PRINCIPLES AND PROGRESS IN HEALTHCARE COST ANALYSIS

APPLICATIONS TO ECONOMIC EVALUATIONS IN COPD

UITGANGSPUNTEN EN VERBETERINGEN VOOR KOSTENANALYSES IN DE GEZONDHEIDSZORG

TOEPASSINGEN VAN ECONOMISCHE EVALUATIES OP HET GEBIED VAN COPD

Proefschrift ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus Prof.dr. S.W.J. Lamberts en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op donderdag 9 december 2004 om 16.00 uur door Jan Berend Oostenbrink, geboren te Eelde.

PROMOTIECOMMISSIE

Promotor
Prof.dr. F.F.H. Rutten

Overige leden
Prof.dr. E.K.A. van Doorslaer
Prof.dr. M.G.M. Hunink
A.H. Briggs, DPhil.

Copromotor
Dr. M.P.M.H. Rutten-van Mölken

This thesis is based on the following publications:

CHAPTER 2

Oostenbrink JB, Koopmanschap MA, Rutten FFH. Standardisation of costs: the Dutch manual for costing in economic evaluations [1]. Pharmacoeconomics 2002;20(7):443-54.

CHAPTER 3

Oostenbrink JB, Buijs-Van Der Woude T, Van Agthoven M, Koopmanschap, MA, Rutten FFH. Unit costs of inpatient hospital days [1]. Pharmacoeconomics 2003;21(4):263-71.

CHAPTER 4

Oostenbrink JB, Rutten-van Mölken MPMH, Al MJ, van Noord JA, Vincken W. One-year cost-effectiveness of tiotropium versus ipratropium to treat chronic obstructive pulmonary disease[2]. European Respiratory Journal 2004;23:241-9.

CHAPTER 5

Oostenbrink JB, Rutten-van Mölken MPMH, Monz BU, FitzGerald JM. Probabilistic Markov model to assess the cost-effectiveness of bronchodilator therapy in COPD patients in different countries [3]. Value in Health; In press.

CHAPTER 6

Oostenbrink JB, Rutten-van Mölken MPMH. Resource Use and Risk Factors in High-Cost Exacerbations of COPD[4]. Respiratory Medicine 2004; 98: 883-91.

CHAPTER 7

Oostenbrink JB, Al MJ, Rutten-van Mölken MPMH. Methods to analyse cost data of patients who withdraw in a clinical trial setting[1]. Pharmacoeconomics 2003;21(15):1103-12.

CHAPTER 8

Oostenbrink JB, Al MJ. Cost analysis in economic evaluations with incomplete data due to dropout[5]. Health Economics; In press.

Reprinted with kind permission of:
1 Adis International
2 European Respiratory Society Journals Ltd
3 Elsevier
4 International Society for Pharmacoeconomics and Outcomes Research
5 John Wiley & Sons, Ltd

The publication in chapter 2 received the 2003 Young Researchers Award of the Dutch Society for Technology Assessment in Health Care. Preliminary versions of the publications in chapter 4 and 7 received Best Contributed Paper Awards at the 2001 ISPOR conference in Cannes for poster and oral presentation respectively.

CONTENTS

CHAPTER 1 Introduction *12*
1.1 Background *13*
1.2 Measurement and valuation of costs *14*
1.3 Event-driven costs *15*
1.4 Incomplete cost data due to premature withdrawal *16*
1.5 Aim of the thesis *17*
1.6 The case of COPD *17*
1.7 Outline of this thesis *19*
CHAPTER 2 Standardisation of costs: the Dutch manual for costing in economic evaluations *20*
2.1 Introduction *21*
2.2 The Dutch manual for costing *21*
2.3 Key issues regarding the standardisation of costs *26*
2.4 Discussion *29*
Appendix 2.1 Example of the use of standard values *31*
CHAPTER 3 Unit costs of inpatient hospital days *32*
3.1 Introduction *33*
3.2 Methods *34*
3.3 Results *36*
3.4 Discussion *40*
3.5 Conclusions *43*
CHAPTER 4 One-year cost-effectiveness of tiotropium versus ipratropium to treat COPD *44*
4.1 Introduction *45*
4.2 Methods *45*
4.3 Results *49*
4.4 Discussion *55*
4.5 Conclusions *58*
CHAPTER 5 Probabilistic Markov model to assess the cost-effectiveness of bronchodilator therapy in COPD patients in different countries *60*
5.1 Introduction *61*
5.2 Methods *62*
5.3 Results *67*
5.4 Discussion *62*
5.5 Conclusions *75*
Appendix 5.1: The calculation of transition and exacerbation probabilities *76*
CHAPTER 6 Resource use and risk factors in high-cost exacerbations of COPD *80*
6.1 Introduction *81*
6.2 Methods *81*
6.3 Results *83*
6.4 Discussion *87*
6.5 Conclusions *90*

CHAPTER 7 Methods to analyse incomplete cost data of patients who dropout in a clinical trial setting *92*
7.1 Introduction *93*
7.2 Methods *94*
7.3 Results *100*
7.4 Discussion *102*
7.5 Conclusions *103*
CHAPTER 8 The analysis of incomplete cost data due to dropout *104*
8.1 Introduction *105*
8.2 Methods *106*
8.3 Results *110*
8.4 Discussion *113*
8.5 Conclusions *119*
Appendix 8.1: creation of samples *121*
Appendix 8.2: creation of DAR and ID in the complete samples *124*
CHAPTER 9 General discussion and conclusions *124*
9.1 Introduction *125*
9.2 Economic evaluation of tiotropium *125*
9.3 Standardisation and comparability of costs in economic evaluations *127*
9.4 Generalisability of economic outcomes across countries *129*
9.5 Analysis of event-driven costs *130*
9.6 Analysis of incomplete cost data due to premature withdrawal *132*
9.7 Applicability of results to other diseases *135*

References *136*
Abbreviations *144*
Summary / Samenvatting *146*
Dankwoord *158*

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Due to medicalisation, ageing of the population, and technological and pharmaceutical developments, Western countries have been confronted with a rapid increase in the costs of healthcare during the last decades. The armamentarium of the medical profession has grown enormously and medications have become available for diseases for which, until recently, treatment was not possible. These developments coincided with increasing pressure on budgets of national governments and the awareness that limits must be set to the growth of the costs of healthcare. Instead of the automatic influx of new technologies, the need arose to assess these technologies in terms of their costs and benefits in order to decide upon registration, reimbursement and pricing (Boer, 2002). These developments have led to a significant increase in the number and variety of economic evaluations in healthcare. Economic evaluations have been performed for many different kinds of health technologies, including organ transplantation, diagnostic devices and treatment with medicines. Economic evaluations were either performed alongside prospective randomised controlled trials, as standalone studies based on retrospective data, or as modelling studies incorporating economic and clinical data from a variety of sources. With the increase of studies, several authors have expressed their worries about the quality and comparability of these economic evaluations. The incomparability may not necessarily be a problem as long as differences in outcomes reflect real differences with regard to the setting, aim or perspective in which the study is performed. However, it is clear *'that some of the observed differences in cost-effectiveness have more to do with study methodology than with the performance of the therapies being evaluated'* (Drummond, 1994). Badia et al. reviewed six economic evaluations of hepatitis B vaccination programs in Spain and found numerous discrepancies between studies that were not related to their aim. They concluded that *'this kind of heterogeneity ought to be minimised, otherwise evaluations of the same problem in the same setting could produce different results, undermining their impact on the decision-maker and even the credibility of the evaluations'* (Badia et al., 1997). Based on a review of 45 economic evaluations conducted alongside clinical trials, Barber and Thompson concluded that *'there is an urgent need to improve the statistical analysis and interpretation of cost data'* (Barber and Thompson, 1998). Graves et al. examined the quality of the cost methods of the same 45 economic evaluations and found these methods to be of poor quality, stressing the need for greater rigour by stating that *'no amount of statistical analysis can compensate for inadequate costing methods'* (Graves et al., 2002). Jacobs and Bachynsky reviewed 48 economic evaluations performed in Canada and concluded that biases occurred in most studies and most cost categories (Jacobs and Bachynsky, 1996).

To meet the shortcomings of economic evaluations, many authors have argued for more standardisation of the methodology of economic evaluations. Drummond et al. have identified three motivations for standardisation in order to: 1) maintain the scientific quality of studies; 2) facilitate the comparison of results of economic evaluations for different healthcare interventions; and 3) assist in the interpretation of results from setting to setting (Drummond et al., 1993). These calls for more standardisation have resulted in numerous methodological papers, standard textbooks, good practice recommendations and checklists for researchers

and reviewers for conducting and assessing studies in health technology assessment (HTA). Methodology has been proposed for many aspects of economic evaluations including the measurement and valuation of health outcomes and costs, the statistical analysis of uncertainty about costs, effects and the cost-effectiveness ratio, and the use of modelling studies.

The call for more standardisation has also been accommodated by the development of national guidelines for (pharmaco) economic evaluations in many Western countries (see for instance: http://www.ispor.org/PEguidelines/index.asp). Hjelmgren et al. classified these guidelines into three broad groups: 1) formalised guidelines as a requirement prior to reimbursement; 2) informal guidelines as a recommendation prior to reimbursement; and 3) guidelines for health economic methods that are intended for use in discussing and improving methodology in health economic evaluations (Hjelmgren et al., 2001). All these guidelines mainly consist of a set of recommendations about the basic principles of economic evaluations. These, for instance, involve the perspective from which the study should be performed, the choice of the outcome measure that should be used, the discounting of costs and effects, and whether to include costs outside the healthcare sector. Nevertheless, for many aspects guidelines provide only global guidance and many areas still lack consensus about methodological standards (Johnston et al., 1999).

1.2 MEASUREMENT AND VALUATION OF COSTS

The measurement and valuation of costs are essential components of every economic evaluation. In empirical studies, costs are usually determined by measuring the use of healthcare resources of individual patients, and multiplying their resource use with estimates of prices of these resources. The measurement and valuation of resource use have been shown to vary widely across studies. Examples are legionary. Some studies have restricted the collection of resource use data to costs within the hospital, whereas other studies have included all inpatient and outpatient costs. Many studies differ with regard to the inclusion of patients' costs and of costs outside the healthcare sector. The valuation of resource use may have been based on list charges used to reimburse hospitals, or on detailed calculations of unit costs. Often, the observed differences cannot be related to the aim of the study. Based on his review of national guidelines, Hjelmgren and colleagues concluded that *'fairly good agreement'* exists about the core aspects of guidelines and that inconsistencies mainly relate to *'the perspective of the analysis and the measurement and valuation of costs. (...) Especially the valuation of healthcare resources appears to be a difficult problem'* (Hjelmgren et al., 2001).

There seems to be wide agreement about the desirability to standardise the measurement and valuation of resource use in order to improve the comparability and generalisability of economic evaluations. Including different resource use items or using different prices for the valuation constitute a direct obstacle for the comparability of economic evaluations. Health outcomes have been standardised by the development of quality of life instruments that can be used in different studies of patients with the same disease. The adoption of generic quality of life instruments and quality adjusted life years (QALYs) as outcome measures may even facilitate the comparison of studies across settings and diagnoses. Such comparisons are only useful if costs can be standardised to the same extent. In order to improve the comparability

of economic evaluations, some health economists have wondered whether general or national guidelines *'should go deeper in their recommendations, or whether a manual of operational procedures should be written that contains all the specific details that must be taken into consideration'* (Rovira and Antonanzas, 1995).

