Distributions of p, r1 and r2 have been changed compared with previous paper.

**: N.A. : not applicable.

The constraint of the above-mentioned methods for sensitivity analysis in decision analytic models is that those methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taken into account the existing relationships between those variables. The objective of this manuscript is to present various experimental methodologies, which considers this type of inter-variable uncertainty for economic evaluations based on decision analytic techniques. The concepts are illustrated using a Markov model, which compares the economic outcomes of a maintenance treatment with selective serotinin reuptake inhibitors (SSRIs) as the first choice therapy for depression over a one year follow up period. Figure 1 shows the overall structure of this model and the clinical and economic data used to construct the model are summarised in Table I. Further details on the structure of the model and the sources of the clinical and economic data as well as the fixed input values for the model, are given in previous papers.[3 4]

# MEASUREMENT OF SENSITIVITY

## Summary of Concepts

Sensitivity analysis is based on the determination of the difference in a clinical or economic outcome when an input value is varied within its range and depends only on the range of a variable. This can be expressed as:

$$S_{V_1}=\Delta Y=a\cdot\Delta V_1 \qquad \text{(Eq. 1)}$$

where $S_{v1}$ is sensitivity of variable $v_1$, the slope $a$ is derived from the appropriate sensitivity curve. In this formula the outcome is either the effectiveness or the economic outcome associated with an initial treatment over a period of follow-up. In our example the outcome was the total medical costs over as period of 1 year. The assumption was that there is linear relationship between the outcome and the input values, which was true for the depression model. In this equation $\Delta Y$ and $\Delta V_1$ represent the range of respectively the output and input values. The *per diem* (pd) was the most sensitive variable using the presented method for sensitivity analysis. The sensitivity of an input variable in current sensitivity analyses only depends on the range of that input variable $v_1$, because the coefficient $a$ seems a constant variable according to Eq. 1. However the coefficient $a$ is actually dependent of all other variables in the model, excluding $v_1$.

$$a_1=F(v_2,v_3,\ldots..,v_n) \qquad \text{(Eq. 2)}$$

Where coefficient $a_1$ of variable $v_1$ is a function of variables $v_2,v_3 \ldots v_n$.

Consequently:

$$S_{V_1}= F(a_1;dV_1)=F(v_2,v_2,\ldots..,v_n;dV_1) \qquad \text{(Eq. 3)}$$

Hence the sensitivity of $v_1$ depends on the range of $v_1$ ($dV_1$) and the fixed input values of all other variables in the model ($v_2,v_2,\ldots.v_n$). Thus the constraint of the standard sensitivity analysis is that it shows the sensitivity of the outcomes to a change through a range of potential values for $v_1$ without taken into account the existing relationships between this measure of sensitivity and the other variables ($v_2,v_2,\ldots.v_n$). The present study presents various methodologies, which considers this type of inter-variable uncertainty for economic evaluations based on modeling techniques.

**Table II. The uniform 2-order sensitivity for all variables in the model for depression.**

| | p1 | p2 | p3 | p4 | consult | diem | r1 | r2 |
|---|---|---|---|---|---|---|---|---|
| $\Delta S_V$ | | | | | | | | |
| **p1** | 0 | 797 | 3298 | 4686 | 20 | 6927 | 554 | 466 |
| **p2** | 797 | 0 | 1028 | 1408 | 6 | 2159 | 244 | 173 |
| **p3** | 3298 | 1026 | 0 | 5628 | 3 | 8916 | 977 | 501 |
| **p4** | 4687 | 1408 | 5628 | 0 | 2 | 11822 | 573 | 9 |
| **consult** | 19 | 6 | 3 | 2 | 0 | 0 | 9 | 53 |
| **diem** | 6927 | 2159 | 8927 | 11822 | 0 | 0 | 2048 | 1076 |
| **r1** | 554 | 245 | 975 | 572 | 6 | 2111 | 0 | 3 |
| **r2** | 467 | 174 | 501 | 8 | 53 | 1075 | 3 | 0 |
| **Δa** | | | | | | | | |
| **p1** | 0.0 | 3985.0 | 16490.0 | 23430.0 | 100.0 | 34635.0 | 2770.0 | 2330.0 |
| **p2** | 7970.0 | 0.0 | 10280.0 | 14080.0 | 60.0 | 21590.0 | 2440.0 | 1730.0 |
| **p3** | 32980.0 | 10260.0 | 0.0 | 56280.0 | 30.0 | 89160.0 | 9770.0 | 5010.0 |
| **p4** | 27898.8 | 8381.0 | 33500.0 | 0.0 | 11.9 | 70369.0 | 3410.7 | 53.6 |
| **consult** | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.7 |
| **diem** | 11.5 | 3.6 | 14.9 | 19.7 | 0.0 | 0.0 | 3.4 | 1.8 |
| **r1** | 2359.5 | 1043.4 | 4152.5 | 2436.1 | 25.6 | 8990.6 | 0.0 | 12.8 |
| **r2** | 3492.9 | 1301.4 | 3747.2 | 59.8 | 396.4 | 8040.4 | 22.4 | 0.0 |

# METHODS

## Uniform Distributions

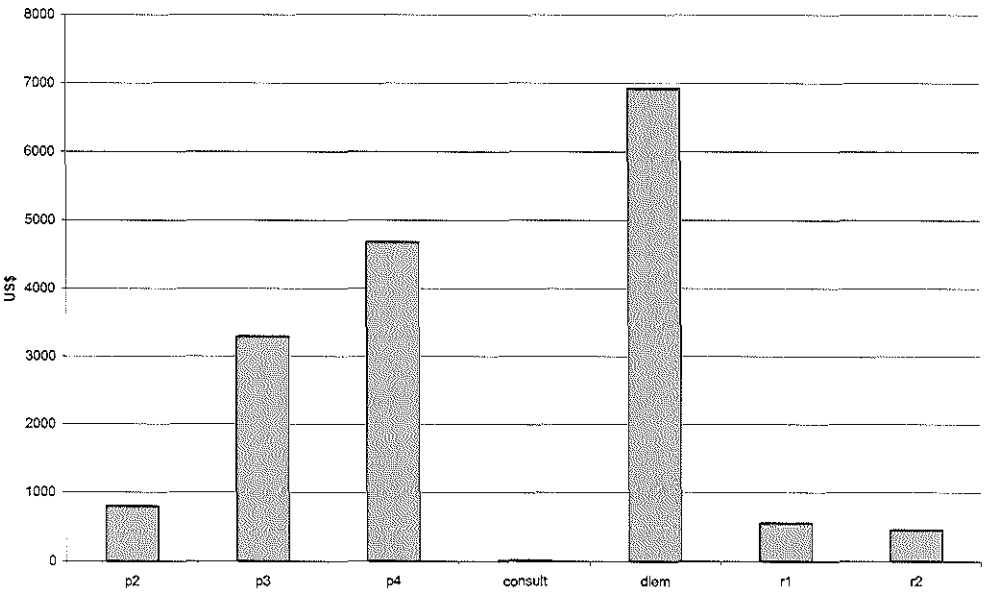
This section presents a mathematical method to determine the dependence of the sensitivity of a variable on other input variables. Initially we assume that all variables have a uniform distribution with an absolute lower and upper limit (Table I). The sensitivity of $v_1$ is determined at the lower value of another input variable $v_2$ ($S_{v_{12min}}$) and subsequently at the upper limit of $v_2$ ($S_{v_{12max}}$). A measure of the impact of $v_2$ on the sensitivity of $v_1$ is the absolute value of the difference ($\Delta S_{v_{12}}$) between the sensitivity of $v_1$ at the lower limit of $v_2$ and the sensitivity at the upper limit of $v_2$:

$$\Delta S_{v12} = \text{abs}\,(S_{v12max} - S_{v12min}) \qquad \text{(Eq. 4)}$$

Measures, which show the "sensitivity of the sensitivity", are defined as second-order sensitivity for the remainder of this article. Consequently the traditional sensitivity of the outcome of a model to change of a variable is defined as a 1-order sensitivity. Figure 2 shows the outcomes for *p1* to the other

input values of the mode: this analysis shows that the sensitivity of *p1* is most sensitive to the per diem. The results for all variables are shown in Table II.

**Figure 2. The 2-order sensitivity of p1 to the other variables in the model.**




A disadvantage of the above mentioned method is that the second-order sensitivity of variable $v_1$ to $v_2$ depends not only on $v_2$, but also on $\Delta v_1$, which is shown in the following equations:

$$S_{v12max}=a_{max}\,\Delta V_1 \qquad \text{(Eq. 5)}$$

$$S_{v12min}=a_{min}\,\Delta V_1 \qquad \text{(Eq. 6)}$$

$$\Delta S_{v12}=(a_{max}-a_{min})\,\Delta V_1=\Delta a\,\Delta V_1 \qquad \text{(Eq. 7)}$$

Consequently, the second-order sensitivity of $v_1$ to $v_2$ is better reflected by the following equation:

$$SS_{v12}=\Delta S_{v12}/\Delta V_1=\Delta a \qquad \text{(Eq. 8)}$$

Where $SS_{v12}$ is the uniform second-order sensitivity of $v_1$ to $v_2$. The results of this analysis are also shown in Table II, which shows that this adjustment does not change the relative impact of the other variables on the sensitivity of another variable (e.g., *p1*), because all second-order sensitivity values are divided by the same value (e.g., $\Delta p1$). An advantage of this measure of second order sensitivity is that it allows a comparison of the impact of one variable on the sensitivity of the other variables. Table II (Figures 3-a and 3-b) can be judged horizontally and vertically: 1) the rows show which variable has the highest impact on the sensitivity of a variable (e.g., the *per diem* has the highest

impact on *p1*) and 2) the columns show the impact of a variable on the sensitivity of other variables (e.g., *pdiem* has the highest impact on *p3*).

**Figure 3-a. The impact of the variables on the sensitivity for each variable.**

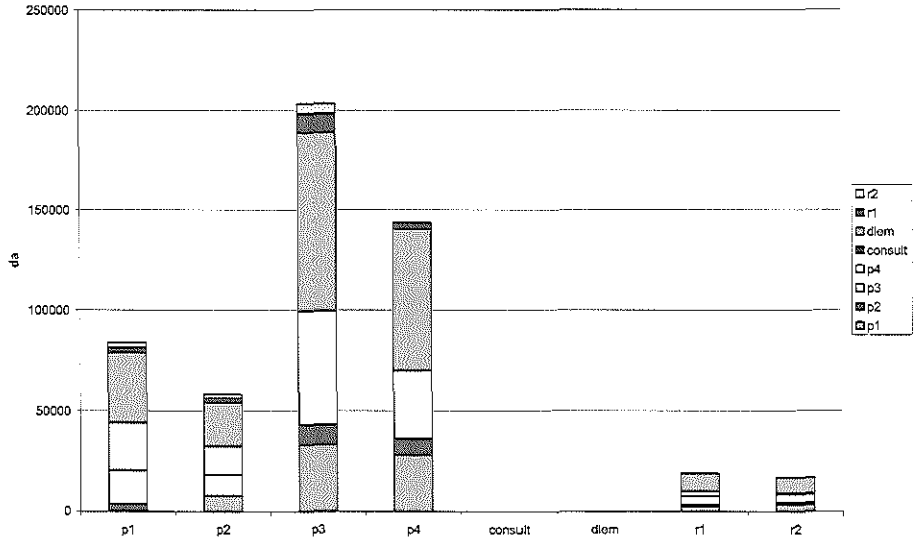
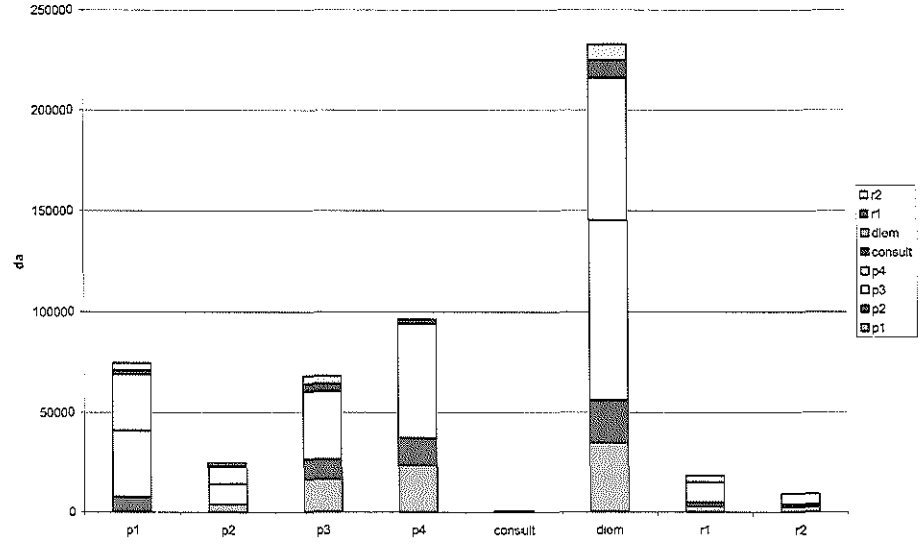



**Figure 3-b. The impact of each variable on the sensitivity of other variables.**



The results of this analysis show that the 2-order sensitivity of consultation does not depend on the per diem, while the per diem heavily influences the sensitivity of the other variables, which can be explained using the following example:

$C = C_1 + C_2 + .. C_k$

Where $C$ is the total cost and $C_1, C_2 \ldots C_k$ are the costs of $k$ pathways in the model, which are based on the following formulae:

$C_1 = P_1 (k_1 \cdot d + m_1 \cdot ce_1 + n_1 \cdot ce_2)$

$C_2 = P_2 (k_2 \cdot d + m_2 \cdot ce1 + n2 \cdot ce_2)$

$C_k = P_k (k_k \cdot d + m_k \cdot ce_1 + n_k \cdot ce_2)$

Where $ce_1$ is the cost of consultation and $ce_2$ is the cost of hospitalisation, which is based on the per diem. $P_1, P_2 \ldots P_k$ are based on a multiplications of all probabilities alongside each pathway ($k$, $m$) and $n$ correspond with number of respectively drug usage, consultations and hospitalisations along a pathway. Consequently:

$C = P_1 (k_1 d + m_1 ce_1 + n_1 ce_2) + P_2 (k_2 d + m_2 ce_1 + n_2 ce_2) .. + P_k (k_k + m_k ce_1 + n_k ce_2)$

We now determine sensitivity for variable $ce_1$:

$C_{min} = P1 (k_1 d + m_1 ce_{11} + n_1 ce_2) + P2 (k_2 d + m_2 ce_{11} + n_2 ce_2) .. + P_k (k_k + m_k ce_{11} + n_k ce_2)$

$C_{max} = P1 (k_1 d + m_1 ce_{12} + n_1 ce_2) + P2 (k_2 d + m_2 ce_{12} + n_2 ce_2) .. + P_k (k_k + m_k ce_{12} + n_k ce_2)$

$\Delta S_{v12} = C_{max} - C_{min}$

$= P_1 (m_1 ce_{11} - m_1 ce_{12} + P_2 (m_2 ce_{11} - m_2 ce_{12}) .. + P_k (m_k ce_{11} - m_k ce_{12})$

Consequently the 2-order sensitivity of $ce_1$ does not depend on $ce_2$, which is excluded from the last equation), while the difference between the two equations does not erase the other variables, because they are in multiplication relationship with consultation. Thus we may distinguish multiplication relationships and summation relationships in a model. The second-order sensitivity of a variable only depends on variables, which are in a multiplication relationship: 1) relationships between probabilities and costs and 2) relationships between probabilities and probabilities. Summation relationships only relate to relationships between costs and costs (prices as well as resource utilization). Hence the second-order sensitivity of consultation does not depend on the *per diem*, which is excluded from the last equation), while the difference between the two equations does not erase the other variables, because they are in multiplication relationship with consultation.

## Real Distributions

A limitation of the above-mentioned equation 1 is that it only based on the range of variables without taken into account the distribution of $v_1$: $a$ is a function of the distribution of variable $v_{2,3,4...k}$. Therefore the original method for the determination of the 2-order sensitivity will be adjusted to include the real probability distributions of all variables. We present various methods to measure the inter-variable relationships in the model for depression.

### *Method 1:Standard Deviation-Based Approach Using Linear Relationships*

The original equation (Eq. 1) is adjusted to:

$$S_m=a_m \cdot \Delta V_m \qquad \text{(Eq. 9)}$$

$S_m$ is the sensitivity of variable $V_m$, which depends on the range of variable $V_m$, but also on the value of $a_m$, which is a function of all other variables, excluding $V_m$.

The following equation is determined:

$$a_m=k \cdot v_n+l \qquad \text{(Eq. 10)}$$

Where $a_m$ is a function of variable $v_n$; $k$ and $l$ are the coefficients. In this example when $v_m$ is $p1$ and $v_n$ is *pdiem*, $k$ is –57,7 and $m$ is –605.

$$a_{p1}= -57{,}7 \cdot \text{pdiem} \qquad \text{(Eq. 11)}$$

Consequently $a$ at the fixed input value of *pdiem* of 500 equals –28245. Figure 4 shows that a linear relationship exists between the $a_{p1}$ and *pdiem*, which we also proved for the other relationships between the $a$ values and the input values: we also determined the corresponding values for $k$ and $l$ for all other variables. A measure of second-order sensitivity of $v_m$ to $v_n$ is the standard deviation of $a_m$ resulting from Eq. 12, which is a function of the standard deviation of $v_n$ *(SD-*$v_n$*)*.

$$SD\text{-}a_m= |k]\ SD\text{-}v_n \qquad \text{(Eq. 12)}$$

where *SD-*$a_m$ is the second-order sensitivity of variable $v_1$ to $v_2$.

An adjustment to our example leads to the following formula:

$$SD\text{-}a_{p1}= |k]\ SD\text{-}pdiem_n= 57{,}7 \cdot = 173$$

Thus the per diem leads to a standard deviation of 173, which reflects the second-order sensitivity of $a_{p1}$ to *pdiem*. The standard deviations of the other variables are shown in Table III, which show that *pdiem* has the highest impact on the sensitivity of $p1$ and all other variables, excluding consultation.

Those results also allow us to determine the impact of each variable on the sensitivity of the other variables; in this example the per diem has highest impact on *p3*.