The generalisability of economic evaluations relates to the question to what extent the findings of an economic evaluation apply to another setting. Although the question of generalisability also applies to studies that have been performed in the same country, the issue becomes even more hazardous when a study is performed in another country. Whereas it may often be assumed that the biological effect of a treatment is more or less the same for patients from different countries, this clearly is not the case for resource use and costs. Many factors that are pertinent to resource use and costs are known to vary across countries. These include demography and epidemiology of the disease, the availability of healthcare resources, variation in clinical practice, incentives to healthcare professionals and institutions, and differences in absolute and relative price levels (Drummond, 1994; Drummond et al., 1992). Hence, *'standardisation of costs may be much more difficult than standardising clinical measures, especially if the participating centres are from different states or countries'* (Rizzo and Powe, 1999) and, clearly, resource use and costs may not be generalised to other countries without appropriate adjustment.

■ 1.3 EVENT-DRIVEN COSTS

The nature of resource use and cost data in economic evaluations has led health economists to the conclusion that *'the outcomes with which they are concerned are awkward to analyse empirically'* (Manning and Mullahy, 2001). Day-to-day costs of treatments for many chronic diseases, for instance, may consist of regular visits to physicians and other healthcare providers, diagnostic and prognostic tests to identify and monitor disease progression, and of treatment with maintenance medication. Costs of this day-to-day treatment may occur in the majority of patients with the same diagnosis and gradually increase with the severity of the underlying disease, with only a modest variation between patients. The picture changes, however, when the disease is not adequately controlled and a patient suffers a medical complication. The occurrence of such medical events may lead to changes in therapy, increased numbers of physician visits and acute admissions to hospital. The cost of this event-related resource use is often disproportionately high when compared to the cost of day-to-day treatment. The typical situation of *'a few patients incurring rare but highly expensive costs and many patients having few or no costs'* (Briggs et al., 2002) causes the distribution of costs to be severely skewed, with important implications for the calculation and analysis of cost data.

The analysis of skewed cost data has already received much attention. An important feature of skewed data is the inappropriateness of common analytic methods to deal with such data. Most methods require a normal distribution to produce valid inferences and only sufficiently large sample sizes may relax this criterion. The use of non-parametric rank tests should be avoided as they compare the distribution of data in terms of shape and location, rather than the mean (Thompson and Barber, 2000). Transformation of cost data has been discouraged because of problems related to the interpretation of data analysed on a trans-

formed scale (Briggs and Gray, 1998; Thompson and Barber, 2000). Currently, bootstrapping seems to be the method of choice for the analysis of cost data (Briggs et al., 1997). However, it has recently been shown that with small sample size, the bootstrap may also be non-robust to deviations from the normality assumption (O'Hagan and Stevens, 2003).

Another implication for the analysis of costs is the high variance by which skewed distributions are characterised. As many empirical economic evaluations are performed alongside randomised controlled trials powered on some clinical outcome measure, the sample size is often not sufficient to detect differences between treatments that are 'statistically significant' at conventional levels. A review of studies reporting patient-level cost data revealed that *'the majority of cost studies identified in this review are grossly under-powered to detect any but extremely large differences in cost'* (Briggs and Gray, 1998). It has also been shown that sample size calculations based on cost-effectiveness outcomes require information about many different parameters of which many may be associated with high uncertainty (Al et al., 1998). The high variance in costs and cost-effectiveness outcomes is one of the driving forces behind the concept that an adequate description of uncertainty is more relevant for decision-making than classical hypothesis testing based on some arbitrary threshold of statistical significance.

Considering the contribution event-related costs may make to total treatment costs, careful consideration should be given to the definition of events and the measurement of event-related resource use. The lack of a uniform definition of events may hinder the comparability of clinical outcomes across studies. Estimates of event-related resource use may not apply to other settings because of the use of different definitions. This incomparability of results from different settings may also affect model-based economic evaluations. Data to populate a model are often derived from multiple sources, and uniform definitions of events are a necessary condition to combine data about event rates and event-related resource use from these different settings.

1.4 INCOMPLETE COST DATA DUE TO PREMATURE WITHDRAWAL

Many economic evaluations of chronic diseases are based on empirical data collected during longitudinal studies. Designed as a stand-alone study or nested within a randomised controlled clinical trial, the aim is to follow patients over a longer period of time. For each patient, data on health outcomes and resource use need to be collected for the entire observation period. In these studies, the occurrence of incomplete data due to premature withdrawal (dropout) of a patient may constitute a serious problem. If for instance resource use is measured over 1 year, and a patient decides to withdraw from the study after 9 months, data for the last 3 months may be lacking. Patients who withdraw from the study are often more severely ill than patients who do not withdraw. Hence, their costs per time interval are generally higher than the costs of the non-withdrawals and, in case of a direct comparison of different treatments or a treatment with placebo, the withdrawal rate may differ considerably between groups. Johnston et al. identified the *'investigation of methods for handling missing and censored data (...) as one of the main methodological issues when there is a lack of consensus'* (Johnston et al., 1999). Surprisingly, little attention has been paid to the problem

of incomplete cost data in economic evaluations. In the review of Barber and Thompson of 45 economic evaluations related to randomised controlled trials, 21 (47%) studies did not provide any information about the completeness of the data, while three papers reported to have no missing data. Of the remaining 21 papers reporting to have missing data, 11 excluded all patients with missing cost data, apparently without any further investigation (Barber and Thompson, 1998).

The limited attention that has been paid to the analysis of incomplete data within economic evaluations strongly contrasts with other areas. Many different methods for the analysis of incomplete data are available, varying from simple methods like case deletion or mean imputation to more advanced methods like generalised linear mixed models and multiple imputation (Little and Rubin, 1987). All these methods differ with respect to how they compensate for the missing data. They also vary largely with regard to the distribution of data and the type of dropout they can handle. Up to now it is largely unknown what the impact is of incomplete data on the analysis of costs in economic evaluations and how available methods for dealing with missing data perform when applied to cost data.

1.5 AIM OF THE THESIS

This thesis deals with the analysis of costs and cost-effectiveness in economic evaluations, with applications in chronic obstructive pulmonary disease (COPD). COPD is a chronic, slowly progressing, respiratory disorder where the problems of skewed cost distributions and dropout that is not completely at random have challenged us for quite some time (Rutten-van Mölken et al., 1994). The economic evaluation of tiotropium in patients with COPD will be used as a critical case to illustrate these problems, and to show principles and progress with regard to the analysis of costs. In addition to the specific aim of each publication, this thesis aims to:

Contribute to the comparability and generalisability of economic evaluations by standardising and improving methods for the calculation and analysis of costs.

Special attention will be paid to the following research areas:
- the standardisation and comparability of costs in economic evaluations;
- the generalisability of the results of economic evaluations to other countries;
- the impact of expensive medical events on the analysis of costs;
- the analysis of incomplete cost data due to dropout.

1.6 THE CASE OF COPD

COPD is a disease of the respiratory system characterized by slowly progressive airflow limitation that is not fully reversible (Pauwels et al., 2003; Siafakas et al., 1995). Characteristic symptoms of COPD include cough, sputum production and dyspnoea upon exertion (Pauwels et al., 2003). By 2020, COPD is estimated to become the fifth most common cause of disability and the third most frequent cause of death in the world (Murray and Lopez, 1997). Major risk factors of COPD include tobacco smoke, occupational dusts and chemicals, and indoor and

outdoor pollution. Cigarette smokers have a higher prevalence of respiratory symptoms, a greater annual rate of decline in lung function and a greater COPD mortality rate than non-smokers (Pauwels et al., 2003). It has been estimated that 80% to 90% of all COPD patients have a history of smoking (U.S. Department of Health and Human Services, 1990). The proportion of smokers who develop COPD is not exactly known and estimates have ranged from 15% to 50% (Celli et al., 1995; Lundback et al., 2003). The prevalence of COPD in the Netherlands in 2000 was estimated to be 24 per 1000 in men and 15 per 1000 in women (Hoogendoorn et al., 2003). Due to the ageing of the population and past smoking behaviour, these figures are predicted to increase to 33 in men and 27 in women in 2025 (Hoogendoorn et al., 2003).

The primary aim of COPD management is to prevent disease progression. Since cigarette smoking is the major risk factor for COPD, the first step in the management of COPD among patients who are still smokers is to offer smoking cessation counseling, whether or not in combination with medication or nicotine-replacement therapy. Smoking cessation has been shown to be the most effective way to reduce the risk of developing COPD and to stop its progression (Pauwels et al., 2003). The mainstay of COPD pharmacotherapy is the treatment with bronchodilators directed to relieve symptoms and to prevent exacerbations of the disease. Tiotropium is a new long-acting anticholinergic bronchodilator for once-daily administration in COPD patients. Randomised controlled trials (RCTs) have shown that tiotropium provided sustained bronchodilation, improvements in dyspnoea and health-related quality of life and was associated with fewer exacerbations than placebo and the short-acting anticholinergic ipratropium (Casaburi et al., 2002; Vincken et al., 2002). Improvements in lung function have also been shown to be significantly better than with salmeterol, a beta-agonist for twice-daily administration (Donohue et al., 2002). To support reimbursement and formulary decision-making and guide the positioning of tiotropium in the treatment of COPD we analysed the one-year cost-effectiveness of tiotropium prospectively as part of the RCT program of tiotropium.

The economic evaluation of tiotropium is a good example of the current state of the art of cost-effectiveness analysis, and the analysis of costs in this economic evaluation clearly shows the importance of the research areas addressed above. Firstly, due to its chronic nature and increasing prevalence, the treatment of COPD may lay a heavy burden on the future drug budgets of developed countries. In addition to tiotropium, several new and competing treatments may soon become available (Barnes, 2003). Healthcare authorities will be forced to decide upon reimbursement of these medications and may need high-quality information about their costs and effectiveness. Comparability of studies and generalisability of the results to other settings and countries will be necessary requirements to facilitate these decisions. Secondly, patients with COPD are known to suffer from exacerbations of the disease. Exacerbations are characterised by increased symptoms of sputum, cough and dyspnoea and patients have reported reduced quality of life during exacerbations (Spencer and Jones, 2003). The frequency and severity of exacerbations are related to the underlying severity of the patient's COPD (Andersson et al., 2002; Rodriguez-Roisin, 2000). Treatment of exacerbations is expensive and is a key driver of the costs of unscheduled care. Hospitalisation for a COPD exacerbation may become necessary when initial outpatient exacerbation therapy has

failed, and failure rates of 12-21% have been reported (Dewan et al., 2000; Miravitlles et al., 2002). Exacerbation-related hospitalisations and medications are the major cost drivers, but it depends on the severity of COPD and the country or region which of the two ranks first (Grasso et al., 1998; Hilleman et al., 2000; Jansson et al., 2002; Miravitlles et al., 2003; Ruchlin and Dasbach, 2001; Rutten-van Mölken et al., 1999; Strassels et al., 2001; Sullivan et al., 2000; Wouters, 2003). Total treatment costs are disproportionately distributed and a relatively small proportion of severely ill COPD patients is responsible for a substantial share of total COPD-related healthcare costs. Thirdly, due to the chronic nature of COPD, the efficacy, safety and cost-effectiveness of new treatments need to be investigated in studies that last for a longer period of time. Studies with a time span of up to three years have been conducted in COPD (Burge et al., 2000; Decramer et al., 2001; Scanlon et al., 2000). Similar to the one-year economic evaluation of tiotropium, a considerable number of patients will withdraw from these studies before the scheduled end data and withdrawal rates up to 50% over 3 years have been reported (Burge et al., 2000). The occurrence of dropout that is completely at random is rare and raises the question how to analyse costs in these data sets.

1.7 OUTLINE OF THIS THESIS

Chapters 2 and 3 address the standardisation and comparability of costs. Chapter 2 presents a description of the Dutch manual for costing in economic evaluations and discusses the key issues in relation to the standardisation of costs and how these have been addressed in the manual. Chapter 3 provides an analysis of the unit costs of inpatient days. Costs of inpatient days are often the main driver of costs in economic evaluations and this chapter shows to what extent estimates of unit costs may differ between wards and hospitals. Chapters 4 to 7 are related to the economic evaluation of tiotropium in COPD patients. In chapter 4 the cost-effectiveness of tiotropium is compared to ipratropium based on an economic evaluation conducted alongside two RCTs in the Netherlands and Belgium. Chapter 5 presents a model-based economic evaluation. In this chapter the costs and effects of three bronchodilators are compared using a Markov model with a time span of one year. The model was specifically designed to compare the costs and effects of these treatments in different countries. These two approaches to economic evaluations provide a good example of the current state of the art of the methodology of cost-effectiveness analysis. Chapters 6 and 7 are based on the empirical economic evaluation of tiotropium that was presented in chapter 4. Chapter 6 presents a further analysis of the resource use, costs and risk factors that are associated with COPD exacerbations. Chapter 7 focuses on the analysis of patients with incomplete data due to premature withdrawal. This chapter discusses the impact of premature withdrawal on the analysis of costs, and determines the sensitivity of the outcomes of the economic evaluation to applying different methods for the analysis of incomplete data. The problem of incomplete data due to premature withdrawal is further explored in chapter 8. In this chapter, the performance of various naïve and principled methods for the analysis of incomplete data are compared. These methods are applied to simulated data sets with various distributions of costs and different patterns of dropout. Finally, chapter 9 discusses the findings from the previous chapters in relation to the research aims of this thesis.

CHAPTER 2

STANDARDISATION OF COSTS: THE DUTCH MANUAL FOR COSTING IN ECONOMIC EVALUATIONS

2.1 INTRODUCTION

A major problem in economic evaluations to date is the quality and consistency of studies and the degree to which results can be compared among studies (Haycox and Walley, 1997; Siegel et al., 1997). The lack of a uniform methodology is often considered a weakness of economic evaluations that hinders the use of such assessments in practice. Therefore, several authors have encouraged the standardisation of methods used in economic evaluations (Drummond et al., 1993; Rovira, 1996) for *'promoting high standards of conduct, scientific credibility and for interpreting and comparing the results of studies in similar and different settings'* (Mason, 1997).

In recent years many publications contributed to the distribution of knowledge on (new) methodologies and to the standardised application of such methodologies (Drummond et al., 1997; Gold et al., 1996). These publications also contributed to the standardisation of costing. There are, however, some serious constraints to the degree in which costing within economic evaluations can be standardised. Because studies are performed in different settings, have different aims and differ with respect to the disease and intervention that are investigated, it is not possible to suggest one standardised approach that is applicable to all studies. A balance must be found between the degree of standardisation that can be achieved and the necessity to tailor the approach to a specific study setting.

Standardisation plays an important role in the field of national guidelines. In some countries, these guidelines are voluntary and formulated by leading health-economists. In other countries these guidelines are formulated and issued by governmental agencies and reflect the formal requirements that have to be met for these studies to be considered when deciding on the reimbursement of new medical therapies (Jacobs et al., 1995; Langley, 1996; Rovira and Antonanzas, 1995; Torrance et al., 1996). In general, these guidelines are rather global with respect to costing. In an attempt to further standardise the costing methods used in pharmacoeconomic analyses, Australia and Canada have issued additional guidelines for cost-calculations (Canadian Coordinating Office for Health Technology Assessment, 1996; Commonwealth Department of Health and Ageing, 2002). Recently, the Dutch *'Manual for Costing: Methods and Standard Costs for Economic Evaluations in Healthcare'* (Oostenbrink et al., 2000) (further referred to as 'the manual') has been published. This manual provides guidelines and recommendations for costing in economic evaluations in the Netherlands.

The objectives of this article were to describe the main content of the manual and to discuss in more detail some key issues of the manual in relation to the standardisation of costs. The Dutch manual is also compared with the Australian and Canadian costing guides.

2.2 THE DUTCH MANUAL FOR COSTING

The manual is formulated and issued in addition to the Dutch *'Guidelines for pharmacoeconomic research'* (Riteco et al., 1999). As in other countries, these guidelines define the general standards and major methodologies of pharmacoeconomic evaluation. These guidelines have been issued by the Dutch Health Insurance Board and are approved by the Minister of Public Health, Welfare and Sport. In the future, the standards and methods described in these guidelines have to be applied in studies that support submissions to acquire reimbursement for new pharmaceuticals. For the calculation of costs, the guidelines refer to the manual.

The aim of the manual is to support researchers and users of economic evaluations with the design, performance and assessment of cost-calculations in economic evaluations and to improve the quality and comparability of these studies. Not only does the manual describe the formal standards that have to be met in case of appraisal studies of new pharmaceutical products, but it can also be applied to the economic evaluation of other technologies. It describes cost-concepts, provides many examples, and includes appendices that contain practical information and data that can be applied directly in cost-calculations.

The manual introduces a six-step procedure for costing (*figure 2.1*). During each step, choices have to be made and these together define the approach taken. These choices depend on the aim and the specific setting of the study. As choices in later stages depend on choices made in previous steps, the six steps have to be passed in chronological order. The six steps concern the: 1) scope of the study; 2) choice of cost-categories; 3) identification of units; 4) measurement of resource use; 5) monetary valuation of units; and 6) the calculation of unit costs.

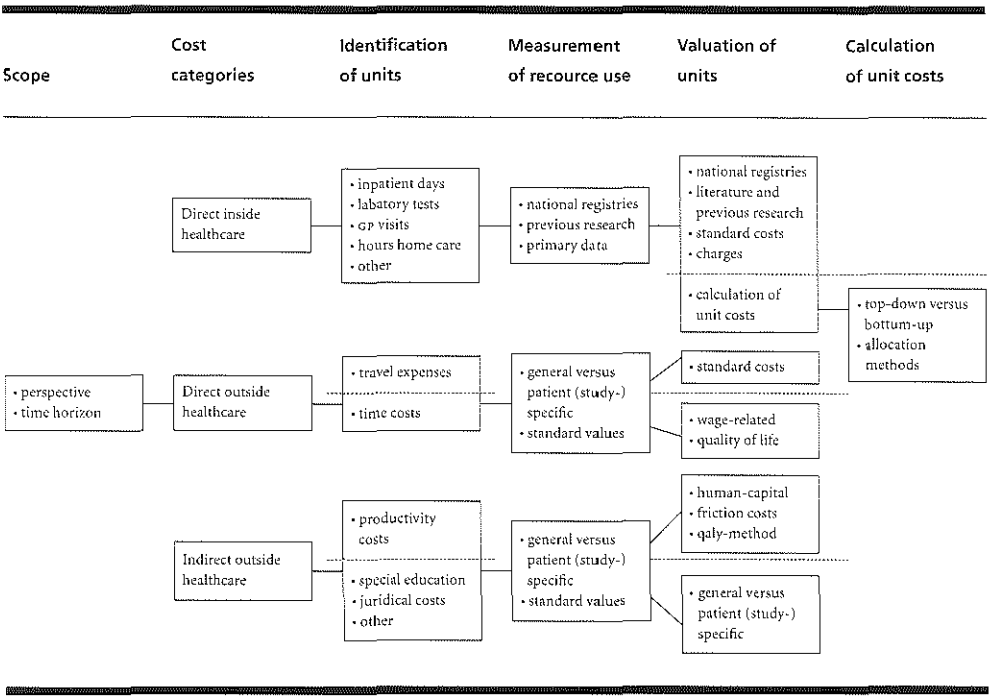


Figure 2.1: six-step procedure for costing.

Step 1: The scope of the study | The scope of the study concerns the choice of perspective and the time span of the study. These two are closely related. A societal perspective, for example, implies a time span allowing inclusion of full consequences of the intervention. If, on the other hand, a study is performed from a healthcare provider's perspective, often a shorter time span will be adopted. The scope of the study affects the entire study design. It is of vital importance for the costing-process, as many of the choices in later stages depend on it. It affects the types of resource use that should be considered and how they are to be measured and valued.

Step 2: The choice of cost categories | In accordance with the Dutch guidelines for pharmacoeconomic research, a distinction is made in direct and indirect costs, and in costs within and outside the healthcare sector. The costs within the healthcare sector include the costs of all healthcare services, irrespective of whether they are borne by the health insurers, government or patients. Examples of direct costs outside the healthcare sector are travel expenses and time costs, which are often borne by patients and their families. Indirect costs contain for example productivity costs, costs of special education, juridical costs, etc.

The choice of which cost categories to include follows directly from step 1. The choice of a societal perspective, for instance, implies that not only the costs within the healthcare sector, but also the direct and indirect costs outside the healthcare sector have to be included in the analysis. If, on the other hand, a study is performed from a health insurers' perspective, this will lead to the exclusion of all costs outside the healthcare sector and of all costs within the healthcare sector that are not reimbursed by the health insurer like (e.g. co-payments).

Step 3: Identification of units | The identification of units raises two questions: what types of resource use are relevant for the disease and the intervention studied; and to what level of detail do they have to be measured and valued separately. If, for example, almost all patients who visit a specialist receive the same diagnostic tests, it is not necessary to cost these tests separately, but it is better to incorporate the average costs of diagnostic tests into the unit price of an outpatient visit. Only when diagnostic tests per visit differ strongly between patients, is it worthwhile to identify these tests as separate units. In the manual we recommend making a description of the treatment process to determine which units should be included in the analysis. We then recommend performing a sensitivity analysis to get insight into the units that have the largest contribution to incremental and total costs. The sensitivity analysis can be used to determine which costs should be measured and valued in detail (micro-costing), and for which units a gross-costing approach is more appropriate.

Step 4: Measuring resource use | There is a wide variety of data sources for determining resource use. These differ strongly with respect to the level of detail in which they provide data. The manual distinguishes four criteria for the selection of data sources in a particular situation. These are the perspective of the study, the contribution of units to total or incremental costs, the availability of data, and the balance between internal and external validity (Jacobs and Baladi, 1996). Internal validity in this case is related to the question of whether the resource use measured reflects actual use in the population being studied, mostly in a clinical trial setting, while external validity is related to generalisability of the results to other settings, especially in daily practice. An increase in the internal validity is frequently offset by a decrease in external validity and vice versa.

Based on these four criteria, a choice has to be made to collect data from either primary or secondary data sources. Secondary data from national registries are often aggregated from several institutions and may therefore have high external validity. This character of secondary data, however, often prevents the use of such data in economic evaluations, especially when a detailed comparison of resource use and costs between treatments is required. The most

well-known type of primary data collection is the RCT. Besides the advantages of randomisation and the controlled situation, a RCT also offers the possibility of integrating data collection into the case report form (CRF). This can contribute importantly to the quality and completeness of the data. A disadvantage of economic evaluations piggy-backed to RCTs concerns the trial induced costs. Because of trial-related resource use, such as scheduled visits to the trial physician, an accurate measurement of the costs that would have occurred in a natural situation is prevented. In many cases, it is unknown in what respect the difference in costs between treatment groups is affected by these trial-induced costs.