**Figure 4. The relationship between $a_{p1}$ and pdiem.**

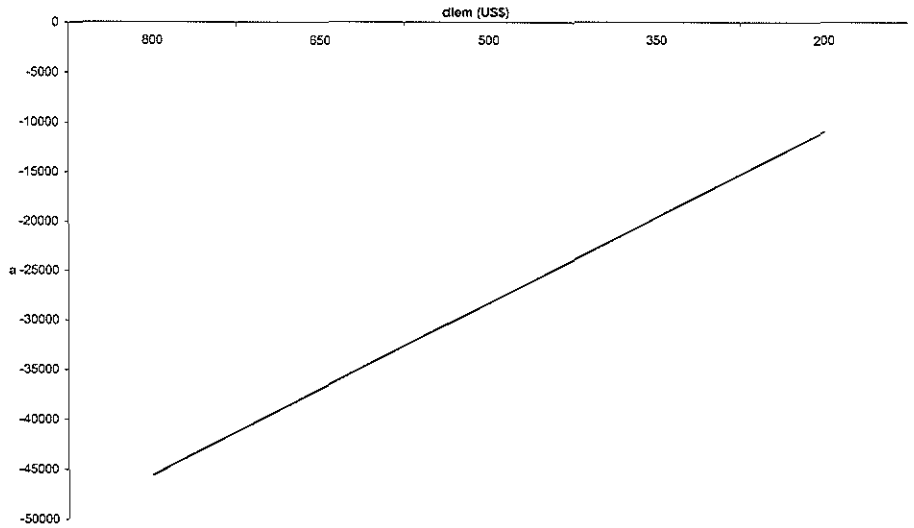



**Table III. The 2-order sensitivity for all variables based on the standard deviation of *a*.**

| | p1 | p2 | p3 | p4 | consult | diem | r1 | r2 |
|---|---|---|---|---|---|---|---|---|
| **SD** | | | | | | | | |
| **p1** | 0.05 | 0.03 | 0.03 | 0.04 | 23.4 | 173.0 | 0.015 | 0.033 |
| **p2** | 0.0 | 1195.5 | 4947.0 | 5578.6 | 29.3 | 9986.4 | 177.0 | 575.1 |
| **p3** | 1992.5 | 0.0 | 3084.0 | 3352.4 | 17.6 | 6225.1 | 155.9 | 427.0 |
| **p4** | 8245.0 | 3078.0 | 0.0 | 13400.0 | 8.8 | 25707.8 | 624.1 | 1236.6 |
| **consult** | 6974.7 | 2514.3 | 10050.0 | 0.0 | 3.5 | 20289.7 | 217.9 | 13.2 |
| **diem** | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 |
| **r1** | 2.9 | 1.1 | 4.5 | 4.7 | 0.0 | 0.0 | 0.2 | 0.4 |
| **r2** | 589.9 | 313.0 | 1245.7 | 580.0 | 7.5 | 7785.0 | 0.0 | 3.2 |

*Method 2: Standard Deviation-Based Approach by Applying Probabilistic Sensitivity Analysis*

The use of a probabilistic sensitivity analysis allows the incorporation of the real distributions of all input variables. The assessment of the sensitivity of the model for $v_1$ is based on a probabilistic sensitivity analysis based on the distribution of *v1* only, while the values of the other variables are kept fixed at the input values. A measure of sensitivity of the model to $v_1$ is the standard deviation resulting from this probabilistic sensitivity analysis ($SD_{v_1}$). Subsequently the 2-order sensitivity of $v_1$ to another variable $v_2$ is assessed by another probabilistic sensitivity analysis based on the distributions of $v_1$ and $v_2$, while the values of the other variables are kept fixed at the input values. This analysis will yield another standard deviation ($SD_{v_{12}}$), which is different from the initial standard deviation ($SD_{v_1}$) because of the incorporation of the extra uncertainty associated with variable $v_2$. Hence a measure of the absolute second-order sensitivity of $v_1$ to $v_2$ ($ASS_{v_{12}}$) is the increase in the standard deviation due to $v_2$:

$$ASS_{v12}=SD_{v12}- SD_{v1} \quad \text{(Eq. 13)}$$

Subsequently the relative second-order sensitivity can be defined:

$$RSS_{v12}= ASS_{v12}/SD_{v1} \quad \text{(Eq. 14)}$$

Where $RSSv_{12}$ is the relative 2-order sensitivity of $v_1$ to $v_2$, which is the proportion of change $SS_{v_{12}}$ compared to $SD_{v_1}$.

Figure 5 shows the absolute second-order sensitivity of *p1* to the other input variables. This figure shows that the sensitivity of *p1* is most sensitive to the *per diem*, which corresponds with the findings of the 2-order sensitivity analysis based on the previous methods. A main difference is that the other variables only have a minor impact on the sensitivity of *p1* compared with the previous methods. Tables IV and V contain respectively the absolute and relative 2-order sensitivity for all input variables in the model. Those tables show that the *per diem* has the most impact on the sensitivity of most variables. Those tables also that *p1* is the only variable, which somehow influences the 2-order sensitivity of the *per diem*.

**Figure 5. The 2-order sensitivity of p1 to the other variables in the model using probabilistic sensitivity analysis.**

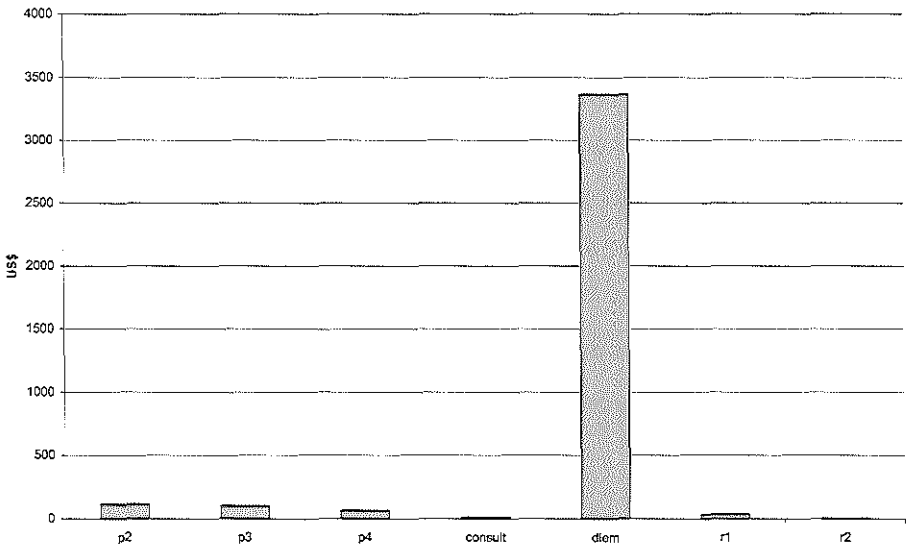



**Table IV. The absolute 2-order sensitivity for all variables in the model for depression using probabilistic sensitivity analyses.**

| | All* | One | Two | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | p1 | p2 | p3 | p4 | consult | diem | r1 | r2 |
| **p1** | 4831 | 1405 | -1405 | 117 | 101 | 65 | 11 | 3366 | 36 | 2 |
| **p2** | 4831 | 533 | 989 | -533 | 225 | 194 | 25 | 3991 | 101 | 8 |
| **p3** | 4831 | 555 | 951 | 203 | -555 | 181 | 18 | 3975 | 92 | 6 |
| **p4** | 4831 | 488 | 982 | 239 | 248 | -488 | 18 | 4024 | 98 | 18 |
| **consult** | 4831 | 152 | 1264 | 406 | 421 | 354 | -152 | 4381 | 201 | 11 |
| **diem** | 4831 | 4490 | 281 | 34 | 40 | 22 | 43 | -4490 | 20 | 15 |
| **r1** | 4831 | 318 | 1123 | 316 | 329 | 268 | 35 | 4192 | -318 | 7 |
| **r2** | 4831 | 62 | 1345 | 479 | 499 | 444 | 101 | 4443 | 263 | -62 |

**Table V. The relative 2-order sensitivity for all variables in the model for depression using probabilistic sensitivity analyses.**

| Two | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **p1** | **p2** | **p3** | **p4** | **consult** | **diem** | **r1** | **r2** |
| **p1** | -1.00 | 0.08 | 0.07 | 0.05 | 0.01 | 2.40 | 0.03 | 0.00 |
| **p2** | 1.86 | | 0.42 | 0.36 | 0.05 | 7.49 | 0.19 | 0.02 |
| **p3** | 1.71 | 0.37 | | 0.33 | 0.03 | 7.16 | 0.17 | 0.01 |
| **p4** | 2.01 | 0.49 | 0.51 | -1.00 | 0.04 | 8.25 | 0.20 | 0.04 |
| **consult** | 8.32 | 2.67 | 2.77 | 2.33 | -1.00 | 28.82 | 1.32 | 0.07 |
| **diem** | 0.06 | 0.01 | 0.01 | 0.00 | 0.01 | -1.00 | 0.00 | 0.00 |
| **r1** | 3.53 | 0.99 | 1.03 | 0.84 | 0.11 | 13.18 | -1.00 | 0.02 |
| **r2** | 21.69 | 7.73 | 8.05 | 7.16 | 1.63 | 71.66 | 4.24 | -1.00 |

Finally a total measure of sensitivity of all other input variables on $v_1$ is determined by the following equation:

$$RSS_{v1234\ldots n} = S_{v1}/ASS_{v1234\ldots n} \qquad \text{(Eq. 15)}$$

Where $RSS_{v1234\ldots n}$ is the relative second-order sensitivity of $v_1$ to all other input variables; $S_{v1}$ is standard deviation for probabilistic sensitivity analysis for $v1$ only and $ASS_{v1234\ldots n}$ is standard deviation for probabilistic sensitivity analysis incorporating distributions of all input variables.

## Point-Sensitivity

The above-mentioned methods for the determination of second-order sensitivity are based on the distribution of variables, which assess the simultaneously the impact of a fixed input value (mean) and its associated range on the sensitivity of another input variable.

$$S_m = \Delta Y_m = a_m \cdot \Delta V_m \qquad \text{(Eq. 9)}$$

Eq. 9 states that sensitivity of variable $V_m$ depends on the range of $V_m$ and the mean fixed input value of $a_m$ only. Consequently it is reasonable to perform a sensitivity analysis, which shows only the impact of the fixed input value of $a_m$ without including the impact of its range. In a previous article, we introduced the concept of point-sensitivity analysis, which measures the responsiveness of the outcome to the input value at a fixed point and does not include any associated range of the input.[3] The definition of point-sensitivity was based on the concept of point-elasticity (Eq. 16) where $X$ and $Y$ represent the fixed input value and the outcome respectively and $\eta_s$ refers to point-elasticity, which describes the % change in the outcome, when the input variable changes 1%.

$\eta_s = dY/dX \cdot X/Y$ (Eq. 16)

In the case of a linear sensitivity relationship (Eq. 17) a more simple formula (Eq. 18) can be used:

$Y = a \cdot X + b$ (Eq. 17)

$\eta_s = a \cdot X/(a \cdot X + b)$ (Eq. 18)

The concept of point sensitivity was in the previous studies[3] [4] applied to the first-order sensitivity of an input variable, but it can also be applied to determine the impact of the mean value of $a_m$ on the second-order sensitivity of another input variable by adjusting the equations:

$\eta_{m,n} = da_m/dv_n \cdot v_n/a_m$ (Eq. 19)

Where $\eta_{m,n}$ is the second-order point sensitivity of $v_m$ to $v_n$. Because of the linear sensitivity relationship (Eq. 20) a more simple formula (Eq. 21) can be used:

$a_m = k \cdot v_n + l$ (Eq. 20)

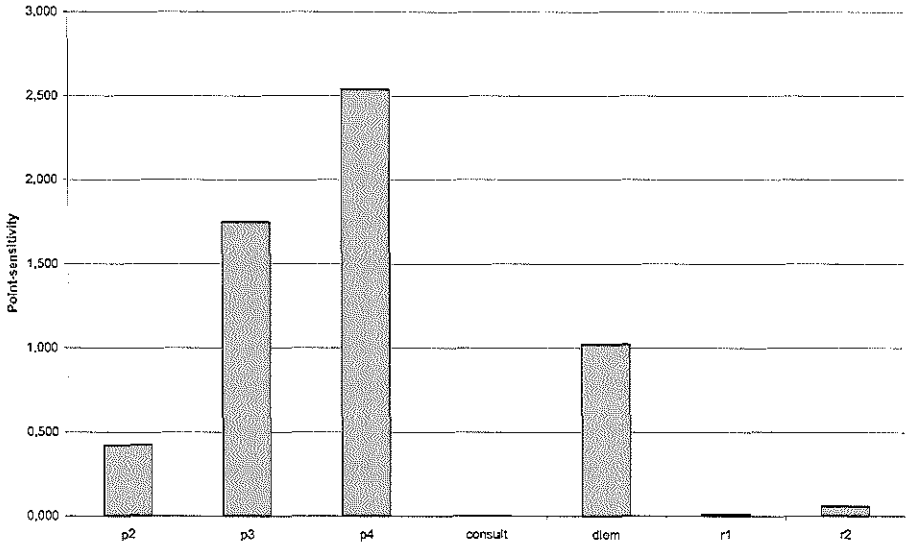
$\eta_{m,n} = k \cdot v_n/(k \cdot v_n + l)$ (Eq. 21)

In this example when $v_m$ is *p1* and $v_n$ is *pdiem*, $k$ is –57,7 and l is –607.7, the second-order point sensitivity is 1,021. Hence increasing the *per diem* by 1 % results only in a 1.201% increase of $a_{p10}$. Figure 6 show the point-sensitivity of *p1* to the other input variables, which shows that *p3* and *p4* now have become more sensitive variables than the per diem. Table VI shows the results of the point-sensitivity analyses for all variables in the model for depression, which show that *p3* and *p4* have become most sensitive variables for most of the other variables, while the impact of the per diem has become lower.

**Table VI. The point-sensitivity outcomes.**

| Two | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **p1** | **p2** | **p3** | **p4** | **Consult** | **diem** | **r1** | **r2** |
| **p1** | na | 0.423 | 1.751 | 2.543 | 0.002 | 1.021 | 0.015 | 0.060 |
| **p2** | 1.325 | na | 1.710 | 2.389 | 0.002 | 0.998 | 0.026 | 0.074 |
| **p3** | 1.331 | 0.414 | na | 2.315 | 0.000 | 1.000 | 0.025 | 0.050 |
| **p4** | 1.427 | 0.429 | 1.714 | na | 0.000 | 1.000 | 0.010 | 0.001 |
| **consult** | 0.109 | 0.034 | 0.017 | 0.011 | Na | 0.000 | 0.003 | 0.074 |
| **diem** | 1.330 | 0.415 | 1.714 | 2.315 | 0.000 | na | 0.025 | 0.051 |
| **r1** | 0.999 | 0.442 | 1.759 | 1.036 | 0.002 | 1.410 | Na | 0.001 |
| **r2** | 1.501 | 0.566 | 1.610 | 0.025 | 0.035 | 0.961 | 0.001 | na |

**Figure 6. The 2-order sensitivity of p1 to the other variables in the model using point-sensitivity analysis.**




## DISCUSSION

The constraint of the standard methods for sensitivity analysis is that those methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taken into account the existing relationships between those variables. The present study describes to present various experimental methodologies, which considers this type of inter-variable uncertainty for economic evaluations based on decision analytic techniques. We presented initially a method assuming only uniform distributions, and subsequently various methods incorporating the real distributions. The results show that this second-order sensitivity of a dependent input variable depends very much on the distribution of the independent input variables. Overall our analysis showed that the most sensitive variables for the outcome of the model (1-order sensitivity), were also the most sensitive for the 2-order sensitivity of the other input variables. Finally we presented a point-sensitivity analysis, which shows only the impact of the fixed input value of $a_m$ without including the impact of its range. The results from this type of 2-order sensitivity analysis resulted to a different ranking of all variables according to their impact on the sensitivity of other variables, which showed the impact of the distribution on the 2-order sensitivity. From a methodological point of view one may argue what is scientifically the most correct measure of second-order sensitivity: the 2-order sensitivity based on the distributions or on the fixed input value only.

The results from the 2-order sensitivity analyses show that the most sensitive variables in the model have also the most impact on the sensitivity of the other input variables. Consequently one can argue that the reduction of uncertainty associated with the most sensitive variables in a model, will lead to an overall reduction of uncertainty in the outcomes in the model because of a reduction in sensitivity of most variables. Hence it may be more important to investigate the uncertainty of the most sensitive variables of the model in more depth, e.g. by means of a meta-analysis, than to investigate the uncertainty associated with all variables. On the other hand, it may be worthwhile to reduce the sensitivity of the most sensitive variable by identifying the variables, which mostly influence its sensitivity by means of 2-order sensitivity analyses. Although a more general mathematical prove is required to support those conclusions, which are based on a depression model example, we may end with the following recommendations: 1) identify the most sensitive variables in the model by means of traditional sensitivity analysis (1-order sensitivity), 2) identify the variables, which have the most impact on the sensitivity of those most sensitive variables by means of 2-order sensitivity analysis, 3) reduce the uncertainty associated the variables, which were identified in the previous steps.

Another approach to judge the benefit of further research to reduce uncertainty in a modeling study is based on the concept of Value of Information. Recently a Bayesian decision theoretic framework has been suggested for the evaluation of health technologies.[5] [6] This approach distinguishes the conceptually separate decision concerning efficient service provision given the level of information available, from the decision concerning to fund further information collection. The basic concept for a modeling study is to measure the uncertainty in cost-effectiveness outcomes and determine the worth of further data collection.[5] Here the decision of further data collection is based upon the expected cost of uncertainty, which is determined by both the extent of the uncertainty surrounding the efficient service provision and the consequences of this uncertainty. The extent of the uncertainty is measured by the error probability associated with the decision and the consequences of uncertainty are measured in terms of health benefits foregone when this uncertainty causes the incorrect decision to be made concerning service provision. Where these health benefits are valued according to society's willingness to pay for certain health outcomes, the approach gives a monetary value for the amount that society is willing to pay to reduce the uncertainty surrounding the service provision. The primary objective of both the concept of 2-order sensitivity and the concept of Value of Information is to identify the impact of input variables in a model on the level of uncertainty within a model and to explore the worth of further research to reduce uncertainty of an input variable. The Value of Information concept quantifies the costs of further research and relates it to reduction of uncertainty in the decision-making process, while the 2-order approach only helps the researcher in prioritizing further research by identifying the key input variables in the mode without incorporating the cost of extra research. The advantage of the 2-order approach is that it takes into account the inter-variable uncertainty based on the existing relationships between the input variables. Although the concept of Value of Information initially was applied to empirical data, it may also be applied to a health

economic model. When an input variable appears to be very sensitive because of its large range, the concept of Value of Information may be applied by assessing the cost of extra research for reducing the confidence interval of that input variable by means of using more data. For example a clinical response rate was initially based on one few studies leading to a large variance, which may be reduced by including the results of more published studies or by performing a formal meta-analysis. The concept of Value of Information in a health economic model may be extended by incorporating inter-variable uncertainty, which may further reduce the extra costs for the reduction of the uncertainty associated with the healthcare decision. For example the 2-order sensitivity analysis may show that the uncertainty associated with the clinical response rate depends mainly on the hospitalization rate. Consequently the extra cost of exploring this rate may be less than the extra costs for reducing the range of the response rate.