Step 5: Valuation of resource use | In the manual we distinguish 5 alternative ways to obtain valid unit prices: 1) prices derived from national registries; 2) prices derived from health economics literature and previous research; 3) standard costs; 4) tariffs or charges; and 5) calculation of unit costs. The choice for a valuation method is often related to the measurement of resource use. For example, it does not make sense to base the resource use measurement on national registries, while much effort is put into detailed unit costing. The criteria for choosing between the valuation methods are similar to the ones described above regarding the measurement of resource use.

The use of standard costs and the calculation of unit costs (step 6) are described in more detail below. The advantage of deriving prices from national registries or previous research is related to the limited efforts needed to collect such data. In the Netherlands, however, the availability of prices from these data sources is limited. Not only is it difficult to obtain prices that are valid for the patient group being studied, but, in many cases, there is also insufficient information about the way these prices have been determined. Consequently, one of the three other valuation methods often has to be used. The attractiveness of the use of charges lies in the extensive list of procedures and services for which charges are available. When the identification of units in a costing study matches with the classification of procedures and services used for the reimbursement, the use of charges for the valuation is relatively simple and less time-consuming than other methods. Especially in valuing hospital services, like laboratory tests, radiographic imaging and surgery, charges are often the most appropriate (and sometimes the only available) method. However, the use of charges is often debated, because they do not always reflect the actual unit cost of a procedure or service, but are merely a vehicle for transferring money from payers to healthcare organisations and providers. Charges should therefore be applied with care, and a researcher should try to get some insight in whether the charge is a good estimate of the actual unit cost of a procedure and can be used in the economic evaluation. Only an analysis from the healthcare insurer's perspective will normally lead to the adoption of charges as the appropriate valuation method.

Step 6: Calculation of unit costs | Because unit cost calculation is much more laborious than other valuation methods, and because it requires a completely different methodology, the calculation of unit costs is distinguished as a separate step in the costing process. It is used for the valuation of units that have a substantial impact on incremental or total costs, and for which no adequate unit cost estimates from other sources are available. Unit costs are calcu-

lated for healthcare services and procedures, such as inpatient days, outpatient visits and surgical interventions, and include the costs of physicians, nursery, medical devices, buildings and equipment, etc. Important choices regarding the calculation of unit costs concern the selection of a specific setting in which unit costs will be calculated, the use of top-down versus bottom-up methods and the allocation of costs of supportive departments, buildings, general equipment, etc.

The selection of a specific setting for the calculation of unit costs is a problematic issue. Several authors have argued that unit costs can differ considerably between healthcare providers, and, consequently, that the choice of centre(s) can seriously affect the cost-calculations (Goeree et al., 1999). We recommend collection of unit costs in more than one centre, and varying unit costs in a sensitivity analysis based on the differences in unit costs that were found or according to estimates obtained from other studies.

In top-down cost calculations the financial administration data of the healthcare provider is the primary source for determining the unit costs per product. Costs of a department are derived from cost-accounting data and assigned to the products or services produced by the department. Top-down calculations can be applied in the case of a department with a relatively homogeneous production. This, for example, is often the case for a nursery ward. Costs of personnel, medical materials, other expenses, and the annual number of inpatient days at that particular ward, can be obtained directly from the central financial and production administration databases to calculate the direct costs per inpatient day. In cases where the production of a department is not homogeneous, the bottom-up method is more appropriate. In bottom-up calculations, unit costs per product or service are determined by measuring actual use of personnel, materials and equipment; for example, by measuring the time a physician spends on a certain procedure for a single patient. Costs are then calculated by valuing the average time measured, using an estimate of the income of the physician. The disadvantage of bottom-up calculations is that they are usually very time-consuming, and a researcher will not always have the opportunity to perform such detailed measurements. In practice, a combination of top-down and bottom-up calculations is often used.

Once direct costs have been calculated, the costs of supportive departments, buildings, equipment and overheads have to be assigned to the healthcare service. For the allocation of these costs several methods are available (Horngren, 1982). The method most often used in economic evaluations is direct allocation, because this method is relatively easy to apply and because accounting systems in hospitals do not often allow the use of more advanced methods. In direct allocation, a distinction is made between departments that directly serve patients (e.g. a ward, a surgery department) and supporting departments (e.g. the kitchen in a hospital, financial department). Costs of supporting departments are not directly assigned to products or patients, but firstly assigned to the departments that directly serve patients. Then, these costs are allocated among the products of these latter departments. For a more extensive description of unit cost calculation we refer to the manual and the relevant literature (Drummond et al., 1997; Horngren, 1982; Oostenbrink et al., 2000).

2.3 KEY ISSUES REGARDING THE STANDARDISATION OF COSTS

Basic principles | Many of the standards in the costing manual are derived from the Dutch pharmacoeconomic guidelines and do not only affect the cost analysis, but are applicable to the entire study design. The most important standards are the use of a societal perspective, a time span allowing inclusion of full consequences of the intervention, the performance of an incremental analysis, and the use of a discount rate of 4%.

A societal perspective takes into account the costs of all stakeholders in society, as opposed to narrower perspectives such as the healthcare sector's or health insurer's perspective. A societal perspective also implies that a time span should be used that includes all costs related to the initial intervention. In many interventions for acute problems, the time span is relatively limited and costs can be determined by direct measurement. In other cases, the time span for costs influenced by the initial intervention exceeds the study duration during which actual resource use is measured, and requires model estimates of costs for the period after the study (sometimes a time span similar to the survival period of the longest living patient).

An incremental analysis implies that, when two or more treatments are compared, outcome measures are expressed as the difference in costs and effects between treatments. The choice for an incremental analysis has important implications for the types of resource use that are included and how they are measured and valued. A discount rate of 4% is chosen in accordance with the Dutch *'Cabinet standpoint on the reconsideration of the discount rate to be used in assessing public sector projects'* (Ministerie van Financiën, 1995).

Basic principles that only consider the measurement and valuation of costs are the use of opportunity costs, marginal costing, the exclusion of costs of non-related diseases during life-years gained, and the inclusion of direct and indirect taxes. A common tool in the valuation of resource use is the concept of opportunity costs, i.e. the value of the foregone benefits because the resource is not available for its best alternative use (Drummond et al., 1997). In perfect competitive markets, prices of inputs are equal to opportunity costs, but this does not hold for most markets in the healthcare sector. Consequently, tariffs and other prices in the healthcare sector should be applied with care, and often other valuation methods are used instead. The concept of opportunity costs can help to find solutions in such cases.

The concept of marginal costing implies that only the increase in costs that is related to the introduction of an intervention is to be considered. Hence, fixed costs do not have to be included. In the long run, however, almost all costs depend on the amount of output produced. Marginal costs in these cases will equal average costs (Gold et al., 1996). Because economic evaluations should include long-term costs, and because the results of such evaluations are often used to support decision-making on a national level, we recommend the use of average costs. Only in specific situations, with, for example, a restricted time span, is marginal costing more appropriate.

In current theory on costing in economic evaluations, no consensus has been reached about whether to include or exclude the costs of non-related diseases during life-years gained (Canadian Coordinating Office for Health Technology Assessment, 1997; Commonwealth Department of Health and aging, 2002). In the Netherlands, it has been decided for practical reasons that these costs should not be included in the economic analysis.

Several authors have argued that direct and indirect taxes and social premiums should be excluded from cost analyses as such taxes and premiums do not represent costs for society (Gold et al., 1996). On the other hand, market prices of products and labour include direct and indirect taxes, and, from an opportunity cost approach, it can be argued that these taxes should be taken into account. Because in most situations it is difficult to exclude (especially indirect) taxes, the Dutch manual has adopted a practical approach and determines that these taxes should be included.

The use of all basic principles outlined in this section is obligatory in the case of studies that support submissions to acquire reimbursement for new pharmaceuticals. The importance of using similar basic principles among studies is obvious. Such principles are related to all stages of the costing process and have important implications for the selection, measurement and valuation of resources.

Methods for the measurement and valuation | Methods for the measurement and valuation can be distinguished into costs within and outside the healthcare sector. Most parts of the manual deal with the measurement and valuation of healthcare resource use for which many alternatives are available. It is difficult to standardise the use of these methods because of the differences between studies. In general, the precision of the cost estimate should reflect the contribution of the resource item to the incremental or total cost, but, as described above, other criteria also play a role. Therefore, these parts of the manual are mainly descriptive, and the choice of any method in a particular situation is left almost entirely to the investigator. The contribution in achieving more uniformity in cost calculations through this type of standardisation is found in standardising the considerations that are followed by the selection of methods and sources and not in prescribing the methodologies that have to be used in a particular setting.

In economic evaluations performed from a societal perspective, direct and indirect costs outside the healthcare sector also have to be considered. It is impossible to standardise the measurement and valuation of all different cost categories that can arise. Only a limited number of cost items, like travel expenses and time and productivity costs, play a role in almost any economic evaluation. In particular, the measurement and valuation of time and productivity costs has led to much discussion among health economists (Brouwer et al., 1998; Gold et al., 1996; Johannesson, 1997).

The manual has adopted the viewpoint that time costs of patients, other than lost productivity, should not be valued in monetary terms, but in terms of quality of life. Only time costs of family and friends for informal care should be valued. For the valuation of informal care, a price of € 8.00 per hour is recommended. Productivity costs are divided into paid and unpaid work. Lost productivity during paid work has to be valued using the friction cost method (Koopmanschap et al., 1995). This method is based on the idea that organisations need a certain time-span (the friction period) to restore the initial production level after an employee becomes absent from work. The amount of production lost due to disease depends on the length of this friction period. Productivity costs are calculated by multiplying the days absent from work with the costs per day; the number of days absent from work is limited to

the duration of the friction period. Unpaid work not only concerns voluntary jobs but also, for example, housekeeping. These costs can be considerable, especially in those with chronic diseases. For the valuation of lost productivity during unpaid work, a price of € 8.00 per hour is recommended, which is identical to the price for informal care.

Use of the friction cost method, and the prices mentioned above, is obligatory in studies designed to obtain reimbursement for new medications. Because estimates of productivity costs have been shown to depend highly on the approach taken (Koopmanschap and van Ineveld, 1992), the use of similar methods and prices will have an important contribution to the standardisation of costs outside the healthcare sector.

Standard costs | Standard costs were first introduced in the Australian costing guide (Commonwealth Department of Health and Ageing, 2002). They are average unit costs of standard resource items and, in the Dutch manual, are presented as one of the available methods for the valuation. A selection of standard costs from the manual is presented in *table 2.1*. In this first version of the manual, only standard costs of the resource use items most often used in

Unit	Standard cost
Inpatient days	
General ward general hospital	240
General ward academic hospital	330
Intensive care unit	1140
Psychiatric hospital	180
Nursing home	135
Outpatient visits	
Specialist in a general hospital	40
Specialist in an academic hospital	70
Other costs within healthcare	
Visit general practitioner	17
Visit physiotherapist	18
Hour of home care	23
Hour of revalidation treatment	80
Travel expenses	
Car (per km)	0.12
Parking costs car (per visit)	1.15
Public transport (per km)	0.12
Taxi (per km)	1.35
Other costs outside healthcare	
Hour of informal care	8.00
Costs per day absent from paid work	variable*

Table 2.1: Standard costs in 1999 euro.
** Depending on age and sex*

economic evaluations are available and do not take into account any differences between patient groups. This, for example, means that only the overall average costs of an inpatient day in a hospital are given. Data that allow differentiation of standard costs according to diagnosis or disease severity are currently not available. The use of standard costs is therefore not obligatory and, for the units with the largest contribution to total or incremental costs, a more detailed costing approach might still be necessary. Despite this limitation, the introduction of standard costs can have an important contribution to the standardisation of costs, by eliminating some of the price differences between studies.