## References

[1] Drummond et al. Current Trends in the Use of Pharmacoeconomics and Outcomes Research in Europe. Value in Health 2(5): 323-332

2 Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health care Technology: A Report on Principles. Ann Intern Med. 1995; 122: 61-70

3 Nuijten MJC, Hardens M. Measuring sensitivity in pharmacoeconomic studies: an integration of point-sensitivity and range-sensitivity. Pharmacoeconomics. 1997 Nov; 12(5): 555-564

4 Nuijten, MJC. Measuring sensitivity in pharmacoeconomic studies: an integration of point-sensitivity and range-sensitivity: a refinement by incorporation of probability distributions. PharmacoEconomics, 1999; 16 (1): 33-41

5 Claxton K. The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. Journal of Health Economics. 1999; 18: 341-64

6 Claxton K, Posnett J. An economic approach to clinical trial design and research priority setting. Health Economics 1996; 5 (6): 513-24

Chapter 11

# Combining a Budgetary Impact Analysis and a Cost-Effectiveness Analysis Using Decision-Analytic Modelling Techniques


## SUMMARY

BACKGROUND: Reimbursement of new drugs is usually based on the budgetary impact of a new drug but there is also increasing demand for cost-effectiveness data on new drugs. OBJECTIVE: To present a modelling technique (methodology) for an appropriate assessment of the budgetary impact of a new drug, which can simultaneously be used for a traditional cost-effectiveness analysis. DESIGN AND PERSPECTIVE: To illustrate the methodology, a model was constructed for a new hypothetical drug in Parkinson's disease, which allowed us to determine the budgetary impact and the cost-effectiveness of this new antiparkinsonian drug from a societal perspective. The methodology consisted of two steps: (i) a simple population model (Markov model) was constructed to validate the epidemiological data by proving the consistency between the prevalence and incidence of Parkinson's disease for the Dutch population; (ii) this model was extended to a more complex model (semi-Markov model) by incorporation of disease progression for Parkinson's disease and all relevant economic and clinical measures. These included all drug utilisation associated with Parkinson's disease, as well as other resource utilisation patterns associated with outpatient and inpatient care for the treatment of Parkinson's disease. RESULTS: The study showed that the difference in epidemiological data between a simple model and a complex model are substantial, which justifies the development of a complex model with a higher external validity. The complex model allowed an assessment of all potential candidates for the new drug and simultaneously allowed the assessment of the cost-effectiveness of the new drug versus usual care. CONCLUSION: One model can be used for an appropriate assessment of the budgetary impact and the cost-effectiveness of a new drug.

## INTRODUCTION

In The Netherlands the main response to rising costs has been the introduction of an impressive number of cost containment policies over the last decade. This resulted in the Drug Reimbursement System, based on the classification of drugs into groups of interchangeable drugs (GVS). There is a fixed refund price being based on the average list price for "therapeutically interchangeable drugs" belonging to the same group (cluster). Any new drug, which cannot be clustered, is not reimbursed, unless there is no treatment for the relevant pathology (e.g. Alzheimer). The use of pharmacoeconomic data was considered for reimbursement of those drugs, which are cost effective. In July 1999, the Minister approved the guidelines ["Richtlijnen voor farmaco-economisch onderzoek" (Dutch guidelines for Pharmacoeconomic Research)].[1] Currently the Ministry is considering to adjust the reimbursement legislation by implementing official requirements for submission of health economic data comparable with Australia and Canada.[2]

Besides the cost-effectiveness of a new drug, reimbursement will also be based on the budgetary impact of a new drug on JOZ [Jaaroverzicht Zorg (Annual National Health Care Budget)], especially the impact of a new drug on the drug budget. Therefore, the authorities require an assessment of the impact of a new drug on the annual healthcare budget, and especially the drug budget. For the financial analysis, data on the following subjects will be required: descriptive epidemiology (data on incidence and prevalence); the patient group that is indicated for the drug and the anticipated substitution effects (i.e. the extent to which the existing treatment will be replaced); the use of the drug (posology, length of the treatment, etc.), the price of the drug; and the total treatment costs. Especially, if reimbursement of pharmaceuticals is going to be based in part on data derived from a financial analysis, it is vital to carefully scrutinise and refine this type analysis.

The primary objective of this manuscript was to demonstrate the concept of an appropriate assessment of the budgetary impact of a new second-line drug using modelling techniques, which simultaneously can be used for a traditional cost-effectiveness analysis.

## METHODS

The financial analysis was based on data from a health economic study and descriptive epidemiology (data on incidence and prevalence). To illustrate the methodology, a model was constructed for a new hypothetical drug in Parkinson's disease, which allowed us to determine the budgetary impact and the cost-effectiveness of this new antiparkinsonian drug (AP) from the perspective of the society in The Netherlands. This new drug was proposed as an add-on therapy to usual care in patients who develop fluctuations.

The methodology consisted of the following steps:

- Step 1: A simple population model (Markov model) was constructed. The objective of this model was to validate the epidemiological data by proving the consistency between prevalence and incidence of Parkinson's disease for the Dutch population.
- Step 2: The simple population model was extended to a more complex model (semi-Markov model) by incorporation of disease progression for Parkinson's disease and all relevant economic and clinical measures. These included all drug utilisation associated with Parkinson's disease as well as other resource utilisation patterns associated with outpatient and inpatient care for the treatment of Parkinson's disease, direct nonmedical costs and indirect costs (working days lost). The primary objective of this model was to estimate the potential number of candidates and the annual drug costs for a new drug in Parkinson's disease. Secondary objectives were (i) to compare the impact of this new drug versus usual care (standard therapy) on the total annual healthcare budget (health insurance perspective); (ii) to compare the impact on total costs from the perspective of society and the patient by including direct non-medical costs and indirect costs; and (iii) to perform a traditional cost-effectiveness analyses comparing the new drug versus usual care.

## Simple Population Model

### *Literature Review*

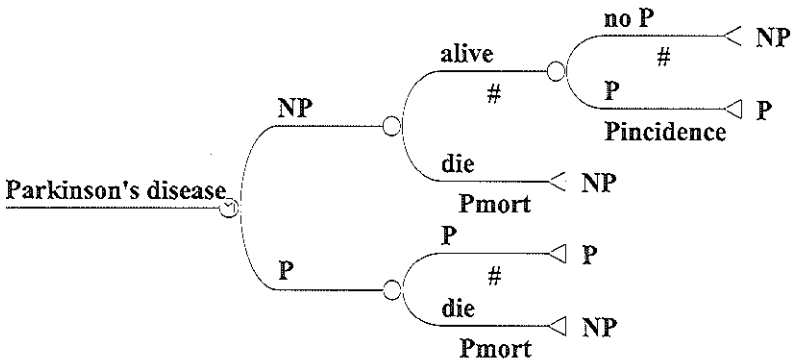
A literature search on incidence, prevalence data and disease-specific mortality for Parkinson's disease was performed on international, and especially Dutch, literature. Search parameters identified as keywords included incidence, prevalence, Parkinson's disease, mortality, life expectancy. The prevalence estimates of idiopathic Parkinson's disease were in the range of 66 to 258 per 100,000, while the annual incidence varies from 4 to 24 per 100,000. Data on national Dutch population statistics were supplied by the Dutch national office of statistics (CBS: Central Bureau Statistics) [3] and included: annual mortality, annual birth rate and life expectancy. The DATATOP study showed that after an average of 8.2 years of observation, the overall death rate was unaffected by type of treatment and was about that expected for an age- and gender matched US population without Parkinson's disease.[4] This study was initially designed to test outcomes in 800 previously untreated patients with Parkinson's disease, treated with deprenyl or tocopherol (vitamin E) over a period of 18 months; however, follow-up was extended to an average of 8.2 years yielding long-term death rates for Parkinson's disease. The results of the literature search are shown in Table I.

### *Model Structure*

A simple population model (Markov model) was constructed. The objective of this model was to validate literature data by proving the consistency between prevalence and incidence of Parkinson's disease. A Markov process is a particular type of decision tree models and represents a convenient way of modelling the long-term evolution of health states over successive time periods: A Markov process

model describes several discrete states of health in which a person can be at time t=n as well as the states of health into which the person may move at time t=n+1.

**Figure 1. Simple Markov population model for Parkinson's disease.**




The progression from t=n to t=n+1 is called a cycle. Probabilities are associated with each change from one health state to another; these are termed transition probabilities (P). A Markov model is deemed appropriate for use in situations where events reoccur (e.g. incidence) over time, and when individual persons of a population move among a finite number of health states (Parkinson's disease or no Parkinson's disease) over the time period.[5,6] Figure 1 shows the structure of the population Markov model for Parkinson's disease. The cycle time chosen for the model was 1 year, which corresponds with epidemiological data on annual incidence, prevalence and mortality and the 1-year horizon for the financial analysis.

For this analysis we defined mutually exclusive Markov states. Initially we defined Markov health states corresponding with the existence of Parkinson's disease as follows: "P": a person with Parkinson's disease; "NP": a person without Parkinson's disease.

In the model all persons start in the health state "NP". After one cycle of 1 year there are three possibilities:

- A person will not develop Parkinson's disease and will stay in health state "NP".
- A person will develop Parkinson's disease and will move to health state "P".
- A person will die and will move to health state "NP", which is based on assumption that the annual mortality equals the annual birth rate.

During subsequent cycles, a persons with Parkinson's disease will have two possibilities:

- A person will not die and will stay in health state "P".
- A person will die and will move to health state "NP".

Hence persons once entered in health state "P" will not return to the health state "NP", because Parkinson's disease is a chronic disease without recovery.

### *Key Model Assumptions*

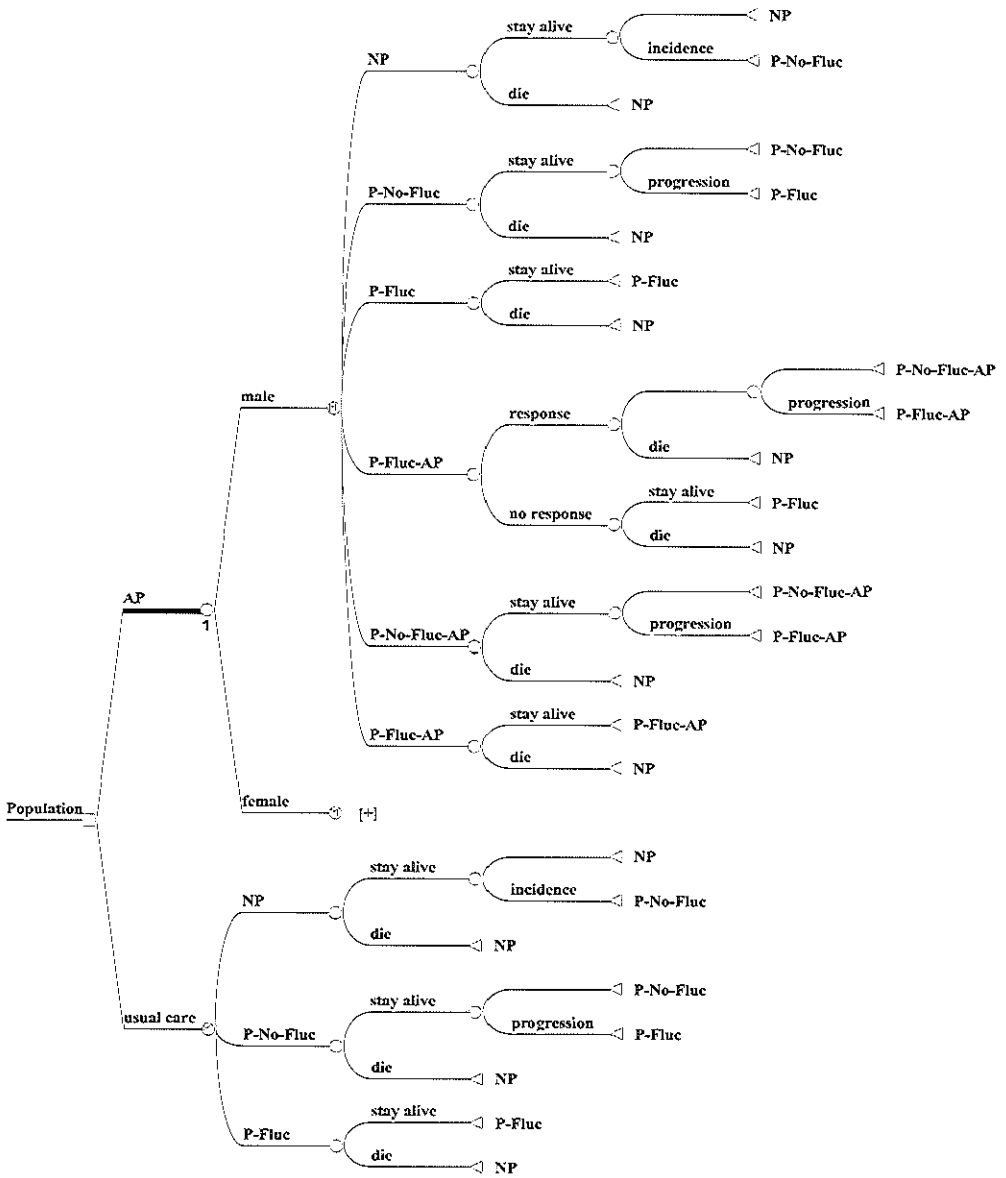
There are a number of key assumptions upon which the model is based. These are:

- The follow-up period was the minimal time period in which the proportion of patients with Parkinson's disease stabilises.
- The population size is constant: the annual mortality equals the annual birth rate.
- The population distribution with regards to age is constant.
- Gender differences are not taken into account and consequently we assumed no differences in mortality rate, prevalence and incidence for Parkinson's disease between males and females, which is contrary to reality: males have a lower life expectancy.
- Age is not incorporated in the model. Consequently, we did not consider that annual mortality will increase by age and that occurrence of Parkinson's disease usually starts at an average age of 62. This assumption is only allowed because of the previous assumptions with regards to constant population size (1) and constant population distribution with regards to age (2).
- The model only incorporated the incidence and prevalence of Parkinson's disease, while other morbidities were only indirectly reflected in annual mortality rate.
- Parkinson's disease is not associated with a lower life expectancy, which was based on The DATATOP study.[4]

**Table 1. Prevalence data.**

| **Source** | **Prevalence per 100,000** | **Population** |
|---|---|---|
| Surcliffe[7] | 66-258 | General |
| Dodel[8] | 100-180 | |
| De Pedro Cuesta[9] | 100 | |
| Range | 66-258 | General |
| | **Incidence (/100,000)** | **Population** |
| Rajput[10] | 5-21 | General |
| Dodel (Ref18-21) | 4-20 | General |
| Koller[11] | 5-24 | General |
| Range | 4-24 | General |
| **Source** | **Population Statistics** | **Population** |
| Annual mortality rate | 0.012834 | General |

**Figure 2. Complex Markov population model for Parkinson's disease.**




### *Analysis*

The analysis determined at which prevalence and incidence data point the proportion of persons in the health state "P", representing the modelled prevalence of Parkinson's disease, reaches an equilibrium after an "infinite" number of cycles (n=1000). Hence the population model allows an assessment and

a validation of the incidence and prevalence data derived from international literature for the Dutch population by comparing the prevalence and incidence data, once the model has reached equilibrium, with literature data (Table I).

## Complex Population Model

While in a standard Markov process model transitions only depend on health state and time, the transitions in a semi-Markov model also depend on co-variates like clinical past history, the current treatment, age and gender, which allow a more realistic simulation of the disease. Figure 2 shows the structure of the semi-Markov population model for Parkinson's disease, which consists of two arms: an arm based on usual care only and an arm which includes the introduction of a new innovative hypothetical drug for Parkinson's disease. The model structure for the usual care treatment arm is identical. In our model a hypothetical cohort of persons with fluctuations will receive AP or usual care. After the first cycle patients using AP may have improved by control of fluctuations: we assumed a response rate of 80%. Responders will continue the treatment with AP for the remainder of the study period even if they progress to experiencing fluctuations, while non-responders will terminate treatment with AP.

In addition, the semi-Markov model was based on the following adjustments to the initial Markov model:

- The cycle time chosen for this model was 6 months, which closely approximates the follow-up period of the clinical trials in Parkinson's' disease, which when doubled corresponds with a time horizon of annual financial budgets.
- We stratified the existing "P" health state in the following states: "P-no fluctuations": a person with no fluctuations; "P-fluctuations": a person with fluctuations.
- We used the health state of the original Markov model: "NP": a person without Parkinson's disease.
- Age-specific mortality was incorporated in the model in order to reflect increasing mortality risk when a patient gets older: The population was stratified in following subpopulations: 0,1,2,3,4,..... 100. Because of difference in life expectancy between males and females, gender was also incorporated in the model by stratifying patients in to males and females.

Our model was based on the following assumptions:

- The Markov health states in this model correspond with severity levels according to the existence of motor fluctuations: "no fluctuations", "fluctuations".
- The follow-up period was the minimal time period which was necessary to stabilise (i) the proportion of patients with Parkinson's disease and (ii) the distribution between Parkinson's health states.

- A 6-month cycle was used because this interval closely approximated the follow-up period of the clinical trials for AP.
- We assumed that AP will not reduce disease progression.
- We assumed that the complete target population for AP will receive AP. Hence the analysis will show the maximum budgetary impact of AP.
- We assumed that that the occurrence of Parkinson's disease only starts in the subpopulation of 62, which is based on the average age of onset (62 years) in the literature.[4 12 13]
- The population size is constant: the annual mortality equals the annual birth rate.
- The model only incorporated the incidence and prevalence of Parkinson's disease, while other morbidities are only indirectly reflected in annual mortality rate.
- Parkinson's disease is not associated with a lower life expectancy.[4]
- A life-time model was constructed; patients were followed until death.
- We assumed no differences in prevalence and incidence for Parkinson's disease between males and females.

## Data Sources

The data sources and collection were based on the following:

- The incidence and prevalence were derived from the simple Markov model.
- The probabilities of disease progression to fluctuations and mortality were based on published literature and population statistics data (CBS).[3]
- Utilities and units of resource utilisation were obtained from a cross-sectional study.[14]
- Direct medical costs, involving estimates of the units of healthcare utilisation and their prices/tariffs (product of unit and price), were derived from official lists.[15]

### *Probabilities*

The initial prevalence and incidence values were based on the simple Markov model and initially all patients with Parkinson's disease started in the health state "no fluctuations". The incidence was based on the general population and was adjusted to the assumption that the occurrence of Parkinson's disease only starts in the subpopulation of 62. The incidence of Parkinson's disease in the general population (Ipop) was calculated as (Eq. 1)

$$I_{pop}=I_{age\text{-}male} \times F_{age\text{-}male}+I_{age\text{-}female} \times F_{age\text{-}female} \qquad \text{(Eq. 1)}$$

Where

$I_{pop}$= incidence of PD in the general population

$I_{age\text{-}male}$ = incidence of PD in males at 62.

$I_{age\text{-}female}$ = incidence of PD in females at 62.

$F_{age\text{-}male}$ = Proportion of males at 62 in total population.

$F_{age\text{-}female}$= Proportion of females at 62 in total population.

We assumed that $I_{age\text{-}male} = I_{age\text{-}female}$ and

Ipop = 4 per 100,000

$F_{age\text{-}male} = 0.04495$[3]

$F_{age\text{-}female} = 0.04508$[3]

Consequently:

$I_{age\text{-}male} = I_{age\text{-}female} = I_{pop}/(F_{age\text{-}male}+F_{age\text{-}female}) = 444.26$ per 100,000. (Eq. 2)

The transition probability between the different health states in our example was derived from a study by Nuijten.[14] The annual probability of 0.147 was transformed to a 6-months probability of 0.075. The probabilities of mortality were based on the annual age-specific mortality quotient for men and women derived from 1998 population data, which were transformed to 6-month probabilities.[3]

### *Utilities*

The assessment of the utility was based on the assignment of fixed value to the health states associated with a 6-month cycle. The utilities for the different health states were derived from a cross-sectional study by Dodel.[16] The health states "P-no fluctuations" and "P-fluctuations" had utilities of respectively 0.73 and 0.49. The classification of disease severity in this study corresponded with the health states in our model. The underlying assumption was that the function between severity and utility may be extrapolated from Germany to The Netherlands, when we assume that clinical outcomes are not country specific.[17]

### *Resource Utilisation*

The resource utilisation for the different health states was based on a German cross-sectional study by Dodel, which was performed over a period of 3 months.[16] We used resource utilisation from the cross-sectional study, to which we applied official Dutch tariffs and prices (Table II). We validated the assumption that the German and Dutch treatment patterns for Parkinson's disease are similar by comparing the distribution of antiparkinsonian drugs and referral patterns.

### *Costing*

Costing was performed from the societal perspective according to the Dutch pharmacoeconomic research guidelines. The cost assessment was based on the assignment of fixed costs to the health states associated with a 6-month cycle. The costs of each health state were determined by the resource utilisation associated with a health state: medical resource utilisation (e.g. medication, consultations, laboratory and diagnostic tests and procedures, hospitalisations, nursing home). In addition, direct

non-medical costs were included. Examples are: equipment, transportation and home modifications. Indirect costs due to working days lost were not included, because the average age of patients with Parkinson's disease is 60-64, while most people in The Netherlands are no longer actively working after 60 years.[3]

Market prices, where available, were used to evaluate opportunity costs (*i.e.* costs of foregone opportunities as a result of investment in a specific healthcare strategy). In the absence of available costs, tariffs were used as an approximation.

- Drug prices were derived from the Taxe (Dutch list of drug prices).[18]
- Consultations, procedures and hospitalisation costs were derived from the College Tarieven Gezondheidszorg (COTG 1999).[15]
- Nursing costs were derived from the COTG 1999.[15]
- Transportation costs were valued according to the national allowance per km (CBS).[3]
- Costs for home medications and other non-medical costs were based on market prices.

**Table II. Costs for Each Health State (in € per 3 months)**

| Stage | Direct medical costs | | | | | |
|---|---|---|---|---|---|---|
| | Drugs | Consul-tations | Physio-therapy | Diagnostic procedures | Hospita-lisation | Total |
| No-fluctuations | 340 | 169 | 463 | 0 | 0 | 972 |
| Fluctuations | 918 | 156 | 839 | 15 | 210 | 2138 |
| **Stage** | **Non-direct medical costs** | | | | | |
| | Social services | Trans-port | Home modif. | Other | Total | |
| No-fluctuations | 0 | 116 | 1 | 0 | 117 | |
| Fluctuations | 405 | 115 | 120 | 10 | 650 | |

## Analysis

Financial and health economic analyses were executed, which were respectively performed at population and patient level. The analyses were performed after stabilisation of the model for usual care resulting in a static distribution between different health states and a subsequent switch to AP for fluctuators.

We forecasted the impact of an add-on therapy with AP on the annual drug budget for each subsequent year after launch of AP until stabilisation of the annual costs of AP. This analysis shows for each year: 1) the maximum number of patients using AP and 2) the annual costs of AP. The base case analysis was based on assumption that all potential candidates for AP will use the drug. A scenario analysis was performed assuming that only patients, who start developing fluctuations, are going to use AP.

We also forecasted the impact of AP on the healthcare budget. This analysis was an extension of the previous one by including other direct medical costs. Finally we also included nonmedical resource utilisation in order to assess the impact on the national budget.

Besides the financial analysis, a cost-effectiveness analysis was performed from the society perspective using the following outcomes:

- The expected average effectiveness: time without severe fluctuations, QALYs.
- The total costs per patient, including cost distribution: direct medical and non-medical costs.

Clinical and economic outcomes were discounted because the time horizon of the model extended beyond a period of 1 year. The present values were calculated for the year 2000 and the discount rate was based on the recommended 4% according to the Dutch pharmacoeconomic guidelines.[1]

# RESULTS

## Simple Population Model

The initial prevalence and incidence values were the minimum values of the ranges 66 and 4 per 100,000. Figure 3 shows the change in prevalence over an infinite period (1000 years). The model takes nearly 300 years to reach the equilibrium, because the initial population at t=0 does not contain any patients with Parkinson's disease: these patients subsequently enter the model based on the incidence. The prevalence increases and stabilises when at 295 per 100,000, which slightly exceeds the upper range derived from literature. Thus we determined the appropriate prevalence and incidence data for the population model in The Netherlands, which was based on the listed key assumptions, including the assumption of constant population size. The results show that only at the minimum value for incidence the prevalence remains near the maximum value taking into account the ranges found in the literature. Hence, we validated the key input data of the population model: the incidence and prevalence of Parkinson's disease. This simple model confirms the relationship between prevalence and incidence data, but the representativeness of the model may be questioned, taking into account the low incidence and high prevalence at which the model stabilises.

**Figure 3. Results of population model for Parkinson's disease.**

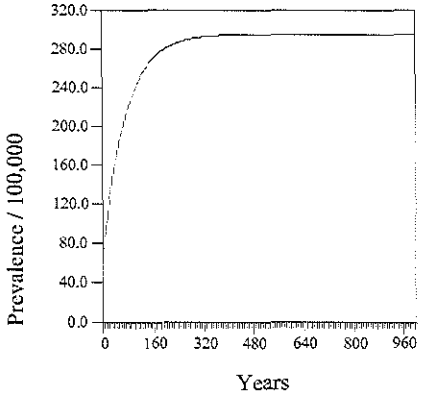


**Figure 4. Results of population model for Parkinson's disease.**

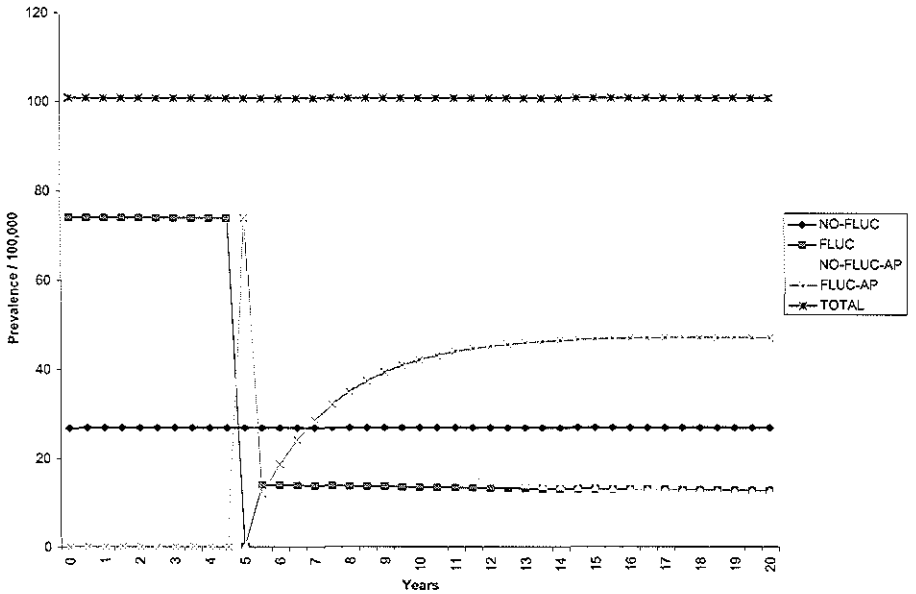
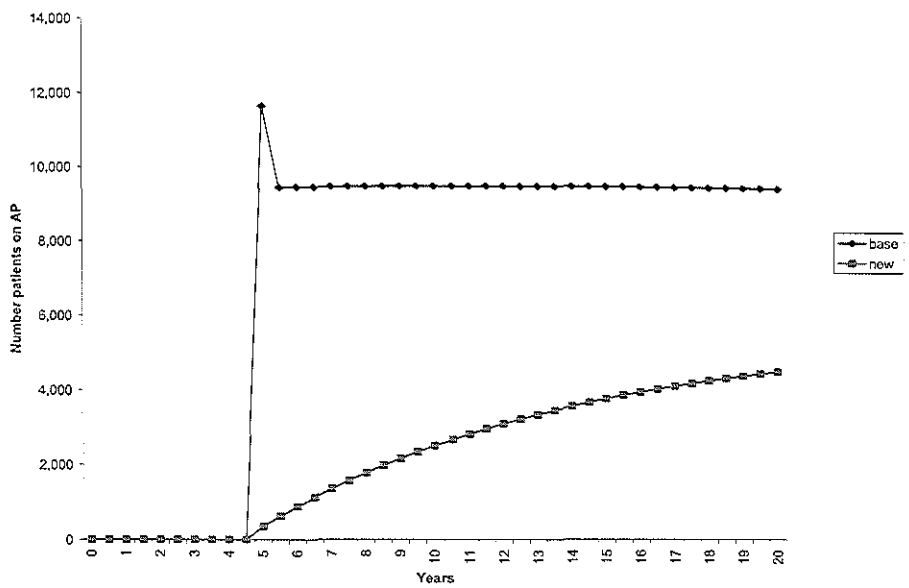



**Figure 5. Results of population model for Parkinson's disease: potential number of patients.**



## Complex Population Model

Figure 4 shows the results of an analysis after stabilisation of the model, which was achieved after an "infinite" number of cycles (n=1000): 5 years before the introduction of AP and the following 15 years. The patient population before launch consisted of patients without fluctuations (NO-FLUC) and with fluctuations (FLUC) with a prevalence of 27 per 100,000 and 74 per 100,000, respectively, yielding a total prevalence of 100 per 100,000. Hence, the total number of patients with Parkinson's disease in The Netherlands is 15,865 based on a total population of 15.7 million, which consists of 4,241 patients without fluctuations and 11,624 patients with fluctuations. Patients with fluctuations switch to AP at t=5 which corresponds with the time of launch. Subsequently, the patient population consisted of two additional patient groups: patients without fluctuations using AP (NO-FLUC-AP) and patients with fluctuations using AP (FLUC-AP). The model stabilises over a period of 20 years, leading to a new distribution of the prevalence.

- NO-FLUC: 27 per 100,000 (unchanged),
- FLUC: 14 per 100,000 (N=2,199),
- NO-FLUC-AP: 14 per 100,000 (N=2,199),
- FLUC-AP: 46 per 100,000 (N=7,223).

Figure 5 shows the number of patients using AP after launch at t=5, which rapidly stabilises at 10,493. The initial drop was due to the 20% non-responders.

Table III shows the development of the cumulative discounted costs after launch of AP compared with the total costs associated with usual care after respectively 1,5 and 10 years of introduction. AP reduces the total costs with €521,758, €2,221,224 and €3,609,027 after respectively 1, 5 and 10 years. AP substantially reduces medical costs, especially due hospitalisation and consultations, and also the non-medical costs, resulting in overall cost savings. In addition, AP also reduces costs within the drug budget. The cost saving of AP decreases the total drug costs from the first year by €115,072, which increased to €507,321 and €786,154 after respectively year 5 and 10.

**Table III. Results of population model for Parkinson's disease for cumulative medical and other costs (discounting).**

| Year | | Costs (€) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Drugs | Consul-tations | Physio-therapy | Diagnostic procedures | Hospita-lisation | Non-medical | Total |
| 1 | AP | 455870 | 102207 | 393670 | 4262 | 35828 | 220141 | 1211977 |
| | Usual care | 570942 | 96851 | 521684 | 9613 | 130490 | 404154 | 1733735 |
| 5 | AP | 2135233 | 471298 | 1850075 | 20958 | 264776 | 1060873 | 5803213 |
| | Usual care | 2642554 | 448266 | 2414569 | 44492 | 603963 | 1870593 | 8024437 |
| 10 | AP | 4014829 | 850949 | 3450614 | 41973 | 550916 | 2060459 | 10969741 |
| | Usual care | 4800983 | 814408 | 4386781 | 80833 | 1097277 | 3398486 | 14578768 |

### Cost-Effectiveness analysis

Table IV shows the outcomes of the cost-effectiveness study. The results of the baseline analysis show that the use of AP with a daily treatment cost (DTC) of €2.7 (NLG 5) is less costly than usual care only with direct medical costs of €53329 versus €54654, respectively, and total costs of €69,917 versus €71,268. AP increased the Time without fluctuations by 3.41 years and was associated with a 0.91 increase (11.77%) in QALYs (8.74 QALYs versus 7.82 QALYs). The use of AP results in an incremental cost-effectiveness ratio of 1,380 € per QALY, when only medical costs are included. The cost per year without fluctuations is €405, when only medical costs are included. The use of AP was dominant for both Time without fluctuations and QALYs, when all costs are included.

**Table IV. Results of the Base Case Analysis: Costs* and Effectiveness.**

| Outcome | Time without fluctuations | QALYs | Costs (€) | | |
|---|---|---|---|---|---|
| | | | Medical | Non-medical | Total |
| AP | 3.41 | 8.74 | 53329 | 16589 | 69917 |
| Usual care | 0.00 | 7.82 | 54654 | 16613 | 71268 |
| Difference | 3.41 | 0.91 | -1325 | -25 | -1350 |
| Cost per year without fluctuations | | | 405 | N.A. | AP:dominant |
| Cost per QALY | | | 1380 | N.A. | AP:dominant |

* discounted at 4%

## DISCUSSION

This study presented a methodology for an appropriate assessment of the budgetary impact of a new drug, which simultaneously can be used for traditional cost-effectiveness analysis. A model was constructed for a new hypothetical drug in Parkinson's disease, which allowed us to determine the budgetary impact and the cost-effectiveness of this new drug (AP) from the perspective of the society in The Netherlands.

The use of a Markov model allowed us to simulate a financial analysis and a cost-effectiveness analysis by extrapolating the clinical outcomes of AP beyond the duration of the AP clinical trial. There are also advantages to combining a budgetary impact analysis and a cost-effectiveness analysis. The analyses are based on the same underlying assumptions and data, which will increase the consistency between the outcomes of both analyses.

Key data for the financial analysis are prevalence and incidence. A simple Markov model was constructed to validate the epidemiological data by testing the consistency between prevalence and incidence of Parkinson's disease for the Dutch population. The simple model can only be used to predict the potential number of patients who are candidates for a first-line treatment, at time of launch. However, the duration of treatment cannot be derived from this model: patients starting on a first line treatment will switch over time to another treatment due to disease progression and the initial

treatment of patients with Parkinson's disease will be adjusted after development of fluctuations. Therefore, the simple model cannot be used to determine the annual budgetary impact of a new first-line drug. In addition, the model cannot be used to determine the annual budgetary impact for a new drug targeting subpopulations (e.g. patients with fluctuations), because the health state "P" only considers patients with Parkinson's disease as a homogenous population without distinguishing between severity of the disease or other stratifications.

Given the assumption of a stable population, the simple model should have been able to predict disease prevalence from incidence. However, it does it poorly, presumably because the data on incidence and prevalence derived from literature are very inaccurate. The literature data on prevalence and incidence vary from respectively 66 to 258 per 100,000 and 4 to 24 per 1000,000 (Table I), which strengthens this explanation. Another reason may be that the external validity of the simple model may be very limited and consequently is not an accurate reflection of reality.

For example in the simple model we used an average mortality while Parkinson's disease is a condition, which occurs mostly in the elderly population and especially in males. Parkinson's disease usually starts at an average age of 62 and consequently patients with this disease will have substantial higher mortality. The incorporation of age may not be necessary for determining the potential number of candidates for a first-line treatment, but it is required for an appropriate estimation of duration of treatment. The more complex model appears to correct the problem by using age-specific and gender-specific mortality rates, and as a result yield a closer relationship between modelled and observed prevalence rates using the observed incidence from the literature. Neither the presence of the various health states nor treatment affects in this model the relationship, because neither affects survival changing the equilibrium.

If reimbursement of pharmaceuticals is going to be based mainly on economic data derived from modelling studies, it is vital to carefully scrutinise and refine this type design. The data being used in a modelling study may come from a variety of sources and are subject to varying degrees of uncertainty. In a previous manuscript, the limitations of the various types of data sources were extensively discussed.[17] The main drawback of all existing data sources is that they have not been developed for economic evaluations, but for either administrative or medical/scientific purposes. Hence they may suffer from incompleteness (e.g. no information on either healthcare utilisation or clinical outcomes, especially QALYs), sufficient external validity and appropriate format of the information, which does not fit the structure of the model.

An extra concern about the financial analysis is the epidemiology data. When the model yields a discrepancy between the relationship between incidence and prevalence compared with literature data, we have to validate the model structure and epidemiological literature or identify other reasons, which may explain the difference. In particular, epidemiological data for other countries may not be applicable to the study country. When the discrepancy cannot be explained we recommend giving more weight to prevalence than to incidence data. Hence, incidence may be varied outside its range in

order to keep prevalence within an acceptable range. Prevalence has a higher impact on outcomes of financial analysis and accuracy of prevalence data is usually higher compared with incidence data.
The model allowed various scenario analyses. In this example we only showed the impact of a scenario assuming that only patients who start developing fluctuations are going to use AP. Other scenarios are possible. Our analyses determined all potential candidates who are going to use the new drug, which assumed a 100% substitution. The current model also allows scenario analyses on the substitution effect. Another assumption is that population size is constant and that the population distribution with regards to age is constant, which is valid when predicting potential candidates of a drug within 10 years after launch. However, the model allowed the incorporation of aging of a population by changing the assumption that mortality rate equals birth rate.

# CONCLUSION

Our conclusion is that one model can be used for an appropriate assessment of the budgetary impact and the cost-effectiveness of a new drug. The main methodological advantage is that the use of one model will increase the consistency between the financial analysis and the health economic analysis by using the same underlying model structure, its assumptions and data sources.