Standard values | Another important feature of the manual is the introduction of 'standard values'. Unlike standard costs, standard values do not reflect actual unit costs of services or procedures, but provide parameters that are used in the calculation of unit costs. Examples of standard values within the healthcare sector are the number of workable hours of employees per year, and the yearly turnover of specialists. Examples of standard values outside the healthcare sector are the average distance of a household to a hospital (to calculate travel expenses), and the duration of a friction period (to calculate productivity costs). A selection of the standard values from the manual are presented in *table 2.2* and an example of the use of these values in the calculation of the unit cost of an outpatient visit is given in appendix 2.1.

Standard values contribute to standardisation by reducing differences in unit cost estimates between studies. Estimates of specialists' costs per hour in the Netherlands, for example, differ considerably and have ranged from € 113500 to € 260000. (Commissie modernisering curatieve zorg, 1994; Groeneveld, 1996). Such an amount of variation can importantly affect unit cost estimates. The use of standard values helps to overcome these types of differences, and is obligatory in studies used to support reimbursement for new medications. Only when the researcher can demonstrate that actual values differ considerably from standard values, are deviations from standard values approved.

Presentation of outcomes | In addition to the six-step procedure for costing, the manual presents some additional topics. One of these concerns the reporting of costs in the methods' and results' sections. The manual defines some minimum standards for the presentation of such sections. Important issues concern the reporting of basic principles, sources and methods that have been used in the costing. For results, resource use and unit costs always have to be presented separately. In addition to mean values, the variation around the mean has to be described.

Several authors have already emphasised the importance of a clear and standardised reporting format for economic evaluations (Mason and Drummond, 1995; Siegel et al., 1996). Standardisation of a reporting format does not directly lead to more standardisation, but contributes to the comparability and generalisability of economic evaluations.

2.4 DISCUSSION

The manual has only recently been issued and its contribution to standardisation and uniformity in costing is therefore yet unknown. After the manual is applied in a number of studies,

Parameter	Standard values
Payroll employees	
Gross salary	collective labour agreement*
Allowance for irregular working hours	collective labour agreement†
Working hours per year employees	1540
Working hours per year residents	2150
Mark up for premiums social security and retirement	32%
Secondary labour costs	3%
Medical staff	
Workable hours per year	2100
Proportion direct patient related time	70%
Yearly turnover self-employed specialist (in 1999 euro)	160000
Yearly income specialist academic hospital (in 1999 euro)	115000
Average distance in km to:	
Hospital	7.0
General practitioner	1.8
Physiotherapist	1.8
Other	
Annual discount rate	4.0%
Long term interest	4.5%
Friction period	123 days

Table 2.2: Standard values from the manual.
** The manual advises use of the salary-scale that is in accordance with the job. Within this scale, one should use the value just above the middle value as an estimate of the average salary; † the allowance for working time outside banking hours is determined by the collective labour agreement and varies from 22% to 72%.*

its use will be evaluated by the Dutch Health Insurance Board. Standard costs and standard values will be updated every two or three years. In future versions of the manual, standard costs will also be further developed. It is intended to provide prices for healthcare services for which, at the moment, no prices are available and to assign prices to diagnostic groups. This is especially true for unit costs of inpatient days and outpatient visits.

So far, guidelines for costing in pharmacoeconomic evaluations have been issued in Australia, Canada and the Netherlands (Canadian Coordinating Office for Health Technology Assessment, 1996; Commonwealth Department of Health and Ageing, 2002; Oostenbrink et al., 2000). Large differences exist between these costing manuals. The Australian costing guide provides an extensive list with standard costs of hospital services. These standard unit costs are obtained by large costing studies in which many hospitals take part and in which unit costs are related to diagnostic-related groups (DRGs). The use of these standard costs is obligatory in case of formal appraisal studies for new medications. The Canadian *'Guidance document for the costing process'* is more concise and does not yet present a standard cost list or standard costs. It presents the basic principles and methods for the measurement and valuation of resource use and is mainly descriptive. More recently, developments have started to formulate a standard

cost list for the Canadian situation as well (Jacobs and Roos, 1999). In the Netherlands, a systematic collection of standard unit costs does not take place and DRG-related prices are not yet available. The use of an extensive list with standard costs as in Australia is therefore not possible. The Dutch manual attempted to strike a balance between the Australian and Canadian costing guides. It is more detailed and gives more guidance than the Canadian costing guide, but is less prescriptive than the Australian manual in the use of standard costs. With the introduction of standard values and the way standard costs are treated, the Dutch manual introduces new elements that may contribute to the standardisation of costs in economic evaluations.

Bottom up unit cost calculation has some disadvantages compared with other types of valuation in that: 1) it is time consuming and expensive to perform; 2) the degree to which results can be generalised to other populations or other centres can be limited; and 3) it will not always be possible to get access to the necessary information within a healthcare organisation. Despite these disadvantages, detailed cost calculations are recommended for those units with the largest contribution to total or incremental costs. This is because, especially for some of the units for which unit cost calculations will often be necessary (e.g. outpatient visits and inpatient days), no acceptable alternatives are available. Charges for outpatient visits consist of a fee for each patient visiting a physician and are independent of the number of visits made by a particular patient. Hence, a fee per visit is not available. Charges for inpatient days are used to match actual income of the hospital with its budget and vary by hospital and by year. Moreover, no distinction is made between inpatient days on a ward and in an intensive care unit. This implies that as long as no good alternatives are available, it will still be necessary to use detailed unit costing in the valuation. Only a systematic collection of unit costs from hospitals and other healthcare organisations might overcome the need for performing detailed unit cost calculations in (pharmaco-) economic evaluations in the Netherlands.

If, in future, other countries develop guidelines for the costing-process, each country will have to find its balance between standardisation of costs and the degree to which standardised costs are valid in a specific setting. Such guidelines also have to fit with current practice and the availability of data in each country. The way in which the Dutch manual tries to reach this balance can perhaps serve as an example. The increase in the number of studies performed in a multinational setting, and the growing demand to extrapolate results from one country to another, also asks for more uniformity of costing among countries. To improve the standardisation and comparability of economic evaluations among countries is another challenge for health economists.

■ APPENDIX 2.1: EXAMPLE OF THE USE OF STANDARD VALUES

The unit cost of an outpatient visit is calculated using the buttom-up method. The costs of the specialists make-up an important share of the costs of this service. The time of the specialist per visit is estimated to be 10 minutes. The standard values for determining the costs of a specialist are an average annual income (turnover) of € 160000, 2100 workable hours a year, and a percentage of direct patient-related time of 70. With the help of the standard values, the specialist's cost per hour can be calculated as € 160000 × 2100 × 70% = € 109. The specialist's cost per visit are estimated to be (10/60) × 109 = € 18.

CHAPTER 3

UNIT COSTS OF INPATIENT HOSPITAL DAYS

3.1 INTRODUCTION

In many countries, the calculation of unit costs is an established method for the valuation of resource use in economic evaluations (Drummond et al., 1997; Gold et al., 1996). This method implies that unit costs are calculated on the basis of financial and production data obtained from healthcare centres participating in medical research. Often, the use of unit cost calculation is necessary because no other cost prices are available. Unit costs taken from earlier trials are frequently not valid for use as they have been determined for a specific patient group or because of lack of clarity about the methods used for the cost calculations.

Several authors have addressed the problem of variation in unit costs among centres. Goeree et al. found wide variations in unit costs of comparable procedures among hospitals (Goeree et al., 1999). The reasons for these differences are often unknown. They may reflect actual differences in the use of resources, differences in the hospitals' administration systems or differences in the methods used in the calculations. These differences imply that the outcomes of the cost calculations are highly dependent on the hospitals selected. Recently, Raikou et al. have argued that resource use depends on unit costs, i.e. that centres with a relatively low cost for a certain procedure are likely to demonstrate higher use of this procedure than centres with high unit costs (Raikou et al., 2000). Consequently, averaging unit costs from different hospitals for the valuation of all resource use in a trial will overestimate actual costs. In spite of these limitations, the calculation of unit costs is frequently the only available method to value resource use.

In many studies, costs of inpatient hospital days are reported to be the main drivers of total treatment costs (Hakkaart-van Roijen, 1998; Polder, 2001). For instance, Sullivan et al. have calculated that 73% of the total healthcare costs for patients with COPD were due to inpatient hospitalisation and emergency room visits (Sullivan et al., 2000). In a study on bone marrow or stem cell transplantation for patients with non-Hodgkin's lymphoma or Hodgkin's disease it was shown that the costs of inpatient days made up 60% of the total treatment costs (van Agthoven et al., 2001). In such cases, the unit cost estimate of an inpatient hospital day is an important factor that influences total costs. In the Netherlands, there is a lack of clear public disseminated information on unit costs of inpatient hospital days. Data from national registries regarding the unit costs of inpatient days are not available. Inpatient hospital day charges show a wide variation between hospitals and between years as they only serve as a tool to match the hospitals' incoming cash flows to their annual budgets. Charges can therefore not be used as approximations of actual unit costs.

Because of the problems in determining unit costs of inpatient days and their importance in many economic evaluations, we determined unit costs of inpatient hospital days from 22 general wards and 11 intensive care units (ICUs) of 15 hospitals from a healthcare provider's perspective. The aim of this paper was to provide data regarding unit costs of inpatient hospital days and to give insight into the extent to which cost categories and total costs differ between hospitals.

Data from this study have been used to derive a standard cost for inpatient days for the Dutch *'Manual for Costing: Methods and Standard Costs for Economic Evaluations in Healthcare' (Oostenbrink et al., 2000)*, which was written in addition to the *'Dutch guidelines for*

pharmacoeconomic research' (Riteco et al., 1999). The manual was approved by the Dutch Ministry of Health and the Health Insurance Council and, as described in the previous chapter, describes the standards, methods and standard costs to be used in studies set up to obtain reimbursement for new pharmaceuticals in the Netherlands.

3.2 METHODS

Data collection | Unit costs were collected from ten wards in eight general hospitals, 12 wards in six university hospitals and from 11 ICUs of which four were in general hospitals and seven were in university hospitals. These hospitals and wards were involved in clinical trials with piggybacked economic evaluations that were performed by our institute during the years 1995-1999. Cost data were collected in co-operation with employees from the financial departments of the hospitals. If data were not accessible from the central accounting system, additional information was obtained from the assessed ward or other departments. This mainly concerned the costs of medical staff, residents (physicians in specialty training) and medication.

All hospitals made use of the uniform accounting scheme designed by the Dutch Hospital Institute (Prismant, 2003). This scheme provides a standard classification in sections and account numbers for Dutch intramural healthcare organisations. A standard form was developed to standardise the extraction of the required financial data from the hospital accounting systems referring to the sections of this accounting scheme. An explanation and instruction for completion of the form were included. For cost categories for which no estimate was provided, hospitals were asked to indicate whether these data were missing or whether these costs were included in any of the other categories. Data on the following cost categories were collected: medical staff (specialists and residents), nursing, direct administrative personnel and management, materials, nutrition, medication, blood components, laundry, cleaning, accommodation (depreciation, energy, maintenance, etc.) and overheads and equipment (including capital costs). For all hospitals, the timeframe for the accounting information was one year and fell between 1995 and 1998. All costs were expressed in euro and Dutch healthcare price indices were used to convert costs to 1998 (Centraal Bureau voor de Statistiek (Statistics Netherlands), 2004).

Calculation methods | A distinction was made between direct and indirect costs. Direct costs are costs that are booked directly to the nursing and medical departments, for instance, wages of nurses and medical materials. All direct costs were calculated by use of an *'account classification'* approach (Horngren, 1982). This approach implies that for each category, the annual costs of the nursing department were derived from the financial accounts of the hospital and divided by the annual number of inpatient days. Nursing costs included wages, social premiums, fees for irregular working hours and the costs of replacement during illness. In Dutch hospitals wages of nursing staff are based on national pay scales for either general or academic hospitals and vary according to age, experience and specialty.

Indirect costs, such as accommodation or overheads, were calculated by allocating costs to the final cost centres and dividing these costs by the annual number of inpatient days. A final cost centre is defined as a department providing healthcare services to patients, for

example, operating theatres, nursing departments or outpatient clinics. The allocation of indirect costs depended on the financial accounting information available. In hospitals that had their own allocation or 'cost price' system, to assign indirect costs to the final cost centres we adopted this hospital allocation system. In hospitals where this was not the case, we made use of direct allocation (Horngren, 1982). This means that the costs of indirect cost centres were assigned to the final cost centres using a weighting methodology, based on various weighting statistics (Davidoff and Powe, 1996). Examples of such weighting statistics are the area surface (m^2) used to allocate costs of accommodation or the number of full time equivalents used to allocate the costs of the personnel department. After assignment of indirect costs to a ward, costs were divided by the annual number of inpatient days to obtain the mean cost per inpatient day. Whether costs of nutrition, blood components, medication and laundry were direct or indirect varied among hospitals according to the sophistication of their accounting systems. Costs of medications were based on the total costs assigned to that department divided by the number of inpatient days. Hence, these estimates were not influenced by the specific patient group the study was originally designed for. Costs of surgical and paramedical procedures, radiographic imaging and laboratory tests were not included in the costs per inpatient day.

In most hospitals, medical staff members were self-employed and, consequently, costs of medical staff could not be obtained from the financial accounts of these hospitals. Instead, specialists and residents were asked to estimate the average time per day spent on a single patient during hospitalisation. Costs were calculated by multiplying the time per inpatient day with fixed cost estimates based on average annual incomes of specialists and residents (Oostenbrink et al., 2000). Only in a few hospitals were costs of residents booked directly to the cost pool of the nursing department. In these hospitals, the calculations were not based on time estimates per day, but the total costs of residents were divided by the total number of inpatient days of that department. The annual income of specialists depended on whether they were self-employed (general hospitals) or in hospital service (university hospitals). The annual income of self-employed specialists in 1998 was estimated to be € 159000 and € 113500 for specialists in hospital service. Annual costs of residents were estimated to be € 45000. These estimates were based on the standard values recommended in the Dutch Manual (Oostenbrink et al., 2000).

Missing data | Several hospitals were not able to provide data on all cost categories that were asked for. Two types of missing data were distinguished. The first type concerned cost categories that were not administrated separately but included in one of the other categories. For example, this was the case when costs of accommodation and cleaning were included in the costs of overheads. In these cases, the proportion between these two cost categories in hospitals of which the values of both categories were known, was used to obtain an estimation for the missing values. The second type of missing data occurred when values of cost categories were unknown and not included in any of the other cost categories. In these cases mean imputation was used, based on the values of hospitals with data for this cost category.

3.3 RESULTS

Missing data | *Table 3.1* shows the percentages of missing data. Costs of residents were missing most often. Hospitals were able to provide an estimate of these costs per inpatient day in only nine (27%) of the 33 cases. Costs of accommodation and cleaning were also frequently missing. In almost half of the cases these costs were included in the costs of overheads and equipment and in six (18%) cases they were missing completely. In 12 (36%) cases no information was obtained on the costs of medication. In only one hospital (university hospital 2) no data were collected on the costs of nursing and materials, because the data from this hospital were no longer used in the original economic evaluation. The hospital was nevertheless included in the current analysis, since data of a few cost categories had already been collected.

Cost category	Included in another category	Missing completely
Specialists	0 (0)	13 (39)
Residents	0 (0)	24 (73)
Nursing	0 (0)	1 (3)
Materials and blood products	3 (9)	1 (3)
Nutrition	8 (24)	4 (12)
Medication	4 (12)	12 (36)
Laundry	11 (33)	6 (18)
Accommodation and cleaning	15 (45)	6 (18)
Overheads and equipment	0 (0)	7 (21)

Table 3.1: Number (%) of missing observations (33 wards and ICUs).

Unit costs of inpatient days | *Table 3.2, 3.3 and 3.4* show the unit costs of inpatient days of general hospitals, university hospitals and ICUs, respectively. Missing cost categories are presented in bold or italics. The bold values indicate costs that were originally included in another cost item. Italicised values indicate costs that were originally missing completely and were not included in any of the other cost items. Total costs per inpatient day ranged from € 120 to € 277 in general hospitals, from € 175 to € 366 in university hospitals and from € 838 to € 1479 in ICUs. In all hospitals, nursing costs were the major cost component and ranged from 32% to 64%. In most wards and ICUs, costs of overheads and equipment were the second highest cost component. All cost categories varied substantially between hospitals and wards.

Specialist costs | The costs of specialists and residents were omitted in the *tables 3.2 to 3.4*, because they were collected in only a few hospitals and because uniform parameters were used for the valuation of these costs, eliminating the variation in costs between wards and hospitals. The mean (n; standard deviation: SD) time specialists spent per inpatient day on a single patient was 12 (4; 2) minutes in general hospitals, 12 (8; 8) minutes in university hospitals, and 24 (8; 12) minutes in ICUs. The resulting costs per inpatient day were € 22, € 15 and € 43 respectively. The mean (n; SD) costs of residents per inpatient day were € 12 (4; 4) in general hospitals and € 19 (3; 12) in university hospitals. Costs of residents in ICUs could

Number*	1	2	3	4	5	6	7	8	9	10
Hospital size†	2	2	3	1	1	3	2	1	2	3
Type of ward	Internal	Pulmonary	Internal	Oncology	Oncology	Internal	Surgery	Surgery	Internal	Internal
No. of nursing days ward	11623	12081	13300	12483	13300	9787	20917	24403	6535	6873
Cost category										
Nursing	61 (32)	62 (34)	55 (46)	165 (64)	96 (51)	114 (41)	59 (39)	45 (33)	92 (45)	131 (51)
Materials & blood products	4 (2)	3 (2)	2 (2)	8 (3)**	6 (3)**	7 (3)	4 (3)	10 (7)	5 (2)	10 (4)
Nutrition	18 (9)	18 (10)	15 (13)	17 (7)	17 (9)	9 (3)	11 (8)**	11 (8)	15 (7)	14 (5)**
Medication	28 (15)	24 (13)	6 (5)	25 (10)***	25 (13)***	25 (9)***	25 (17)***	6 (5)	36 (18)	49 (19)
Laundry	4 (2)	4 (2)	6 (5)	5 (2)**	5 (3)**	1 (0)	4 (3)**	5 (4)***	7 (3)	4 (2)**
Accommodation & cleaning	29 (15)	27 (15)	18 (15)	17 (7)**	17 (9)**	14 (5)	17 (11)**	20 (15)***	25 (12)	18 (7)**
Overheads & equipment	48 (25)	42 (23)	18 (15)	22 (8)	22 (12)	107 (39)	30 (20)	37 (28)***	23 (11)	31 (12)
Total costs	192 (100)	180 (100)	120 (100)	259 (100)	188 (100)	277 (100)	150 (100)	134 (100)	203 (100)	257 (100)

Table 3.2: Unit costs (%) of general hospitals in 1998 euro.

***, *** indicate missing values: ** indicates values that were originally included in another cost item; *** indicates values originally missing completely and that were estimated by mean imputation.* Ward 1 and 2 were from the same hospital; ward 3 and 4 were from the same hospital; † based on the number of beds at the hospital: 1=0 – 300; 2=301 – 600; 3=601 – 900; 4=901-1200.*

Number*	1	2	3	4	5	6	7	8	9	10	11	12
Hospital size†	4	4	4	4	3	3	3	4	4	4	4	4
Type of ward	Gynaecology	Oncology	Surgery	ENT‡	Hematology	ENT‡	Hematology	ENT‡	Pulmonary	Internal	Internal	Oncology
No. of nursing days	7891	8318	35016	11570	6419	14060	3760	8756	7999	12688	15366	10665
Cost category												
Nursing	129 (40)	153 (50)***	112 (38)	175 (48)	139 (48)	184 (51)	94 (39)	86 (49)	112 (48)	118 (43)	81 (36)	124 (53)
Materials & blood products	5 (2)	23 (7)***	12 (4)	30 (8)	19 (6)	33 (9)	16 (6)	10 (6)	10 (4)	29 (10)	28 (13)	9 (4)
Nutrition	20 (6)	10 (3)	21 (7)	21 (6)	15 (5)***	15 (4)	15 (6)***	9 (5)**	14 (6)	14 (5)	11 (5)	10 (4)
Medication	14 (4)	14 (4)***	14 (5)	9 (3)	14 (5)***	22 (6)**	14 (5)***	8 (5)	9 (4)	15 (5)	15 (6)	14 (6)***
Laundry	8 (2)**	10 (3)	7 (3)	4 (1)	8 (3)***	8 (2)***	8 (3)***	5 (3)**	9 (4)	9 (3)	7 (3)	9 (4)
Accommodation & cleaning	53 (16)**	35 (11)	55 (19)	24 (7)	35 (12)***	35 (10)***	35 (15)***	21 (12)**	28 (12)**	33 (12)**	22 (10)**	47 (20)
Overheads & equipment	94 (29)	61 (20)***	70 (24)	103 (28)	61 (21)***	61 (17)***	61 (26)***	36 (21)	50 (22)	58 (21)	60 (27)	22 (9)
Total costs	323 (100)	306 (100)	291 (100)	366 (100)	291 (100)	358 (100)	243 (100)	175 (100)	232 (100)	276 (100)	224 (100)	235 (100)

Table 3.3: Unit costs (%) of university hospitals in 1998 euro.

***, *** indicate missing values: ** indicates values that were originally included in another cost item; *** indicates values originally missing completely and that were estimated by mean imputation. * Ward 3 and 4 were from the same hospital, ward 5 and 6 were from the same hospital; ward 8-11 were from the same hospital; † based on the number of beds of the hospital: 1=0 – 300; 2=301 – 600; 3=601 – 900; 4=901-1200; ‡ ENT: otorhinolaryngology.*

Number*	1	2	3	4	5	6	7	8	9	10	11
Hospital size†	4	4	3	3	3	4	4	3	2	3	3
Type of hospital	University	University	University	University	University	University	University	General	General	General	General
No. of nursing days ICU	3526	4694	3413	5643	2256	3208	10337	5555	1520	3061	6904
Cost category											
Nursing	539 (49)	560 (47)	409 (48)	394 (47)	474 (51)	608 (56)	460 (49)	688 (56)	403 (47)	871 (59)	538 (56)
Materials & blood products	183 (17)	242 (20)	112 (13)	201 (24)	117 (13)	172 (16)	78 (8)	96 (8)	70 (8)	122 (8)	79 (8)**
Nutrition	23 (2)	22 (2)	16 (2)**	16 (2)**	18 (2)***	18 (2)***	11 (1)	9 (1)	15 (2)**	15 (1)**	15 (1)**
Medication	101 (9)	142 (12)	97 (11)***	97 (12)**	97 (10)***	62 (6)	83 (9)	95 (8)***	95 (11)***	95 (6)**	95 (10)**
Laundry	18 (2)	5 (0)	8 (1)**	8 (1)**	9 (1)***	9 (1)***	7 (1)	5 (0)	5 (1)**	5 (0)**	5 (0)**
Accommodation & cleaning	74 (7)	51 (4)	60 (7)**	35 (4)**	61 (7)***	61 (6)***	87 (9)**	28 (2)	82 (10)**	111 (8)**	68 (7)**
Overheads & equipment	164 (15)	171 (14)	150 (18)	87 (10)	158 (17)***	158 (14)***	217 (23)	307 (25)	192 (22)	260 (18)	158 (17)
Total costs	1102 (100)	1193 (100)	852 (100)	838 (100)	934 (100)	1088 (100)	943 (100)	1228 (100)	862 (100)	1479 (100)	958 (100)

Table 3.4: Unit costs (%) of Intensive Care Units in 1998 euro.