### References


1 Riteco JA, Heij LJM de, Luijn JCF van, Wolff I (1999) Richtlijnen voor farmaco-economisch onderzoek. College voor zorgverzekeringen: Amsterdam
2 Drummond MF, et al. Current Trends in the Use of Pharmacoeconomics and Outcomes Research in Europe. Value in Health 2(5):323-32
3 CBS. 1999
4 Mortality in DATATOP: A multicenter trial in early Parkinson's disease. Parkinson's Study Group. Annals of Neurology. Vol 43 No 3. March 199
5 Barr J, Schumacher GE. Using decision analysis to conduct pharmacoeconomic studies. In Spilker B (Ed) Quality of life and pharmacoeconomics in clinical trials, pp 1197-1214 , Lippincott-Raven, Philadelphia, 1996
6 Sonnenberg, F, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993; 13: 322-338
7 Surcliffe RLG et al. Parkinson's disease in the district of the Northampton Health Authority, United Kingdom: a study of prevalence and disability. Acta Neurol. Scand. 1985; 72: 363-79
8 Dodel R. Et al. Costs of drug treatment in Parkinson's disease. Movement Disorders. 1998; 13 (2): 249-254
9 De Pedro Cuesta J. Parkinson's disease occurrence in Europe. Acta Neurol. Scand. 1991; 84: 357-365
10 Rajput AH et al. Epidemiology of parkinsonism: incidence, classification, and mortality. Ann Neurol. 1984; 16: 278-82
11 Koller WC et al. Handbook of Parkinson's disease, 2nd ed. New York, NY: Marcel Dekker Inc. 1992
12 Effect of Lazabemide on the progression of disability in early Parkinson's Disease. The Parkinson Study Group. Ann Neurol. 1996: 40; 99-107
13 Safety and efficacy of pramipexole. The Parkinson Study Group. JAMA. 1997: 278; 125-130
14 Nuijten MJC. Bridging decision analytic modelling with a cross-sectional approach. PharmacoEconomics. 2000,Mar: 17 (3): 227-236
15 Central Tariffs for consultations and procedures in healthcare. Utrecht. College Tarieven Gezondheidszorg 2000
16 The Economic Impact of Parkinson's Disease: An estimation based on a 3-month prospective analysis. Dodel R, Singer M et al. PharmacoEconomics, September 1998; 14(3): 299-312

17 Nuijten MJC. Data management in modelling studies: the selection of data sources". PharmacoEconomics. March 1998; 3(3): 305-316
18 Taxe. s'-Gravenhage: KNMP, 2000.

Chapter 12

# Reporting Format for Economic Evaluation: Focus on Modelling Studies


## SUMMARY

This article presents the first version of a reporting format for modelling studies which is based on a general reporting format by our taskforce, which was published in the previous issue of this journal. The use of decision-analytical models for economic evaluations is increasing because, in practice, it is not always possible to derive information from prospective studies. However, the acceptance of modelling studies is generally lower than prospective studies not only because of the use of secondary data, but also because the reports of modelling studies do not always have sufficient transparency. Hence, a standardised reporting format may improve the transparency and, consequently, the acceptance of modelling studies. This article presents an example of a reporting format for economic evaluation based on modelling studies, which may facilitate the development of future guidelines for modelling studies. The format consists of a number of headings, which are followed by a brief recommendation on the content. This format does not deal with methodology and data management, but especially addresses validation and quality assurance, which may increase the transparency of the report.

## INTRODUCTION

Economic evaluation is beginning to be used widely to support decision makers in health services for allocation of scarce healthcare resources. Because economic evaluation is used for pricing and reimbursement issues by authorities or third party payers, it is vital that these analyses be performed according to generally accepted and standardized methods. In order to increase the scientific quality and integrity of pharmacoeconomic studies, national guidelines have been developed by various countries (Australia and Canada).[1,2,3] However, in practice it is not always possible to derive information from scientifically sound prospective studies. In these cases decision-analytic models may be used to provide some of the missing information. Decision analysis is an explicit, quantitative, prescriptive approach to medical decision-making and allows both clinical and economic consequences of medical actions and attitudes to be analysed under conditions of uncertainty.[4] The model resulting from the decision analysis must correspond, as much as possible, to the real life situation of the disease and should reflect actual treatment patterns with input values (probabilities and items of healthcare utilization) deviating as little as possible from population values. Models may take the form of simple decision-analytic trees or they may be very complex Markov models of underlying disease processes and treatments.

The acceptance of modelling studies is generally lower than prospective studies because of the use of secondary data, but also because the reports of modelling studies do not always have sufficient transparency. Hence a standardised reporting format may improve the acceptance of modelling studies. A general reporting format for economic evaluation was described in a previous issue of this journal.[5] Since, contrary to clinical trial data, there are no accepted methods for data collection and analysis for modelling studies, a more disaggregate reporting presentation is required for modelling studies.

This paper presents an example of a reporting format for economic evaluation based on modelling studies (table I), which may facilitate the development of future guidelines for modelling studies. The format consists of a number of headings, which are followed by a brief recommendation on the content. The report is organised according to the ranking of the headings. If the structure of the report is different, markings made directly on the report indicate where in the report the content of each heading can be found. All headings are addressed in the report. If a heading is not relevant, it is not omitted, but a brief reason is given.

## REPORT INTRODUCTION

The introduction of our report for a modelling study contains all the necessary relevant epidemiological, clinical and economic information, identical to what is already required for a standard prospective study. The purpose of any introduction is to describe the rationale of the study and to present the information on which the study hypothesis is based. However, for a modelling

study we also incorporate information on the reasons for choosing a modelling study and a justification for the design and methods employed in the study. Table II outlines the information used for each step in the decision-making process for a modelling study.

**Table I. Reporting format for a modelling study.**

REPORT INTRODUCTION
- Epidemiology and treatment
  - Prognosis
  - Disease progression
  - Local treatment pattern
- Economic impact

STUDY DRUG
HYPOTHESIS
OBJECTIVES
DESIGN
- Analytic framework
- Patient population
- Comparator
- Analytic horizon
- Perspective
- Setting
- Clinical measures
- Effectiveness measures
- Economic measures

METHODS
- Healthcare system
- Model description
- Data sources
- Data collection
- Probabilities
- Healthcare utilisation
- Data analysis
- Sensitivity analysis
- Discounting

RESULTS
- Intermediate results
- Final results

CONCLUSION
DISCUSSION
VALIDATION AND QUALITY CONTROL
- Validation
- Quality control
- Software

RELATIONSHIPS
APPENDICES
REFERENCES

## Epidemiology and Treatment

The introduction for a modelling study also contains, in addition to estimates on the prevalence and incidence of the particular disease, information on age, gender and risk factors. The clinical

information consists of a description of the pathology, including underlying pathophysiological mechanisms, and on the prognosis, disease progression, and existing treatment options, all of which are relevant to the design of the model.

### *Prognosis*

A patient may fully recover without any higher risk for recurrence or other morbidity than the general population. On the other hand the patient may have a worse prognosis, a possibility which is discussed in the next section on disease progression. The information on patient prognosis is needed for the choice of the type of model (decision tree or Markov model) that will be employed in the modelling study.

### *Disease Progression*

In many diseases, patient may not achieve a complete recovery. There may be: (i) an increased risk for a relapse (e.g. depression); (ii) incomplete recovery (e.g. stroke); or (iii) other morbidity (e.g. myocardial infarction due to angina pectoris). In addition, the disease progression may be characterised by: (i) episodes with intermittent symptom-free periods; (ii) exacerbations, from which there may be only partial recovery, leaving the patient in a worse health state after each episode (e.g. Multiple Sclerosis (MS); (iii) a more continuous chronic character (e.g., Parkinson's Disease). There may also be risk factors that may act as triggers (e.g. fever in MS). This information is needed to decide on the characteristics of a Markov model (health states, cycle time, follow-up period) and effectiveness measure(s).

### *Local Treatment Pattern*

The current treatment options (usual care) are discussed with their onset of action, efficacy, side-effects, serious adverse events and compliance. Information from related clinical studies may be presented. In addition to drug therapy, other relevant treatment strategies (e.g. surgery), including no therapy, prevention, and screening are also presented. This information is needed to decide on the comparator(s) and clinical events which will be incorporated into the model structure. Information on the clinical events may also be used to decide on effectiveness measure(s). In addition to the first line therapy, treatment options for treatment failures are described, and this information may need to be incorporated into the model, depending on local treatment patterns. This section also contains a description of relevant international or national guidelines, which may be used for the development of the model structure. The information on the local treatment patterns may be used to decide on Markov characteristics (Markov states and cycle time). Finally, information on the type of physician (GP or specialist) and setting (in-patient or out-patient) associated with each treatment option, may help to define the perspective of the study and in the definition of economic outcomes.

### Economic Impact

The economic information refers to any previous related studies on the relevant pathology (cost-of-care studies) and associated therapies (cost-effectiveness studies). Of special interest may be country specific studies, because results from studies in other healthcare settings may not be directly extrapolated to the study country. Information on direct medical costs, direct non-medical costs and indirect costs may permit a judgement as to the type of costs to be included and on the perspective(s) of the study. Previous studies may also yield information on main cost drivers and cost structure.

## STUDY DRUG

This section contains a detailed description of the characteristics of the study drugs: indication, onset of action, efficacy, side-effects, serious adverse events, intermediate outcomes and compliance. The drug is also be compared with other therapies. A summary of the clinical trials is given, including information on the design, study population, follow-up period and clinical outcomes. The indication and onset of action will determine the study population. The information on clinical events associated with the study drug is needed for decisions on model structure and in defining the effectiveness measure(s). If the study drug is already on the market, information on market shares, niche and prescribers (e.g. GP or specialist) may be helpful in deciding on the perspective of the study and in defining economic outcomes.

## HYPOTHESIS

The hypothesis for a modelling study will be based on information in the "introduction" and "study drug" sections. The rationale behind the hypothesis is described in detail. This section does usually not contain any more information than that provided for a prospective pharmacoeconomic study.

## OBJECTIVES

The description of the objectives of a modelling study will be based on the hypothesis and does not usually differ from the objectives of a prospective pharmacoeconomic study. Any limitation of the scope of the objective due to the availability of relevant data is mentioned.

## DESIGN

The choice of a modelling study to prove the study hypothesis is justified, because the acceptance of modelling studies is lower than studies based on primary data. The reasons for doing a modelling study instead of a naturalistic study may be scientific (chronic disease) or purely practical. For example, pharmacoeconomic data may not have been required during the period of a clinical trial, but if such data have become required at submission, the only practical solution is a modelling study. After the decision has been made to use a modelling study, the choice of the type of model is justified; decision tree models or Markov models. Decision tree models are indicated for describing the

treatment pattern for an acute episode of illness leading to full recovery or death (e.g. community-acquired pneumonia) while a Markov model represents a convenient way of modelling chronic diseases. In an ideal Markov model, a life-time follow-up is recommended. The choice of the type of model employed is justified, especially in cases where there are deviations from the above mentioned general rule. For example, when a study is performed from the hospital perspective, a decision tree model may be sufficient to determine the costs and effectiveness of a treatment during hospitalisation.

### Analytic Framework

The main analytic framework of the study is described: cost-minimisation, cost-consequence, cost-effectiveness, cost-utility or cost-benefit. The choice of the framework is justified, especially when it is has been determined by the limitations of a modelling study design. The data sources for development of the analytic framework are given and explained (e.g. clinical textbooks, other literature data and expert opinion). All assumptions are listed in a table, including the process of the creation of the assumptions and the validation process.

### Patient Population

The hypothetical study population of the model is described, which will be based on the clinical trial(s) study population. Hence the most relevant inclusion and exclusion criteria of the clinical trial(s) are mentioned. The underlying assumptions for the selection of the hypothetical study population, and the validation process for these assumptions, are also described. Subsequently a judgement is made as to the dissimilarities between the hypothetical study population, which should have a high external validity, and the limited clinical trial population(s). Factors that can limit the application of the results are discussed (e.g. differences in resistance across countries in community-acquired pneumonia).

### Comparator

The choice of comparator is primarily based on the country-specific guidelines (e.g. most widely used, cheapest etc.) and this choice is justified. For example, there may only be clinical data available for one other drug, information which can be incorporated into the model. In the event that there are no local guidelines, a justification of the choice of the comparator is given.

### Analytic Horizon

The general rule for the determination of the study period in a modelling study is that a decision tree model covers the period from treatment initiation until recovery or death, while in a Markov model, ideally a life-time follow-up is recommended. The choice for the study period is justified, especially in cases of deviation from the above mentioned general rule, when the main reasons are discussed. For instance, in chronic diseases, a study period from one to five years may be sufficient, especially when extrapolation of short-term clinical trial results may not be acceptable from a clinical perspective.

### Perspective

The choice of perspective ideally depends on the treatment pattern (type of physician, setting), the rules of reimbursement, and country specific pharmacoeconomic guidelines. Previous economic studies and market information may help in defining the perspective. For example, if another study showed that indirect costs have a high impact, the society perspective may not be abandoned. The choice of the perspective is justified, and the rationale is explained, which may be due to the limitations associated with modelling. For example, if reliable information on working days lost is not available in the literature or other data sources, a society perspective may have to be abandoned since the credibility of expert opinion for this type of data is limited.

### Setting

The model will be developed for the healthcare setting of the study country. The setting will depend on the perspective of the study and the local treatment patterns, including the indications of the study drug and comparator. Previous economic studies and market information may help in identifying the perspective. A justification is given for any deviation form this rule, that might arise because of limitations of the modelling design. For instance, patients with depression are usually treated by GPs and psychiatrists. However, when the database, which was used for the study only contains information from patients being treated by psychiatrists, then the setting may be limited to that of the psychiatrist.

### Clinical Measures

The choice of clinical measures, which will be presented in the cost-of-consequences table, are explained and assessed on clinical events associated with local treatment, including the study drug (and comparator) and management of patients with a treatment failure. In addition, information on clinical measures which may have been used in previous pharmacoeconomic studies in the pathology under investigation, is also presented. Clinical events (e.g. adverse events) may be neglected, when they not differ statistically between study drug and comparator.

### Effectiveness Measures

The choice of effectiveness measures is explained, and may be based on disease progression, clinical events associated with local treatment, including the study drug and comparator, and scales of measurement, used in previous clinical and pharmacoeconomic studies in the pathology under investigation. The preferred effectiveness measures are mentioned, which would have been chosen in a regular prospective study. When the model design does not allow the incorporation of preferred clinical measures, the reasons are explained in this section. For instance quality of life, which deals with subjective feelings and emotions, cannot always be incorporated into a model structure. In this situation, alternative effectiveness measures which can be used for a modelling study are described

[e.g. time without symptoms/toxicity (TWIST)]. Subsequently the choice of the effectiveness measure being used in the model is justified, including a description of the validation process.

### Economic Measures

The choice of the relevant economic measures will be primarily be based on the pathology under investigation and the perspective of the study, and is ideally identical to the economic measures which would have been chosen in a prospective study. Previous economic studies and market information may support the choice. When the model design does not allow the incorporation of the preferred economic measures, this is explained in this section as well as the steps taken to deal with this limitation.

## METHODS

### Healthcare System

This section contains a general description of the healthcare system. The general and specific rules for reimbursement for the pathology under investigation are described. This section is not dependent on the design of the study (prospective or modelling).

### Model Description

This section contains a complete description of the structure of the model, including a figure of the model. The description allows the reader to follow every patient from the initial treatment until the end of the study period, including patients with treatment failure. In the case of a Markov model, this section contains a description of the Markov states (e.g. disease severity only or combinations of severity and treatment) and a justification. For instance, the Markov states may need to be based on health states (disease severity) and treatment, when the transitions do not depend only on health state, but also on treatment and previous treatment failures. Definition of Markov states will depend on information on disease progression and local treatment patterns, which may be supported by previous modelling studies or treatment guidelines. Validation processes are also be added. The patients' progression through the Markov states will be divided into cycles, which is described and justified (e.g., the cycle time corresponds to the time of the sequential therapeutic stages). In Markov models the duration of the cycle is justified, including the validation process. The choice of cycle time is usually based on information on disease progression, and local treatment patterns (e.g. duration of a therapy), which may be supported by previous modelling studies or treatment guidelines. The data sources for model structure development are given and explained (e.g. clinical textbooks, other literature data and expert opinion). All assumptions that were used in structuring the model are listed in a table, including the process of the creation of the assumptions and the validation procedures.

**Table II. Relationship between study choices and information.**

| | Epidemiology and treatment | | | Economic impact | STUDY DRUG |
|---|---|---|---|---|---|
| | Prognosis | Disease progression | Local treatment pattern | | |
| HYPOTHESIS | X | X | X | X | X |
| OBJECTIVES | X | X | X | X | X |
| DESIGN | X | | | | |
| • Analytic Framework | | | | | X |
| • Patient Population | | | | | X |
| • Comparator | | | X | | X |
| • Analytic horizon | | X | | | X |
| • Perspective | | | X | | X |
| • Setting | | | X | | X |
| • Clinical measures | | X | X | | X |
| • Effectiveness measures | | X | X | | X |
| • Economic Measures | | | X | | X |
| METHODS | | | | | |
| • Choice for model | X | | | | X |
| • Type of model | X | | | | X |
| • Markov states | | X | X | | X |
| • Cycle time | | X | X | | X |

## Data Sources

The sources of model variables (probabilities, healthcare utilisation, therapeutic choices, costs/tariffs) are mentioned; clinical trials, literature (e.g. meta-analysis), databases, medical records, Delphi panels and/or official tariff lists. The data sources are described in sufficient detail. For example, the description of clinical trials contains the results of the trial and details of the design, such as the study population, follow-up period of patients, follow-up of drop-outs and treatment failures (intention-to-treat), method of randomisation and primary and secondary clinical outcomes with confidence intervals. The strengths, weaknesses and possible sources of bias, that may be inherent in the data sources used in the analysis, are described. Selection criteria for studies and databases are discussed and an indication is given of the direction and magnitude of potential bias in the data sources which were used. The search strategy is described: Keywords, Journals, Language, Timeframe (e.g. 1987 to date), databases (e.g. MEDLINE, HealthSTAR, HSRProj, HSTAT etc.).

Because of criticism of Delphi panels, the use of this data source is justified by explaining that no other data sources were available. The use of Delphi panels does not follow a clear and consistent pattern; different researchers apply the terms Delphi, modified Delphi and expert panel in different ways. Moreover the basic methodology is often not followed, with both errors in the application of the techniques and in the justification for their use. Therefore this section contains a detailed description of the methodology (e.g. the inclusion criteria for the participating experts) and the process (e.g. number of rounds, interviews or mail).

## Data Collection

The methods and processes for data collection (e.g. for a Delphi panel) and data abstraction (e.g. for a database) is described and explained. The data collection forms which were used in the study are included in the appendix of the report (e.g. the questionnaire for the Delphi panel, or the abstraction protocol for the database). The rationale behind the data collection forms is explained by showing the relationship between the model (structure and variables) and the form. The procedures used in the creation of these tools and the validation process are described. For example, the structure of a questionnaire may have been developed with the help of one clinical opinion leader, and validated by an advisory board of experts.

## Probabilities

The methods for the determination of the probabilities are described. In Markov models the adjustment of all probabilities to the cycle time of the model are explained (e.g. the actuarial method). Assumptions and methods used in extrapolation of short term results to the follow-period in the model are described and justified, including the validation process.

In addition to determination of the above mentioned fixed input values for the model, the method used for determination of the range for each probability is described and justified (e.g. first and third quartile in the case of limited data).

The appendix contains a table showing all probabilities, listing the fixed input plus method and assumption, minimum and maximum plus method, data source and assumptions. In addition, for a Markov model the methods and associated assumptions used for calculation of the transition probabilities are listed. Finally the quality process employed for data entry (e.g. double entry or single entry with visual control) and data processing (e.g. visual control of all formulae) is described.

## Healthcare Utilisation

The methods used for the determination of the units of healthcare utilisation (units of direct non-medical costs and working days lost) are described in detail. In Markov models the allocation of healthcare utilisation to Markov states and transitions is described and any assumptions, including validation processes, are mentioned. This section also contains a description of the costing information, including the data source, for each unit of healthcare utilisation, which will depend on the perspective of the study.. The units of healthcare utilisation are defined in tables for each Markov state and transitions in a Markov model as well as for each branch in a decision tree model. The costing information may consist of actual costs, prices or tariffs, reimbursement percentages and co-payment, and data sources.

In addition to determination of the above mentioned fixed input values for the model, the method used for determination of the range for each unit of healthcare utilisation and price/tariff is described and justified (e.g. first and third quartile in cases of limited data).

The appendix contains a table showing all units of healthcare utilisation, listing the fixed input plus method and assumption, minimum and maximum plus method, data source and assumptions. A similar table is shown for the costing information. Finally, the quality process is described for data entry (e.g. double entry or single entry with visual control) and data processing (e.g. visual control of all formulas).

### Data Analysis

The standard analysis used is described. This will usually be a foldback analysis, which is an analysis based on the fixed input variables and yields the average effectiveness and costs for each treatment strategy. In situations where the study drug leads to a higher effectiveness and higher costs, an incremental analysis is performed (e.g. cost per life-year gained).

### Sensitivity Analysis

The execution of a sensitivity analysis and the interpretation of the results is not yet standardised, and this may lead to a subjective interpretation of the results which may consequently weaken the value of the economic evaluation. Hence the choice of the variables on which a sensitivity analysis is performed is justified and the rationale for the interpretation of the results of such an analysis defined. Sensitivity analyses are performed for the clinical variable on which the study hypothesis is based, and on the main cost driver.

### Discounting

This section will usually not depend on the design of the study (prospective or modelling), and will therefore not differ from country-specific or international guidelines.

## RESULTS

### Intermediate Results

This section starts with a reference to the appendix, which contains all the intermediate results of the study. This may be spreadsheets of all data collection forms, including results of data processing. For example, the estimations of all individual opinion leaders may be given, together with the fixed input value and range, after determination of the median and the first and third quartile. Results from the literature search (e.g. meta-analysis) may be presented in a similar way. The appendix also includes a table with the prices (or tariffs) of units of healthcare utilisation from the society and study perspective, including the origin of the data source.

### Final Results

The presentation of the final results may start with a description of the country-specific treatment patterns, based on the therapeutic choices after initial treatment with the study drug or the comparator. Subsequently, the average clinical and economic outcomes are listed in a cost-consequence table, in

which the economic measures are expressed, both in units of healthcare utilisation and costs. Another table presents the effectiveness and economic outcomes, including incremental cost-effectiveness outcomes, if appropriate. Also the total cost structure associated with each treatment strategy is shown. The disease progression over time and the timing of the different costs over time are presented in a graph or table, which may be combined in one figure to show the relationship between clinical and economic outcomes for each initial treatment. The results of all sensitivity analyses are presented using graphs or tables.

## CONCLUSION

This section describes the interpretation of the results and concludes if the study hypothesis has been proven by the results of the study. Secondary conclusions may also be addressed e.g. main cost driver, cost structure. The robustness of the conclusion may be addressed by discussing the results of the sensitivity analyses.

## DISCUSSION

This section contains a validation of the results of the model by comparing results from other studies or expert opinion. The observed differences with other studies are addressed and explained, along with the impact of the limitations and level of uncertainty in the modelling study. The discussion may conclude with recommendations to solve any existing uncertainty with regards to the final conclusion (e.g. a future prospective study might be recommended).

## VALIDATION AND QUALITY CONTROL

### Validation

If relevant, the general validation methodology and processes are described, in addition to the specific descriptions included under each heading.

### Quality Control

The process of Quality Control is described, dealing with subjects such as quality control of data entry and data processing (e.g. calculation of transition probabilities).

### Software

The software being used is listed, together with a description of the activities for which it was used. (e.g. Data 5.1 for modelling, Excel 5.0 for determination transition probabilities, Word 6.0 for report writing). It is also indicated if the software was validated and, if so, by whom, and when.

## RELATIONSHIP

This section discloses relationships between the performer of the study and the sponsor. Actions taken to avoid any potential bias are described (e.g., using a Medical Advisory Board, or validation by independent experts).

## APPENDICES AND REFERENCES

Especially in modelling studies the disclosure of relevant appendices to reports will increase the transparency of the results. The appendices may cover the following information, which has been mentioned in relevant headings: list of variables and data sources, list of assumptions, list of formulae, the names and addresses of participating experts.

### References


1 Commonwealth Department of Human Services and Health. Guidelines for the pharmaceutical industry on preparation of submission to the pharmaceutical benefits advisory committee. Canberra: Commonwealth Department of Human Services ,1995

2 Australia Commonwealth Department of Health, Housing and Community Services. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Commonwealth Department. 1992

3 Canadian Coordinating Office for Health Technology Assessment (CCOHT). Guidelines for economic evaluation of pharmaceuticals: Canada. 1st ed. Ottawa: CCOHTA. 1994

4 Weinstein MC, Fineberg HV. Clinical decision analysis. Philadelphia (PA): WB Saunders Co., 1980: 228-65

5 Nuijten MJC, Brorens MJA., Hekster YA, Van der Kuy A, et al. Reporting Format For Economic Evaluation: Part I: Application to the Dutch Healthcare System. Pharmacoeconomics 1998: 14 (2): 159-63

Chapter 13

# Discussion

# INTRODUCTION

The objective of this thesis was to identify and explore various types of uncertainty in modelling studies. Also methodologies were introduced, which may reduce the level of uncertainty associated with modelling, and which may consequently increase the acceptance of modelling studies. The concepts were empirically illustrated using Markov models in chronic diseases: depression, Parkinson's disease and multiple sclerosis.

The review by Buxton *et al* explored the role of modelling in economic evaluation and concluded that the concerns about models mainly relate to the trade-off between internal and external validity.[1] In this thesis I explored various types of uncertainty associated with the data used in the model, which relate to the internal-external validity issue raised by Buxton. I also examined in depth the measurement of uncertainty by means of sensitivity analysis and the worth of collecting extra data.

# DATA SOURCES: INTERNAL VERSUS EXTERNAL VALIDITY

## Validity of Data Sources

**Chapter 4** contains an assessment of the validity of various data sources for the variables being used in a model. In general most of the data sources (clinical trials, meta-analysis), which yield the clinical data for the model lack external validity, because those data are usually derived from, randomised clinical trials with a high internal validity. The data sources, which yield data on healthcare utilisation also suffer from lack of external validity, when those data are derived from randomised clinical trials or clinical databases. Alternative data sources for healthcare utilisation are administrative databases, which contain data on healthcare utilisation with a high external validity. Those databases often suffer from several limitations, because they were not developed primarily for health economic evaluations. Consequently the format of the information may not always fit the structure of the model and the data set may not be complete, for example when inpatient data have not been collected for reasons of reimbursement. Also there is in general a lack of appropriate databases in Europe, contrary to the US. The use of medical records suffers from similar limitations and access may be complicated because of legislation associated with privacy protection.

**Chapter 4** also contains a number of recommendations for good practise, which operationalises the concept of transparancy. This chapter proposes to include in the report a very detailed description of all data identified as potentially relevant for the model together with some critical appraisal of each item of data, finished with a formal justification of the choice of each item of data. Such an approach has been recently welcomed by other authors.[2]

## Use of Cross-Sectional Data

In **Chapter 4** the advantages and disadvantages of different sources of data are discussed. A main concern is the use of a panel of experts to gather data, which cannot be derived from the literature,

because there is no guarantee that such an assessment of resource utilisation and utilities accurately reflects reality. Many health economic modelling studies are still based on data derived from Delphi panel, especially in Europe due to lack of appropriate other data sources (e.g. databases). Considering that the results from modelling studies are increasingly used in reimbursement decisions for new drugs, it is obvious that potential bias due to the use of Delphi panels should be minimised. **Chapter 5** describes a cross-sectional study as an alternative data source for a modelling study. This hybrid design is aimed at bridging the probabilities derived from literature and clinical trials with information on costs and utilities from a cross-sectional study.

**Chapter 7** shows that the use of a hybrid design introduces another type of uncertainty due to the impact of confounding variables: the costs and utilities derived from the cost-of-care study, may not only be a function of the defined health states in a Markov model, but also of other variables, which may act as confounding variables when they are not taken into account and consequently reduce the external validity of the model. A methodology was presented to increase the external validity of a Markov model by the incorporation of confounding variables in the model. This study showed that the inclusion of confounding variables substantially changed the outcome of the model. A severe potential limitation is that the incorporation of multiple variables necessitates an increase of the sample size of the cross-sectional study in order to obtain statistically significant results, which may limit the feasibility of this approach.

The negligence of a confounding variable lead to uncertainty relating to generalisability of results as proposed by Buxton and Briggs, which deals with the extent to which the results of a model, as they apply to a particular population, hold true for another population.[1] [3] **Chapter 7** showed that the distribution of confounding variables in patients in a clinical trials or cross-sectional study may not correspond with the actual distribution in the population. Consequently the cost-effectiveness results of a model, which is based on those data sources, may be severely biased.

Finally the ignorance of a confounding variable can also be considered an oversimplification, which is according to Eddy the most common error in modelling and which occurs by omitting important variables.[4]

## Use of Registers before Launch

The statistical constraints may be solved by the use of registers instead of cross-sectional studies. A register and a cross-sectional study are both observational studies, but the register has a much larger sample size and the data collection in a register is more comprehensive. Registries use large longitudinal, observational studies designed to measure the impact of a particular disease or condition on clinical and patient-specific outcomes, and to document the outcomes associated with different treatments or settings of care. Patients are followed prospectively and data are collected on disease severity and clinical outcomes as reported by clinicians, as well as resource use, functional status and quality of life as reported by the patient. Currently targeted longitudinal observational databases, or

patient registries, are being designed, which reflect the current treatment patterns without influencing the treatments or interventions and consequently the observational study is fully naturalistic without any intervention with real practice (e.g. no randomisation) and has a high external validity. The use of a register suits perfectly with the concept of evidence-based medicine (EBM), which means that clinical encounters should be supported by scientific conclusions based on data as much as possible. The large size of a register allows the development of statistically solid multiple regression equations with high external validity, which can be incorporated in a health economic model. Summarising, the cost-effectiveness of a new product can be modelled by combining data on the probabilities derived from literature and clinical trials with information on costs and utilities from a register, which is primarily developed for the collection of health economic data in daily practice. Thus a register also yields data with a high external validity like the cross-sectional study, but its large sample size also reduces the statistical constraints of a cross-sectional study.

Although the incorporation of healthcare utilisation data from a register into a model substantially increases the external validity and reduces the uncertainty of the input data, this methodology remains only a simulation of the cost-effectiveness of a new drug. The input transition probabilities for the new drug are still based on efficacy and safety data derived from the phase III clinical trial, which suffer from limited external validity. Another concern is that the study population in a trial is often not representative of the population under treatment, meaning that results of trials cannot necessarily be generalised to the population actually being treated. For example, until recently, women were underrepresented in many trials. Even today the elderly are frequently underrepresented and those with co-morbidities are often excluded. Although the guidelines accept modelling techniques to extrapolate efficacy to effectiveness outcomes, the extrapolation from cost-efficacy to cost-effectiveness outcomes introduces an extra type of uncertainty. In **Chapter 6** we extrapolated the short-term clinical trial data on progression in the multiple sclerosis clinical trials to a life-time effectiveness outcome (QALY's). This method of extrapolation was validated by comparing the modelled life-expectancy with the life expectancy from an observational study, which was not used in the model construction. Eddy defined this type of validation as a 3-order validation.[4]

## Use of Registers after Launch

The register can yield real-life data for the comparator in the health economic model, which are based on data from daily practice. However the register usually does not contain any data on efficacy and safety of the new drug, when it is not yet reimbursed. Therefore reimbursement decisions for a new drug will always be constrained by this paradox, which requires effectiveness data before the new drug is used in daily practice. When decisions on reimbursement of pharmaceuticals will be based predominantly on health economic data, this paradox does not allow a final decision for the reimbursement of the drug from a methodological point of view, especially because of the high consequences for the patient population. Therefore temporary acceptance of an innovative drug to the

reimbursement package might be considered. A conditional acceptance would permit initial decision-making on reimbursement based on the cost-effectiveness of the new drug derived from modelling data, followed by validation through subsequent prospective data collection after reimbursement of the new drug by means of a naturalistic study. However such a study would require a much larger sample size than a Phase III trial in order to show statistically significant differences in health economic outcomes, because of the more heterogeneous patient population. Consequently the set-up of a naturalistic study would be constrained by logistical requirements. Alternatively the initial register may be used to yield real-life data for the new drug during the period of temporary reimbursement and continue data collection after launch of the new drug. The new data can replace the initial input data for the new drug in the health economic model. Consequently the model can be used to reassess the cost-effectiveness of the new drug based on real life data, which corresponds with a 4$^{th}$ order validation.[4] However this approach is only meaningful, when the conditions under which an intervention is actually implemented, closely reflect those assumed in the model.

The register may also yield more statistically solid safety data with high external validity because of the large sample size of the register compared with the clinical trial. The large sample size of the register may also allow the identification of any type of covariance, which could not be incorporated in the initial model because of lack of power in the clinical trial and lack of sufficient real-life data on the new drug. The new study drug may not only show an improvement on the primary efficacy outcome in the clinical trial, but the new drug may also be superior (or inferior) on other clinical outcomes compared with standard therapy. Such other clinical outcome may act as a confounding variable, when it is not incorporated in the model, because the transition probabilities and the costs/utilities of the health states may also be a function of this other clinical outcome. The impact of confounding variables has been illustrated in **Chapter 7**, which showed that the outcomes of a health economic model for Parkinson's disease are severely biased, when another measure of disease severity (Hoehn&Yahr scale) is not taken into account in addition to only the level of fluctuations.

Observational data, however, may also have important limitations through the non-random decisions of clinicians that introduce bias. [5] In observational data, patients would often be treated differently based on the underlying condition. In such cases, comparing outcomes across patients would confound the effect the treatment with the effect of the underlying condition. This confounding, often referred to as a selection problem in the econometric literature, certainly exists in the case of catherisation of patients with an acute myocardial infarction (AMI).[6] By contrast, in a well-executed RCT, patient condition is independent of treatment, and one can therefore reasonably attribute observed effects to the particular variation in treatment being studied. In observational data the treatment is not allocated randomly. As a result, the characteristics of those obtaining the treatment will generally differ from the characteristics of those who do not get standard care (the controls). The differences may be observable characteristics such as age, in case a regression equation can potentially control for them.[7] Subsequently this regression equation can be incorporated in the health economic model for the 4$^{th}$

order validation of the outcomes of the initial model. A more serious concern is that the differences may be unobservable (to the analyst) characteristics that affect patients under treatment. For example, patients who are more severely ill in ways known to their physicians but not to the analyst might not get the treatment, or vice versa. If so, the effect of the treatment on the outcome is confounded with the severity of illness. This effect is well understood by clinicians and epidemiologists and is a principal reason why RCTs are regarded as the gold standard in clinical research. The use of observational data for health economic studies has also been questioned previously.[8]

However new methodologies have been developed, that may reduce the need for a RCT. For example, Newhouse presented an econometric technique, instrument variables, that can be useful in estimating the effectiveness of clinical treatments in situations when a controlled trial cannot be done.[6] This technique relies upon the existence of one or more variables that induce substantial variation in the treatment variable, but have no direct effect on the outcome variable of interest. The concept was illustrated with an application to aggressive treatment of acute myocardial infarction (AMI) in elderly, which showed that some of the differences in observed mortality between patients receiving catherisation and no catherisation could result from differences in the capabilities of the hospitals and physicians treating the two groups of patients, independently of the procedures given to them. For example hospitals with a cathererisation unit are more likely to have other sophisticated treatments available to their patients, and their physicians and nurses may be more highly trained. There are other examples, where observational data have resulted in conclusions about the effectiveness of therapy, which have subsequently been shown to be wrong such as the prophylactic administration of lidocaine to patients with AMI.[9]

## SENSITIVITY

Sensitivity analysis aims at providing information on the degree of uncertainty in economic evaluations. A sensitivity analysis is based on the modification of the basic clinical and economic estimates of parameters to judge the effect on study results of alternative assumptions for the range of potential values for uncertain parameters. The methods used, the choice of parameters and the range of these parameters must be stated and substantiated. If this procedure is followed for all the estimates in turn, then we refer to it as a 'univariate sensitivity analysis' or one-way sensitivity analysis. The conventional approach is to vary one of the input variables from its baseline values and to observe the effect on the outcome of the model. A limitation of the univariate sensitivity analysis is that it does not incorporate the probability distribution of an input variable and does not provide information about how likely the outcome will deviate from the base case values.

Probabilistic sensitivity analysis (PBA) permits the analyst to assign a range and distribution to input variables. A multivariable PBA examines the effect of simultaneous changes in different variables on the outcomes of the study, which is based on generation of random distributions around each parameter and -- in conformity with the distributions -- then tries to arrive at a new estimate for each

parameter. For each combination of estimates a new estimate of the costs, effects and cost-effectiveness ratio is therefore obtained. By repeating this procedure many times a random distribution can then be presented. In case of a monovariable PBA, the variation of only one input variable is incorporated.[10 11 12]

Agro *et al.* analysed the extent of reporting of sensitivity analysis in the health economic, medical and pharmacy literature.[13] They found that in 59% sensitivity analysis was conducted and 39% of the 59% stated explicitly that a sensitivity analysis was being performed. It is interesting to notice the distribution: 80% of the health economic papers contained a sensitivity analysis; those figures are respectively 70% and 20% for the medical and pharmacy journals. The simple sensitivity analysis was employed most frequently. In health economic journals a simple sensitivity was performed in 50% of the cases, while a probabilistic sensitivity analysis was performed in only 10%. Other types of sensitivity analyses included threshold analysis and analysis of extremes. The authors also state that health economic guidelines contain limited information and recommendations regarding sensitivity analysis methods and that there is a large variability between the guidelines. Briggs also found that a disappointingly low percentage of studies adequately handled uncertainty.[14] Sheldon found that models often do not carry out sensitivity analysis thoroughly. Influential variables about which there really is uncertainty are often not varied as part of the sensitivity analysis.[15] This could be because the authors endow it with false certainty or because the variables are hidden within the model. In those cases the sensitivity analysis is not only incomplete, but also it will underestimate the true uncertainty. In addition it is not sufficient to perform sensitivity analysis only on those variables, that are considered *a priori* very sensitive. There may also be variables, which appear to be sensitive after conducting the sensitivity analysis. Because of the complexity of the model *a priori* guesses of the sensitivity of input variables may lead to errors and consequently important variables may be excluded from the sensitivity analysis. A more important argument is that the selection of variables for the sensitivity analysis based on an *a priori* estimated sensitivity is in conflict with the objective of a sensitivity analysis, which is to identify the most sensitive variables. For example an antidepressant model did not vary critical assumptions on compliance, while the antidepressant model in **Chapter 3** showed the high impact of compliance on the outcomes of the model.[16] These finding relate to a recommendation in **Chapter 8** for univariate sensitivity analysis on every input variable of the model in order to avoid any bias and error in the selection of variables for the sensitivity analysis. In **Chapter 8** a more objective method was presented for sensitivity analyses reducing the amount of potential subjectivity. This new procedure, integrating point-sensitivity and range-sensitivity in a measure of "overall sensitivity", allows an objective judgment of the sensitivity of all variables in a model, permitting the variables to be ranked according to the degree of overall sensitivity. When this method is compared with the existing methods for sensitivity analyses, this method can be considered a special type of univariate sensitivity analysis, because the sensitivity analysis is performed on each input variable separately. A difference is that the standard univariate sensitivity analysis is based on

the determination of the difference in outcome (absolute univariate sensitivity) when an input value is varied within its range and depends only on the range of a variable, while the overall sensitivity also includes the responsiveness of the outcome to the input value at a fixed point (relative univariate sensitivity). Consequently the overall sensitivity yields more comprehensive information for the interpretation of the sensitivity of the input variables, which may be helpful in prioritising further research for reducing the range of certain input variables.