***, *** indicate missing values: ** indicates values that were originally included in another cost item; *** indicates values originally missing completely and that were estimated by mean imputation.*

** ICU 1 and 2 were from the same hospital, ICU 3 and 4 were from the same hospital, ICU 6 and 7 were from the same hospital; † based on the number of beds of the hospital: 1=0 – 300; 2=301 – 600; 3=601 – 900; 4=901-1200.*

not be obtained. To be able to calculate the average costs of ICUs, we assumed that costs of residents in ICUs were twice as high as the costs on a ward in a university hospital (€ 38). This assumption was based on the proportion of time spent per inpatient day by specialists in university hospitals and ICUs.

Mean unit costs | The individual unit costs in *tables 3.2 to 3.4* were used to calculate the mean costs per inpatient day in general and university hospitals and ICUs respectively. These mean costs are presented in *table 3.5*, in which the costs of medical staff and residents were also added. Unit costs in university hospitals were approximately 40% higher than in general hospitals. Unit costs of ICUs were almost five times higher than the unit costs of wards in general hospitals and approximately 3.5 times higher than in university hospitals. *Figure 3.1* shows that the higher costs of ICUs corresponded with an increase in the relative proportion of nursing costs. Thirty-eight percent of the total costs in a ward in a general hospital constituted nursing costs, whereas this proportion in an ICU was 48%. The joint costs of overheads, equipment, accommodation and cleaning made up 25% of the total costs of a ward in a general hospital, 30% of a ward in a university hospital and 22% of the unit costs of an ICU.

Type of hospital	General N=10	University N=12	ICU N=11
Cost category			
Specialists	22 (18-27)	15 (6-39)	43 (6-58)
Residents	12 (9-27)	19 (6-30)	38 (12-60)
Nursing	88 (45-165)	126 (81-184)	540 (394-871)
Materials & blood products	6 (2-10)	19 (5-33)	134 (70-183)
Nutrition	14 (9-18)	14 (9-21)	16 (9-23)
Medication	25 (6-49)	25 (8-22)	96 (62-141)
Laundry	5 (1-7)	8 (4-10)	8 (5-18)
Accommodation and cleaning	20 (14-29)	35 (21-55)	66 (28-111)
Overheads and equipment	38 (18-107)	62 (22-103)	184 (87-307)
Total costs	230 (154-311)	323 (209-400)	1125 (919-1560)

Table 3.5: Mean (range) unit costs of inpatient days in 1998 euro.

3.4 DISCUSSION

In this study, the unit costs of inpatient days of a number of general and university wards and ICUs were collected. The mean costs per inpatient day were € 230 in a general hospital and € 323 in a university hospital. The mean costs per inpatient day on an ICU were € 1125. About 38% to 48% of the total costs were attributable to the nursing costs. All cost categories showed wide variations between hospitals.

One strength of the current study was the large number of hospitals in which unit costs were collected. To our knowledge, such an analysis has never been performed before. The results provide insights into the range of costs of inpatient days, into the differences between

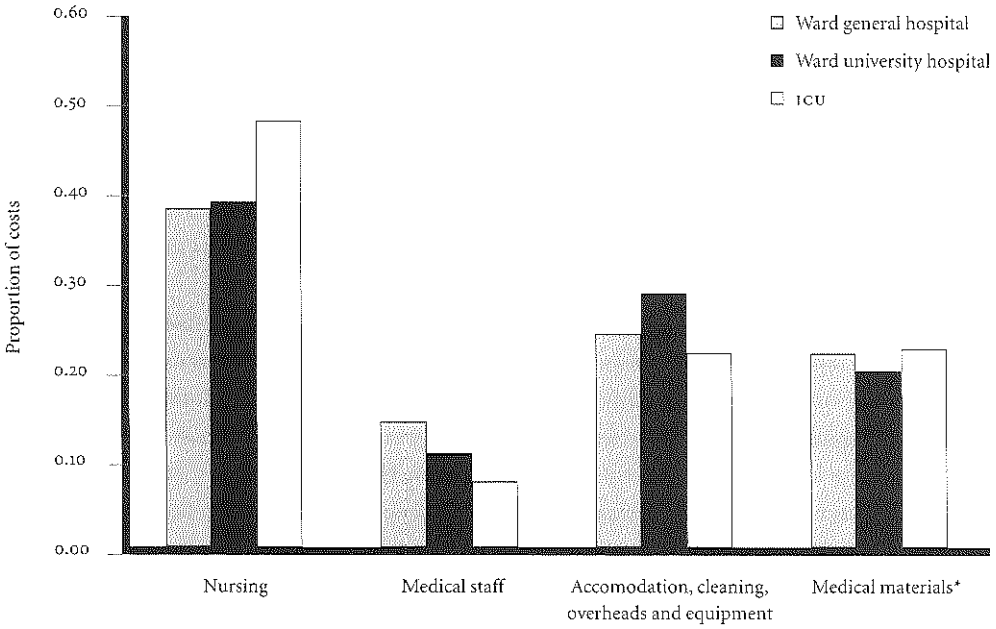


Figure 3.1: Relative contribution per cost category.
** Including materials & blood products, nutrition, medication and laundry*

general hospitals, university hospitals and ICUs and the variation in cost categories between hospitals. A limitation of this study is that hospitals and departments were not randomly selected, but included when they were involved in an economic evaluation performed by our institute. However, the selection of hospitals in this study included hospitals from different regions and from towns that differed in size. Moreover, the average hospital size (i.e. the number of beds) of general and university hospitals was comparable to all Dutch hospitals. We therefore think that another selection of hospitals would not have led to very different outcomes. Another drawback of this study was the considerable amount of missing data. Because only a limited number of characteristics of each ward and hospital were collected and no clear relationship between characteristics and costs of the hospital was found, only simple methods like mean imputation were used to replace missing values. We did not report standard deviations, as these were artificially reduced by the mean imputation method that we applied. Instead we reported the range of the observed values to give insight into the dispersion of unit costs. The results of this study give rise to the question: which factors were responsible for the differences in costs between wards and hospitals? To obtain insight into these factors we performed two multiple least squares regressions with type of hospital (general or academic), hospital size (number of beds), ward size (number of nursing days) and type of ward (internal yes/no) as the independent variables. In the first analysis we used total costs as the dependent variable. Because nursing cost had almost no missing values, this variable was used as the dependent variable in a second analysis. None of the independent variables showed a relationship with total or nursing costs that came near to significance, and therefore we have not presented the results of these analyses.

In addition to the variables distinguished above, other factors such as patient case-mix, differences in the measurement and valuation of costs, the sophistication of the hospitals' accounting systems, and the possible inclusion of incidental costs may have influenced the differences in unit costs between hospitals. Unfortunately, there is no detailed information available in the Netherlands about the case-mix of patients per hospital or hospital department. Therefore, we were not able to determine the impact of patient case-mix. Nevertheless we expect this to be an important factor contributing to the observed differences in unit costs between hospitals.

Another factor that is often mentioned as a possible source of variation in unit costs between settings is the measurement and valuation of costs (Goeree et al., 1999). Because the data collection was performed similarly in all hospitals by using a standardised form and because all hospitals involved used the same accounting scheme, we think that we have been able to control at least for these potential sources of variation. However, a possible source of variation that is closely related to the measurement and valuation is what we have called 'the sophistication of the hospitals' accounting systems'. For example, costs that were booked directly on a final cost centre in one hospital (considered to be direct costs) were sometimes booked to a central cost pool in another hospital. This, for example, was true for the costs of medication. In most academic hospitals, the costs of actual medication use were known for each nursing department, whereas in smaller hospitals, costs of all mediations were mostly booked to the central cost pool 'Pharmacy'.

Another factor related to the hospitals' accounting systems was the information that was available on weighting statistics for allocating indirect costs to final cost centres. Only a few hospitals were able to provide data on the measured area (m^2) of the different hospital departments to assign costs of accommodation and cleaning to the nursing departments. Part of the differences in overhead costs and the considerable amount of missing data amongst the indirect costs can be explained by the varying sophistication of the hospitals' accounting systems.

A final factor that might influence the variation in costs between hospitals is the extent to which costs of 'inefficiencies' or 'coincidences' were incorporated in the unit costs. For instance, a high illness rate of employees might lead to a considerable occasional increase in costs. In particular, fluctuations in the occupancy of a ward might strongly influence the annual cost per hospital day between years. It is unknown to what extent these possible sources of variation have implicitly been incorporated in this study.

In current practice of economic evaluations, unit costs are usually determined for the specific setting in which the study is performed and often depend on the availability of data. In 1993 the *'Manual of Resource Items and their Associated Costs'* (Commonwealth Department of Health and Ageing, 2002) was published in Australia, thus introducing the concept of 'standard costs' for use in economic evaluations. In recent years, Canada (Institute of Health Economics, 2000; Jacobs and Roos, 1999), the Netherlands (Oostenbrink et al., 2000) and the UK (Ferguson, 2001) have also adopted the idea of standard costs and have made efforts to build a database with unit costs. The results of this study have been used to develop standard costs for inpatient days in the Netherlands (Oostenbrink et al., 2000). The main reason

for introducing standard costs was to eliminate the proportion of cost differences due to measurement and variance in hospitals' accounting systems, and other factors that do not reflect actual differences in treatment between patient groups. Hence, standard costs can have an important contribution to the comparability and generalisability of economic evaluations.

However, there are at least three arguments against the use of standard costs. First, standard costs reflect average costs rather than marginal costs thus they give a poor estimate of the actual resources consumed. Second, it can be questioned whether the mechanical use of standard costs to value hospital days reflect the real opportunity costs as recommended in almost all textbooks. This is especially true in the situation where a (new) treatment prevents or shortens hospital admissions and empty beds are not occupied by other patients. Third, several authors have emphasized the relationship between resource use and unit costs and recommend the measurement of both in the same setting to avoid bias (Goeree et al., 1999; Raikou et al., 2000). We do not argue the accuracy of these points from a theoretical perspective. However, in the current practice of economic evaluations it is often not possible to collect unit costs in all centres involved. In addition, our study has shown that even if unit costs are collected the same way, it will be hard to eliminate all differences that are not related to actual differences in resources consumed. Finally, we think that the lack of comparability between (pharmaco-) economic evaluations is one of the major obstacles preventing their use in policy decisions within healthcare. Standard costs can therefore be welcomed as one of the instruments that can help to overcome this obstacle.

3.5 CONCLUSIONS

The aim of this study was to provide data about unit costs of inpatient hospital days in the Netherlands and to give insight into the extent to which cost categories and total costs differ between hospitals. This study showed that unit costs of inpatient days vary considerably between hospitals and that it appeared to be difficult to obtain information on all cost components from all hospitals. In addition to the type of hospital and the type of ward, we identified at least four other factors that may influence the difference in unit costs between hospitals. These were patient case-mix, differences in the measurement and valuation of costs, the sophistication of the hospitals' accounting systems and the possible inclusion of incidental costs. The use of standard costs may eliminate some of the differences that are not due to actual differences in the resources consumed and thus contribute to the standardisation and comparability of economic evaluations.