Another concern is the determination of the range to be used in the sensitivity analysis. In order to avoid bias in the selection of the range, the study should provide a source of explanation for the ranges used in a univariate sensitivity analysis instead of an arbitrary range. Therefore the range of an input variable may be derived from the stochastic variation of an input variable, for example the 5-95% confidence interval, where the 5% and 95% values are respectively the lower and upper limit for the sensitivity analysis. When the data do not allow the construction of a confidence interval, the range can be based on the minimum and maximum values. In **Chapter 8** this approach was used for the determination of the range of the input variables.

A limitation of this method is that the overall sensitivity measure is based on a subjectively chosen range, which excludes the impact of values outside the range on the overall sensitivity. The uncertainty about the value of point estimates of parameters reflecting stochastic variation is best represented as a probability distribution. **Chapter 9** presents a refinement of the initial method, as presented in **Chapter 8**, by the incorporation of the probability distribution, which allow a more accurate assessment of the level of uncertainty in the model. This method can be considered a special type of probabilistic sensitivity analysis (PBA). The refined method for the determination of the overall sensitivity can be considered a type of monovariable PBA, because it incorporates the probability distribution of one input variable, while the other input variables remain constant. On the other hand, there is a large difference in the outcomes. The level of sensitivity of a monovariable PBA is presented as an absolute statistical measure (SD, confidence interval). Contrary, the overall sensitivity is a relative measure, which includes the change of the input variable and consequently may be defined a relative monovariable probabilistic sensitivity analysis.

A constraint of the above-mentioned and other methods for sensitivity analysis is that these methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taking into account the existing relationships between these variables. **Chapter 10** introduces the concept of 2-order sensitivity, which captures this inter-variable uncertainty and shows the practical implications of it for the execution of sensitivity analysis. We presented initially a method assuming only uniform distributions, and subsequently various methods incorporating the real distributions. The findings from the 2-order sensitivity analyses can prioritise any subsequent prospective the data collection.

## Application

The results from sensitivity analyses can be used to judge the benefit of further research in order to reduce uncertainty in a modelling. The standard sensitivity analyses and the presented method of the overall sensitivity can be used to rank the input variables according to their level of uncertainty and prioritise further research for the most sensitive variables. The results from the 2-order sensitivity analyses show that the most sensitive variables in the model have also the most impact on the sensitivity of the other input variables. Consequently one can argue that the reduction of uncertainty associated with the most sensitive variables in a model, will lead to an overall reduction of uncertainty in the outcomes in the model because of a reduction in sensitivity of most variables. Hence it may be more important to investigate the uncertainty of the most sensitive variables of the model in more depth, e.g. by means of a meta-analysis, than to investigate the uncertainty associated with all variables. Oakly and Hagan investigated the relative importance of an input variable in driving the uncertainty of the outcome of the model.[12] They compare the uncertainty of the outcome due to the variance of all input variables with the uncertainty due to the variance of all input variables, excluding the variable under investigation. Subsequently all variables can be ranked according to their relative impact, which will guide the efforts to reduce uncertainty. For example if it was possible learn the true value for two input variables and the cost for obtaining those values would be the same, the variable with the highest relative impact is chosen. An alternative approach to judge the benefit of further research to reduce uncertainty in a modelling study is based on the concept of Value of Information.[17 18] This approach distinguishes the conceptually separate decision concerning efficient service provision given the level of information available, from the decision concerning funding further information collection. Here the decision of further data collection is based upon the expected cost of uncertainty, which is determined by both the extent of the uncertainty surrounding the efficient service provision and the consequences of this uncertainty. The extent of the uncertainty is measured by the error probability associated with the decision and the consequences of uncertainty are measured in terms of health benefits foregone when this uncertainty causes the incorrect decision to be made concerning service provision. Where these health benefits are valued according to society's willingness to pay for certain health outcomes, the approach gives a monetary value for the amount that society is willing to pay to reduce the uncertainty surrounding the service provision. Although the concept of Value of Information initially was applied to empirical data, it may also be applied to a health economic model. When an input variable appears to be very sensitive because of its large range, the concept of Value of Information may be applied by assessing the cost of extra research for reducing the confidence interval of that input variable by using more data. The primary objective of both the concept of 2-order sensitivity and the concept of Value of Information is to identify the impact of input variables in a model on the level of uncertainty within a model and to explore the worth of further research to reduce uncertainty of an input variable. The concept of Value of Information in a health economic model may be extended by incorporating inter-variable uncertainty,

which may further reduce the extra costs for the reduction of the uncertainty associated with the healthcare decision. For example the 2-order sensitivity analysis may show that the uncertainty associated with the clinical response rate depends mainly on the hospitalization rate. Consequently the extra cost of exploring this rate may be less than the extra costs for reducing the range of the response rate.

## CONCLUSION

Summarising, a model based on a hybrid design bridging clinical trial data and data from a register, seems to be the optimal design for yielding scientifically sound cost-effectiveness data at the time of reimbursement. I recommend a temporary reimbursement because of the above-mentioned potential sources of uncertainty, which can only be handled at the time of reimbursement by means of sensitivity analyses. The outcomes of the sensitivity analyses can guide the strategy for collecting real life data of the new drug, which can be used to replace the initial data in the model. Subsequently a final decision on reimbursement can be made based on the new cost-effectiveness outcome. This approach would minimise the logistical and methodological concerns related to current policy and it would also reduce the concern of industry that health economic evaluation guidelines would delay product launch, shortening the period of useful patent life and the return on R&D investment. New drugs would be made available more quickly if prospective data collection were not required prior to reimbursement. From the perspective that results from modelling studies are important for decision makers, it is obvious that recognising the relevance of uncertainty and appropriately dealing with it is required for obtaining unbiased results from health economic modelling studies, especially when those data are being used for reimbursement decisions. This thesis clearly showed the relevance of various types of uncertainty using specific examples. Another key research subject is the feasibility of the use of real-life data in the model by means of observational data collection (e.g. registers). Considering the disadvantages of observational data, further research is required in order to provide scientifically sound methods and appropriate data sources for health economic modelling.

### References


1. 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 Econ 1997 May-Jun;6(3):217-27
2. Brennan A, Akehurst R. Modelling in health economic evaluation. What is its place? What is it value?. Pharmacoeconomics. 2000 May;17(5):445-59. Review
3. Briggs A, Sculpher M and Buxton M. Uncertainty in the economic evaluation of healthcare technologies: the roel of sensitivity analysis. Health Economics. Vol. 3: 95-104: 1994
4. Eddy DM. Assessing medical technology. Washinton DC: National Academy Press; 1985; Technology assessment: the role of mathematical modelling
5. Rosenbaum PR. Discussing hidden bias in observational studies. Ann Intern Med 1991;115(11):901-5
6. New House JP, McCellan. Econometrics in outcomes research: the use of instrumental variables. Annu. Rev. Public Health. 1998. 19: 17-34
7. Bowden RJ, Turkington DA. 1984. Instrument Variables. Cambridge: Cambridge Univ. Press.
8. Simon G, Wagner E, Vonkorff M. Cost-effectiveness comparisons using "real world" randomized trials: the case of new antidepressant drugs [see comments]. J-Clin-Epidemiol 1995;48(3):363-73 issn: 0895-4356

9. Hine LK, Laird N, Hewitt P and Chalmers T. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. Archives of Internal Medicine 1989; 149: 310 1433-1438
10. Critchfield G, Willard K and Conelly D. Probabilistic sensitivity analysis methods for general decision models. Computers and Biomedical Research 1986; 19: 254-65
11. Doubilet P, Begg CB,Weinstein MC, et al. Probabilistic sensitivity analysis using Monte Carlo simulation. Med Decis Making 1985;5:157-77
12. J Oakley, O"Hagan A. Probabilistic sensitiviyty analysis of complex moels: A Baysian approach. Submitted to Journal of the Royal Statistical Society, Series B
13. Agro K, Bradley C, Mittman N, iskedjan M, Lane Ilersich and Einarson R. Sensitivity analysis in health economic and pharmacoeconomic studies. An appraisal of the literature. Pharmacoeconomics. 1997 Jan 11 (1): 75-88
14. Briggs A and Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. Health Economics. 1995; 4: 355-371
15. Sheldon TA. Problems of using modelling in the economic evaluation of healthcare. Health Economics. Vol. 5: 1-11. 1996
16. Jonsson B, Bebbington PE. What price depression. The costs of depression and the cost-effectiveenss of pharmacological treatment. British Journal of Psychiatry 1994. 164;665-673
17. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of healthcare technologies. Journal of Health Economics. 1999; 18: 341-64
18. Claxton K, Posnett J. An economic approach to clinical trial design and research priority setting. Health Economics 1996; 5 (6): 513-24

Summary/samenvatting

# In Search for More Confidence in Health Economic Modelling: Reducing Uncertainty Associated with Modelling Studies

Every government is eager to control the increase of expenses by the implementation of central cost containment policies particularly in relation to pharmaceuticals. For the most part, those measures have relied on budgeting or price controls, including negotiated prospective budgets for hospitals, centralized negotiated budgets for ambulatory physicians including drug prescriptions, and limitations on payments for particular medications. Because those traditional central cost containment measures were only partially successful, due to lack of incentives, the health authorities in Europe started to establish incentives for efficient healthcare delivery. Both traditional and recent containment measures focus especially on the pharmaceutical drugs sector in many countries, as these constitute a health technology that is relatively easy to introduce and implement compared to other forms of care.

Financing prescription medicines in ambulatory care has been a central responsibility based on the traditional clinical trial outcomes used for registration: efficacy, safety and quality parameters. Although there is large variety between the various countries, there are three related trends: decentralisation of the healthcare decision-making process, prescription restrictions, and extra data requirements. We can distinguish various extra data requirements all relating to the use of the drug in real daily practice, while the traditional clinical trial outcomes are only derived from randomised clinical trials. At a central level the demand for cost-effectiveness and budgetary impact data is increasing. The requirement for health economic data has resulted in some countries already to formal reporting requirements (e.g. Canada, Australia, The Netherlands, The UK, Portugal and Finland). Although the most evident impact of health economic studies is expected for central reimbursement audiences, evidence for the use of health economic studies by other audiences is expected to increase (e.g. patients, hospitals, insurers, formulary committees).

Pricing and reimbursement have been based, until recently, on the traditional clinical trial outcomes (efficacy, safety and quality parameters) used for registration, which are called the first three hurdles. Consequently the growing burden on manufacturers to demonstrate the cost-effectiveness of their products is called the fourth hurdle in drug development. This fourth hurdle may have considerable consequences for all players involved. Health economic data may be collected alongside a Phase III clinical trial of a new drug. However data derived from this so-called "piggy-back" trial suffer from external validity, because a clinical trial has strict inclusion and exclusion criteria and treatments are protocol driven. An alternative design is the naturalistic economic trial, which is a randomised trial with the primary objective of gathering "real world" or representative cost and effectiveness data. However since study drugs are usually not approved for registration at the time of performing economic studies, the use of prospective naturalistic trials is limited, making submission of data on effectiveness and expected costs at the time of reimbursement not feasible. On the other hand, projections about a drug's effectiveness and expected costs can be modelled using realistic and explicit assumptions based on data from clinical studies. In addition modelling often helps overcome the practical limitations of prospective studies, particularly for chronic conditions like Parkinson's disease that may require longer-term extrapolations of drug effects. In order to increase the scientific quality

and integrity of pharmacoeconomic studies, national guidelines have been developed by various countries. Although various groups have published recommendations on the good practice of economic evaluations [1 2 3], those guidelines mainly focus on prospective studies and contain only a limited number of recommendations for the execution of modelling studies. This may partly explain why the acceptance of modelling studies has generally been lower than prospective studies, which is mainly due to the level of uncertainty associated with modelling studies. If reimbursement of pharmaceuticals will be based predominantly on economic data derived from modelling studies, it is vital to scrutinise and refine the modelling approach carefully and especially the handling of uncertainty.

Therefore the objective of this thesis was to identify and explore various types of uncertainty in modelling studies. Methodologies are presented, which may reduce the level of uncertainty associated with modelling, and which may consequently increase the reliability of health economic outcomes of modelling studies. This may improve the acceptance of modelling studies, even in the absence of formal guidelines. The concepts are empirically illustrated using Markov models in chronic diseases: depression, Parkinson's disease and multiple sclerosis. Published real data are used, whenever possible, but extrapolation methods are used in absence of real data in order to illustrate the relevant issues.

The following types of uncertainty in modelling studies are explored:

- Uncertainty associated with the data sources providing the input data for the model. The reliability of the estimates depends on the choice of the data sources (selection criteria, external validity). The data may come from a variety of sources and are subject to varying degrees of uncertainty. The following issues relating to the selection of data sources were addressed: 1) Classification of data being used in modelling studies, 2) Assessment of the various data sources, including advantages and disadvantages, 3) General strategy on determining the appropriateness of a data source for a model, 4) Recommendations on a strategy for data source selection and a transparent reporting format for data source selection. A main concern in a modelling study is the use of a Delphi panel to gather data not available from actual existing data sources. There is no guarantee that the panel assessment of resource utilisation, utilities and patient preferences is an accurate reflection of reality. Therefore a cross-sectional study is presented as an alternative data source for a modelling study. It can be used to determine resource utilization and utilities for Markov health states. The overall design may be considered a hybrid between a naturalistic prospective study and a modelling study by maximising the pros and minimising the cons of both types of design.
- Uncertainty reduction through sensitivity analysis: Sensitivity analyses are generally based on only a limited number of variables without justification of the choice of selected variables or the chosen range of each variable, which may lead to subjectivity in the execution of a sensitivity

analysis. An objective method is presented based on the integration of point-sensitivity and range-sensitivity resulting in an "overall sensitivity". This approach avoids the subjective selection of variables for the sensitivity analysis and the potential bias in judging the degree of sensitivity in most current economic studies.

- Inter-variable uncertainty: The constraint of standard sensitivity analyses is that those methods only show the sensitivity of the outcomes to a change through a range of potential values for one or more variables without taken into account the existing relationships between those variables. A methodology is presented, which considers this type of inter-variable uncertainty, which is defined as a 2-order sensitivity. The practical implications of this inter-variable uncertainty for the execution of sensitivity analysis are shown. The results from the 2-order sensitivity analyses show that the most sensitive variables in the model have also the largest impact on the sensitivity of the other input variables. Consequently the reduction of uncertainty associated with the most sensitive variables in a model, will lead to an overall reduction of uncertainty in the outcomes of the model because of a reduction in sensitivity of most variables. On the other hand, it may be worthwhile to reduce the sensitivity of the most sensitive variable by using 2-order sensitivity analyses to identify the most influential variables.
- Uncertainty about confounding variables: The costs and utilities may not only be a function of the defined health states in a Markov model, but also of other variables, which may act as confounding variables when they are not taken into account. A strategy is presented for the incorporation of a confounding variable in Markov health states by means of health state specific relationships between the confounding variable and costs as well as time-dependent values of the confounding variable. Also the sensitivity of the outcomes of a model to the incorporation of a confounding variable is determined. The results show that the outcomes of a health economic model can be severely biased, when a confounding variable is not taken into account. This proves the need for incorporating confounding variables into a health economic model.
- Uncertainty associated with epidemiological data: A methodology is presented for an appropriate assessment of the budgetary impact of a new drug, which simultaneously can be used for a traditional cost-effectiveness analysis. A Markov model is constructed to validate the epidemiological data by proving the consistency between prevalence and incidence of Parkinson's disease. The analysis shows the substantial discrepancy between the prevalence found in the literature and the prevalence derived from the model, which is based on the incidence data from the literature. Consequently a population-based model has an additional uncertainty in epidemiological data compared with the patient-based cost-effectiveness model.
- Uncertainty associated with the reporting of a modelling study. Contrary to clinical trial data, there are no accepted methods for data collection and analysis for modelling studies. Therefore a more disaggregate reporting format is required for modelling studies, which addresses the various types of uncertainty in modelling studies and contains a justification of the choices in the selection of

data sources, model design and its assumptions, managing confounding variables and execution and interpretation of sensitivity analysis.

The conclusion of this thesis is that various types of uncertainty can be distinguished in modelling studies. Until now the standard approach of dealing with uncertainty is the execution of sensitivity analyses, which only address the uncertainty associated with the range of the input variables (statistical uncertainty). This thesis explores in more detail the execution of a sensitivity analysis and identified other sources of uncertainty. An objective method for sensitivity analyses is presented, which minimises the amount of potential subjectivity.

Uncertainty associated with sensitivity analysis: Sensitivity analyses are generally based on only a limited number of variables without justification of the choice of selected variables or the chosen range of each variable, which may lead to subjectivity in the execution of a sensitivity analyses. An objective method is presented integrating point-sensitivity and range-sensitivity, which allows an assessment of the "overall sensitivity". In addition inter-variable uncertainty in sensitivity analysis is explored. This study shows the practical implications of this inter-variable uncertainty for the execution of sensitivity analysis. The results from the 2-order sensitivity analyses show that the most sensitive variables in the model have also the most impact on the sensitivity of the other input variables. Hence it may be more important to investigate the uncertainty of the most sensitive variables of the model in more depth, e.g. by means of a meta-analysis, than to investigate the uncertainty associated with all variables in order to reduce the total uncertainty of the model.

Other sources of uncertainty: In addition to this statistical uncertainty other types of uncertainty in modelling studies are identified (data sources, confounding variables, epidemiological data) and strategies are presented how to deal with them appropriately in the model. Finally a reporting format is presented, which addresses all types of uncertainty in a model and consequently increases the transparency of the model. Summarising this thesis shows that the moderate acceptability of modelling studies seems justified taken into consideration the various types of uncertainty associated with modelling studies, while the current methodologies only handle part of the total uncertainty being the uncertainty associated with the range of the input variables.

## Op zoek naar grotere betrouwbaarheid in gezondheidseconomische modellen: de vermindering van onzekerheid inherent aan modelmatige studies

Overheden zien zich steeds meer gedwongen de groei van kosten in de gezondheidszorg te bestrijden ondermeer door budgettering en prijsregulatie. Deze traditionele methodes van kostenbeheersing waren echter slechts gedeeltelijk succesvol, vooral vanwege het ontbreken van voldoende prikkels. Daarom zijn in verschillende Europese landen van overheidswege initiatieven op gang gekomen die moeten leiden tot een doelmatigere gezondheidszorg. Daarbij richten de maatregelen voor kostenbeheersing zich met name op de farmaceutische sector.

Hoewel er grote verschillen zijn tussen de landen onderling, zijn er drie met elkaar samenhangende ontwikkelingen te onderscheiden: decentralisatie van het besluitvormingsproces in de gezondheidszorg, het opleggen van beperkingen aan het voorschrijven van geneesmiddelen en de eis extra gegevens in te dienen bij de aanvraag voor vergoeding van een nieuw geneesmiddel. De vergoeding van geneesmiddelen is tot voor kort voornamelijk gebaseerd geweest op informatie die essentieel is voor de registratie van geneesmiddelen en toelating tot de markt: werkzaamheid, veiligheid en kwaliteit. Deze uitkomsten staan bekend als "de drie hordes". De toenemende druk op de fabrikanten om ook de doelmatigheid van hun produkten aan te tonen, wordt dan ook als een extra en dus de "vierde horde" beschouwd, die genomen moet worden, voordat het geneesmiddel vergoed kan worden. Terwijl werkzaamheid, veiligheid en kwaliteit afkomstig zijn uit gerandomiseerde klinische studies (*randomized clinical trials*) met een hoge interne validiteit, is bij doelmatigheidsstudies vooral de externe validiteit van belang: Een gezondheidseconomische studie moet inzicht geven in de doelmatigheid van een nieuw geneesmiddel op basis van het te verwachten gebruik in de dagelijkse praktijk. Naast de vraag naar gegevens over doelmatigheid van een nieuw geneesmiddel, is er ook en belangrijk rol voor financiële analyses die in kaart betrekking welk beslag een nieuwe geneesmiddel zal leggen op het gezondheidszorgbudget.

Er zijn verschillende methodes om gegevens voor gezondheidseconomische studies te verzamelen. Bij de z.g *piggy-back trial* worden gegevens over medische consumptie verzameld tijdens de fase III klinische studie, die echter primair voor registratie is opgezet. Een nadeel van deze methode is een gebrek aan externe validiteit omdat klinische studies strikte inclusie- en exclusie criteria hanteren en omdat behandeling verloopt volgens een protocol. Een alternatief is de naturalistische economische studie. Dit is, net als een klinische phase III trial, een gerandomiseerde studie, maar één die primiar wordt opgzet om de doelmatigheid van een geneesmiddel te bepalen door middel van het verzamelen van klinische en economische gegevens, die representatief zijn voor de dagelijke behandelpraktijk.

Helaas is de bruikbaarheid van prospectieve naturalistische studies beperkt, omdat het verzoek tot vergoeding in principe direct na registratie wordt ingediend, waardoor er dus geen gegevens omtrent doelmatigheid voor handen zijn. Modelmatige studies vormen een alternatief zonder de praktische beperkingen van prospectieve studies, in het bijzonder in het geval van chronische ziektes zoals de

ziekte van Parkinson, waarbij extrapolatie van de effecten van geneesmiddelen naar de lange termijn nodig is. In verschillende landen zijn nationale richtlijnen opgesteld om de wetenschappelijke kwaliteit en betrouwbaarheid van farmacoeconomische studies te bevorderen. Hoewel verschillende groepen aanbevelingen hebben gepubliceerd voor het volgens de regels der kunst uitvoeren van economische evaluaties, hebben zulke richtlijnen vooral betrekking op prospectieve studies en hebben slechts een gering aantal aanbevelingen betrekking op het uitvoeren van modelmatige studies. Wellicht ligt daarin een verklaring voor de lagere acceptatie van modelmatige studies vergeleken met prospectieve studies, wat verder wellicht samenhangt met de grote mate van onzekerheid in modelmatige studies. Omdat het in veel gevallen onvermijdelijk is om modelmatige studies te gebruiken voor het bepalen van de doelmatigheid van een nieuw geneesmiddel, zal de vergoeding van geneesmiddelen in sterke mate gebaseerd worden op uitkomsten van modelmatige studies. Daarom is het van groot belang om de modelmatige benadering nauwkeurig te onderzoeken en te verfijnen; bovenal dient duidelijk te worden hoe om te gaan met onzekerheid in modelmatige studies.
Het doel van dit proefschrift was daarom om verschillende soorten onzekerheid die zich bij modelleren voordoen te identificeren en nader te onderzoeken. Methodes worden gepresenteerd die ontworpen zijn om de onzekerheid bij modelmatig studies te verminderen, en zo hun betrouwbaarheid te vergroten. Zo kan een bijdrage worden geleverd aan een grotere acceptatie van modelleren bij gezondheidseconomische evaluaties, ook al ontbreken er vooralsnog formele richtlijnen. De ontwikkelde concepten worden empirisch geïllustreerd aan de hand van Markov modellen voor chronische ziekten: depressie, de ziekte van Parkinson, en multiple sclerose. Waar beschikbaar is gebruikt gemaakt van gepubliceerde "echte" data; in andere gevallen is extrapolatie gebruikt om de relevante aspekten te kunnen illustreren.

De volgende soorten onzekerheid werden onderzocht:

- Onzekerheid die samenhangt met de databronnen die dienen als "*input*" van het model. De betrouwbaarheid van de schattingen hangt af van de keuze van de databronnen (selectiecriteria, externe validiteit). Gegevens kunnen afkomstig zijn van een verscheidenheid aan bronnen en zijn onderhevig aan verschillende soorten van onzekerheid. De volgende aspekten van het selecteren van databronnen werden besproken: 1) een classificatie van gegevens die voor modellering worden gebruikt; 2) beoordeling van de verschillende databronnen en hun voordelen en nadelen; 3) een algemene strategie om te beoordelen of een bepaalde databron geschikt is voor het betreffende model; 4) aanbevelingen voor een strategie voor het selecteren van databronnen en voor een transparante en uniforme wijze van rapporteren van databronselectie (verantwoording). Een ernstig bezwaar dat bijzondere aandacht verdient bij modelmatige studies is het gebruik van Delphi panels voor het verkrijgen van data die niet aan bestaande databronnen ontleend kunnen worden. Er bestaat geen garantie dat de schattingen door het panel van medische "consumptie", utiliteiten en patiëntenvoorkeuren de werkelijkheid op accurate wijze weergeven. Daarom wordt

in dit proefschrift (Hoofdstuk 4) een *cross-sectional* studie voorgesteld als alternatieve databron voor een modelmatige studie. Gegevens over medische consumptie en utiliteiten kunnen door een dergelijke studie geleverd worden voor de verschillende Markov gezondheidstoestanden. Een zo ontworpen studie vormt een brug tussen enerzijds een naturalistische prospectieve studie en een zuivere modelmatige studie, zodanig dat de voordelen van beide types studie worden gecombineerd en de nadelen geminimaliseerd.

- Verminderen van onzekerheid met behulp van een sensitiviteitsanalyse (gevoeligheidsanalyse). Sensitiviteitsanalyses richten zich over het algemeen op slechts een beperkt aantal variabelen, zonder dat de keuze van variabelen of de intervallen waarover ze worden gevarieerd verantwoord worden. Dit kan leiden tot een grote mate van subjectiviteit bij het uitvoeren van een sensitiviteitsanalyse. In dit proefschrift (Hoofdstuk 8) wordt een objectieve methode gepresenteerd die gebaseerd is op integratie van punt-sensitiviteit (*point-sensitivity*) en interval-sensitiviteit (*range-sensitivity*) tot een "totale sensitiviteit" (*overall sensitivity*). Deze benadering vermijdt de subjectieve keuze van variabelen bij senstitiviteitsanalyse en de mogelijke vertekening (*bias*) bij het inschatten van de "gevoeligheid" van variabelen die daarvan in de meeste huidige economische studies het gevolg kan zijn.
- Onzekerheid die het gevolg is van samenhang tussen variabelen (*inter-variable uncertainty*). Een belangrijke beperking van sensitiviteitsanalyses zoals die standaard worden uitgevoerd, is dat alleen het effect op de uitkomst wordt onderzocht wanneer de waarde van een variabele wordt veranderd zonder dat daarbij rekening wordt gehouden met de mogelijke samenhang tussen variabelen. In hoofdstuk 10 van dit proefschrift wordt een methode gepresenteerd die de onzekerheid die voortvloeit uit de onderlinge samenhang van variabelen onderzoekt. Dit verschijnsel wordt gedefinieerd als $2^{dc}$-orde sensitiviteit. De praktische gevolgen worden getoond wanneer bij sensitiviteitsanalsyses met dit type onzekerheid rekening wordt gehouden. Resultaten verkregen met deze $2^{dc}$-orde sensitiviteitsanalsyses laten zien dat de meest "gevoelige" variabelen in een model ook de grootste invloed hebben op de gevoeligheid van het model voor andere variabelen. Daarom zal vermindering van de onzekerheid in de meest gevoelige variabele de totale onzekerheid in een model via twee wegen terugdringen: zowel direct, als indirect door het effect op andere variabelen. Anderzijds kan met behulp van een $2^{dc}$-orde sensitiviteitsanalyse de sensitiviteit van de meest gevoelige variabele verminderd worden door de variabele(n) te identificeren, die de grootste invloed hebben op de sensitiviteit van de meest gevoelige variabele.
- Onzekerheid ten gevolge van *confounding* variabelen. Kosten en utiliteiten hoeven niet alleen een functie te zijn van de in een Markov model gedefinieerde gezondheidstoestanden, maar ze kunnnen ook beïnvloed worden door andere variabelen die kunnen fungeren als *confounding* variabelen wanneer er geen rekening mee wordt gehouden. In Hoofdstuk 7 wordt een strategie gepresenteerd die het mogelijk maakt zulke *confounding* variabelen "in te bouwen" in een Markov model door het bepalen van een relatie tussen de *confounding* variable en kosten voor elke

gezondheidstoestand, waarbij ook rekening gehouden wordt met de tijdsafhankelijkheid van een *confounding* variable. Ook werd de gevoeligheid van de uitkomsten van het model voor het inbouwen van een *confounding* variabele bepaald. De resultaten laten zien dat de uitkomsten van gezondheidseconomische modellen sterk vertekend kunnen worden wanneer geen rekening wordt gehouden met *confounding* variabelen. Dit bewijst het belang van het incorporeren van *confounding* variabelen in gezondheidseconomische modellen.

- Onzekerheid die samenhangt met epidemiologische data. In Hoofstuk 11 wordt een methode gepresenteerd die het mogelijk maakt om enerzijds een juiste schatting te maken van de invloed van de introductie van een nieuw geneesmiddel op het geneesmiddelenbudget (*budgetary impact*), terwijl anderzijds tegelijkertijd een traditionele doelmatigheidsanalyse kan worden uitgevoerd. Een Markov model werd geconstrueerd om de epidemiologische data te valideren door de consistentie tussen prevalentie en incidentie van de ziekte van Parkinson te testen. De analyse brengt een aanzienlijke discrepantie aan het licht tussen de prevalentie zoals die gerapporteerd wordt in de literatuur en de prevalentie die volgt uit het model en die berekend is op grond van de incidentie uit de literatuur. Hieruit kan geconcludeerd worden dat een populatie-model een extra bron van onzekerheid herbergt samenhangend met epidemiologische data in vergelijking met een doelmatigheidsmodel dat uitgaat van de individuele patiënt.
- Onzekerheid die samenhangt met de wijze van rapporteren van modelmatige studies. In tegenstelling tot gegevens uit klinische studies, bestaan er geen algemeen geaccepteerde standaarden voor het verzamelen en analyseren van gegevens voor modelmatige studies. Er dient daarom gepleit te worden voor een meer gedetailleerde en gespecificeerde (gedisaggregeeerde) wijze van rapporteren van modelmatige studies, waarbij expliciet wordt ingegaan op de verschillende types van onzekerheid die inherent zijn aan modelmatige studies: de selectie van variabelen, het ontwerp van het model en aannames die er aan ten grondslag liggen, het hanteren van *confounding* variabelen en het uitvoeren van senstitiviteitsanalyses en het interpreteren van de resultaten daarvan.

De conclusie van dit proefschrift is dat er verschillende types onzekerheid onderscheiden kunnen worden in modelmatige studies. Tot op heden bestaat de gebruikelijke benadering van het omgaan met onzekerheid uit het uitvoeren van sensitiviteitsanalyses die alleen betrekking hebben op de onzekerheid die samenhangt met het interval waarbinnen de waarde van een *input* variabele kan liggen (statistische onzekerheid). Dit proefschrift gaat dieper in op het uitvoeren van sensitiviteitsanalyses en identificeert andere bronnen van onzekerheid. Er wordt een objectieve methode gepresenteerd voor het uitvoeren van sensiviteitsanalyses, die de mate van mogelijke subjectiviteit minimaliseert.

De onzekerheid die samenhangt met sensitiviteitsanalyse: Over het algemeen worden in sensitiviteitsanalyses slechts enkele variabelen onderzocht zonder dat de keuze van variabelen of het interval waarover de variabelen worden gevarieerd verantwoord worden. Dit kan leiden tot

subjectiviteit bij het uitvoeren van sensitiviteitsanalyses. De objectieve methode die in dit proefschrift werd gepresenteerd integreert punt-sensitiviteit en interval-sensitiviteit en maakt het zo mogelijk een oordeel te vormen over de "totale" sensitiviteit. Voorts werd onzekerheid die voortkomt uit de samenhang tussen variabelen onderzocht. De resultaten van de daartoe ontwikkelde $2^{de}$-orde sensitiviteitsanalyse laten zien dat de meest "gevoelige" variabelen in het model, die op deze wijze geïdentificeerd kunnen worden, tevens de belangrijkste factor zijn die de gevoeligheid van de overige variabelen bepalen. Dit maakt het mogelijk om gericht vooral aandacht te besteden aan het terugbrengen van de onzekerheid in die variabelen, bijvoorbeeld door het uitvoeren van meta-analyses; op die manier kan de totale onzekerheid in het model sterker en op meer efficiente wijze worden verminderd dan door uitputtende analyses op elk van de overige variabelen afzondelijk.

Overige bronnen van onzekerheid: Naast statistische onzekerheid werden andere soorten van onzekerheid in modelmatige studies geïdentificeerd (databronnen, *confounding* variabelen, epidemiologische gegevens), en er werden methodes en strategieën gepresenteerd om op de juiste wijze hiermee om te gaan in een model. Tenslotte werd een gestandaardiseerde wijze van rapporteren voorgesteld die rekening houdt met alle types van onzekerheid in een model, waardoor de transparantie vergoot wordt.

Samenvattend, laat dit proefschrift zien dat de matige acceptatie van modelmatige studies gerechtvaardigt lijkt, wanneer rekening word gehouden met de verschillende soorten onzekerheid inherent aan modelmatige studies en het feit dat in de huidige praktijk slechts rekening wordt gehouden met een deel van de totale onzekerheid en wel de statistische onzekerheid samenhangend met het interval van een variabele.

## References


1 Clemens K, Townsend R, Luscombe F, Mauskopf J, Osterhaus P, Bobula J. Methodological and conduct principles for pharmacoeconomic research. Pharamcoeconomics 1995; 8 (2): 169-74

2 Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996; 313: 275-83

3 Jefferson T, Demichelli V. Are guidelines for peer-reviewing economic evaluation necessary? A survey of current editorial practice. Health Econ 1995; 4: 383-8

## ACKNOWLEDGMENTS

In summer 1999 Professor dr. Frans Rutten and I met during a health economic training course for pharmacists. During the dinner the idea for the current thesis emerged. I submitted a study proposal to him, which was the basis for the current thesis started in 2000. We based the thesis on a number of published articles by me dealing with uncertainty and titled it "In search for more confidence in health economic modelling: reducing uncertainty associated with modelling studies". Subsequently we explored in more detail various topics related to uncertainty in modelling studies, which resulted in some joint published or in press manuscripts.

Frans, I want to express my gratitude to you for giving me the opportunity to write this thesis in my sparse part-time in addition to my daily work for MEDTAP International. Although months of silence passed due to other commitments, we succeeded in successfully finalising this project through an efficient collaboration. I look back with pleasure to our progress meetings over the last years.

I also enjoyed the collaboration with the other researchers at iMTA, and especially Dr. Maiwenn Al. Early 2002 Dr. Maiwen Al joined this PhD-project as co-promoter. She played an important role in the finalisation of the thesis. Besides her methodological contribution to the last manuscript, she also contributed to the general sections (foreword, discussion and summary), for which I am very grateful to you, Maiwenn.

The older initial papers, which were the starting point for this thesis originated during collaboration with Martin Hardens at The Benefit Research Group. In retrospect the initial seeds for this thesis were sown during this period, when we wanted to assess a priori the possible outcomes of a prospective naturalistic study for which we build our first Markov model, which later was published. I am grateful for your role as "sparring partner" in those early days.

I want to express my gratitude to Prof. Hunink, Prof. Van Hout and Dr. Koopmanschap for their critical reading of the manuscript. I am also grateful to them and Prof. Donker, Prof. Bonsel and Prof. Leufkens for their willingness to challenge me today on my thesis. Also I would like to thank all co-authors on my publications for their reviews of my initial draft manuscripts and their helpful suggestions for improvements.

I would like to thank John Hutton and Ruth Brown of MEDTAP International for their critical review of various sections of this thesis. I also owe many thanks to Bryan Luce, founder of MEDTAP International, who supported me in my personal and scientific development. I also would like to thank

Peter Engelfriet, Jozsef Kosa, and other colleagues and friends for their support and interest during the making of this thesis.

I also would like to thank the Dutch Health Insurance Board and the Ministry of Health, who used articles from this thesis for the development of the Dutch pharmacoeconomic research guidelines. This has been a stimulus for further exploration of a number of methodological topics.

I am grateful to my parents, who gave me all opportunities for my personal development, which has been the basis of this thesis.

Last but not least, I thank Anna and Charlotte for their patience and mental support over the last years.

**Mark J.C. Nuijten**

studied medicine at the University of Amsterdam and was trained as a physician at the Erasmus University in Rotterdam (1990). He worked in clinical research before obtaining his international MBA from the Erasmus University in Rotterdam (1991-1993). He founded the Continental MEDTAP International office in Amsterdam in 1999 and is Director of Business Development for Europe. Prior to joining MEDTAP he was Managing Director of Benefit BV consisting of Benefit Research Group, Quintiles and The Lewin Group. Since 2001 he has been a member of the Editorial Advisory Board of Value in Health. In 1998 he has contributed to the development of the Dutch pharmacoeconomic guidelines (1999) and the Dutch health economic costing manual (1999). He is also a co-founder of Prosper, a Dutch task force promoting the use of health economics in health care decision-making. Since June 2002, he is Board member of ISPOR. His research focuses on methods of health economic modelling and broadening their application in health care, especially the use in the pricing and reimbursement process.