CHAPTER 4

ONE-YEAR COST-EFFECTIVENESS OF TIOTROPIUM VERSUS IPRATROPIUM TO TREAT COPD

4.1 INTRODUCTION

COPD is one of the leading causes of death and its prevalence is steadily increasing (Murray and Lopez, 1997). In the coming decades, general practitioners, respiratory physicians and other healthcare providers will be confronted with an increasing share of their patient population being COPD patients (Soriano et al., 2000). This is primarily a result of the aging of the population and their smoking behaviour in the past (Feenstra et al., 2001). There is an acute need for more effective treatment options to reduce the burden of this disease for patients, caregivers and society.

Tiotropium is a new inhaled bronchodilator for patients with COPD with a sustained duration of action indicated for once daily dosing (Hvizdos and Goa, 2002). Recent trials showed that tiotropium has superior efficacy compared to ipratropium, salmeterol and placebo (Casaburi et al., 2002; Donohue et al., 2002; Vincken et al., 2002). Based on the favourable results of these studies, it is suggested that *'if the cost is not prohibitive, bronchodilation in moderate COPD could move to once-daily tiotropium'* (Rees, 2002).

This paper addresses the health economic aspects of tiotropium as compared to ipratropium. It is the first pharmacoeconomic analysis of tiotropium, conducted to assess whether the benefits of this new therapy are achieved at reasonable costs. Such information is useful to support reimbursement and formulary decision-making and guide the positioning of tiotropium in the treatment spectrum for COPD.

The cost-effectiveness analysis was performed alongside the ipratropium-controlled clinical trials reported by Vincken et al. (Vincken et al., 2002). These clinical trials showed highly significant differences in the primary outcome measure trough forced expiratory volume in one second (trough FEV_1), defined as the mean of the two pre-dose measurements (i.e. 23-24 hours after the preceding dose of tiotropium, or 8-9 hours after the preceding dose of ipratropium). Trough FEV_1 was improved above baseline by 120 millilitre (ML) after one year for patients receiving tiotropium, whereas it was declined by 30 ML for patients receiving ipratropium. Tiotropium was also found to be more effective in improving dyspnoea, exacerbations and health-related quality of life.

The aim of the cost-effectiveness analysis was to compare tiotropium and ipratropium with respect to healthcare utilisation and costs, and to relate the difference in cost to the difference in COPD exacerbations and quality of life over a period of one year. As recommended in current guidelines, a comprehensive societal perspective was adopted (Drummond et al., 1997; Gold et al., 1996; Riteco et al., 1999).

4.2 METHODS

Design of the trials | This cost-effectiveness analysis was performed alongside two randomised controlled, double-blind, double-dummy, parallel group trials comparing 18 microgram (μg) tiotropium inhalation capsules administered once daily in the morning via the HandiHaler® device with ipratropium two puffs of 20 μg administered four times daily via the metered dose inhaler (MDI) in patients with airway obstruction due to COPD (Vincken et al., 2002). All drugs and devices used to administrate the drugs were supplied by Boehringer Ingelheim, Ingelheim am Rhein, Germany. The studies were conducted at 29 centres in the

Netherlands and Belgium between October 1996 and June 1998. Since the design of both trials was identical, the cost-effectiveness analysis was based on the combined data. The trials were co-ordinated by Boehringer Ingelheim Pharmaceuticals Inc. in cooperation with the participating centres. Analysis and interpretation of the data and the writing of the manuscript are the sole responsibility of the authors.

Patients | Current or exsmokers with relatively stable COPD and a $FEV_1 \leq 65\%$ of predicted normal (Quanjer et al., 1993) and $FEV_1 \leq 70\%$ of forced vital capacity (FVC) were included. Bronchodilator responsiveness was not an entry criterion. Patients were also required to be aged > 40 and to have a smoking history of at least 10 pack-years. Patients with a history of asthma, patients requiring regular supplemental oxygen and patients with a recent upper respiratory tract infection or a significant disease other than COPD were excluded. Patients were randomised per centre to either tiotropium or ipratropium in a ratio of 2:1 using a randomisation list with a block size of three. The sample size of the studies was based on the primary clinical outcome parameter trough FEV_1. To detect with 90% power and a type I error of 5% a change of 0.075 litre over one year, 240 patients per study were required. The trials were approved by the medical ethics committees of all participating centres and all patients gave written informed consent.

Data collection | Patients were followed for one year. After a two-week run-in period, patients were seen at baseline (start of study medication) and at weeks 1, 4, 7, 10, 13, 19, 26, 32, 39, 45 and 52 for scheduled regular visits. At all regular visits, data on healthcare utilisation, study drugs, concomitant therapy, and adverse events including COPD exacerbations were recorded in a CRF. Disease specific quality of life questionnaires were administered at baseline and after 1, 7, 13, 26, 39 and 52 weeks of treatment. All patients who completed at least one scheduled visit after randomisation were included in the cost-effectiveness analysis.

Health outcomes | Prespecified outcomes for the cost-effectiveness analysis were the number of COPD exacerbations and the number of patients with an improvement of at least four units on the St. George's Respiratory Questionnaire (SGRQ). An exacerbation was defined as a complex of respiratory symptoms (i.e. new onset or worsening of more than one symptom such as cough, sputum, dyspnoea or wheeze) lasting for ≥ 3 days. The SGRQ is a disease-specific questionnaire designed to measure the impact of chest disease on health-related quality of life and well-being (Jones et al., 1991). The questionnaire contains 50 items which can be aggregated into an overall score and three sub-scores for 'symptoms', 'activity' and 'impact'. An improvement of four units on the total score is considered to be the minimum clinically important difference (Jones, 2002; Jones and Lasserson, 1994). The impact of including other thresholds for the minimum clinically important difference is investigated in a sensitivity analysis. Additional outcomes to be considered secondarily in the cost-effectiveness analysis were the proportions of patients with an improvement in trough FEV_1 of at least 12% (American Thoracic Society, 1991) and the proportion of patients with an improvement of at least one unit on the transitional dyspnoea index (TDI) over one year (Mahler et

al., 1984; Witek and Mahler, 2003). The TDI is an interviewer-administered questionnaire designed to improve the clinical evaluation of dyspnoea over time. The TDI consists of three components, i.e. functional impairment, magnitude of task and magnitude of effort. A TDI focal score is obtained by adding the scores of the three components and ranges from + 9 (indicating a major improvement) to – 9 (indicating a major deterioration).

Resource use | All resource use, irrespective of its reason, was recorded prospectively in a detailed pharmacoeconomic section of the CRF. Resource use included hospital admissions (ICU and non-ICU days), emergency room (ER) visits, unscheduled visits to respiratory physicians, GPs and other healthcare providers, pulmonary function tests, imaging tests, laboratory tests, puffs of rescue medication (salbutamol, one puff = 100 µg), and concomitant medication. In addition, the number of days patients were unable to perform the majority of their usual daily activities was recorded. If hospitalisation continued after the end of the study, the total length of stay included the days after the one-year study period. Dates of resource use were recorded to establish a link between resource use and adverse events, which were recorded in another section of the CRF. In the base case analysis, only the respiratory-related resource use was included. This was defined as resource use related to adverse events that were classified as either: 1) COPD and lower airway complaints; 2) upper airway complaints; and 3) side-effects of study medication. The impact of including all resource use instead of respiratory-related resource use was investigated in a sensitivity analysis. Except for study medication, all protocol-driven resource use was excluded from all analyses.

Costs | In the base case analysis costs were calculated by multiplying the respiratory-related resource use of each patient with Dutch 2001 unit costs expressed in euro (*table 4.1*). All costs within the healthcare sector were taken into account, regardless of whether they were borne by government, health insurers or patients. Average unit costs of inpatient hospital days and outpatient visits were obtained from the costing study described in chapter 3 that aimed to set standard costs for economic analyses in the Netherlands. This study included seven internal (including pulmonary) wards and five outpatient internal clinics of general and university hospitals. All unit costs included the costs of nursing, materials, hotel costs and the costs of buildings, equipment and overhead. Costs of respiratory physicians were included in the unit costs of inpatient hospital days, ER visits and outpatient visits, and were based on average time-estimates of 30 pulmonologists involved in the trials. Costs of pulmonary function tests, imaging tests and laboratory tests were based on charges. Costs of medications were based on list prices and included value added taxes and a mark-up of € 6.02 per prescription to cover pharmacist fees. The price of tiotropium was determined assuming the annual use of one pack containing 10 units and the device and 11.83 packs with 30 units (refill). The price of ipratropium was based on a pack size of 200 units, administrated via the metered dose inhaler, the device used in the trials. Harmonised consumer price indices were used to convert unit costs of previous years to a 2001 price level (Centraal Bureau voor de Statistiek (Statistics Netherlands), 2004). Because the period of data collection covered only one year, no discounting was used.

	The Netherlands	Belgium*
Day general/pulmonary ward	222	256
Day intensive care unit	1110	769
Visit to pulmonologist	52	52
Visit to GP	17	15
Visit to nurse/physiotherapist	19	14
Visit to emergency room	98	70
Complete spirometry	34	33
Chest radiographs	42	13
Tiotropium (public price per day)	1.57	1.80
Ipratropium (public price per day)†	0.33	0.29

Table 4.1: Unit costs of the most important types of healthcare utilisation in 2001 euro.
** In the base case analysis, trial-wide resource use is multiplied with Dutch unit costs. Because Belgian unit costs are used in sensitivity analyses, these costs are also reported in this table; † Price of ipratropium based on administration by the MDI.*

Cost-effectiveness | The pre-specified incremental cost-effectiveness ratios were the healthcare costs per exacerbation avoided and the healthcare costs per patient with an improvement of at least four units on the total score of the SGRQ. In addition, we calculated the cost per patient with an improvement of at least 12% in trough FEV_1 and the cost per patient with an improvement of at least one unit on the TDI.

Missing data | In order to deal with missing data of patients not completing the study, multiple imputation was used. Multiple imputation is a technique that, instead of imputing one value for each missing observation, replaces each missing observation with a set of m (in this case ten) plausible values (Rubin, 1987; Rubin and Schenker, 1991). This resulted in ten complete data sets for which the overall mean and variance were estimated. The variance between data sets was combined with the variance within data sets and can be considered as the added uncertainty that results from missing values. Imputation within each of the ten data sets was performed using the propensity score method (Rosenbaum and Rubin, 1984). In this method, imputed values are drawn at random and with replacement from patients who are comparable on demographic and baseline characteristics and on costs and health outcomes in periods before dropout. Imputation was used for health outcomes and resource use and was performed in both treatment groups separately.

Analysis | Healthcare costs and measures of effectiveness were expressed as the mean (standard error: SE) costs and effects per patient and year. The 95% confidence intervals (CI) of the differences between treatment groups were calculated taking into account the between variance of the imputed data sets and assuming a normal distribution of the differences. To examine whether the normal distribution assumption held, we bootstrapped the major cost items of the individual data sets. This resulted in almost exact replicates of the 95% CIs as obtained with the normal approximation. The incremental cost-effectiveness ratios were calculated as the difference in costs between tiotropium and ipratropium divided by the difference

in effects. Due to statistical problems associated with the calculation of CIs for ratios, the uncertainty surrounding the cost-effectiveness ratio is presented graphically on a cost-effectiveness plane (CE plane) (van Hout et al., 1994). A CE plane is an x-y plot where the horizontal axis shows the difference in effects between the treatment arms (tiotropium minus ipratropium) and the vertical axis shows the difference in costs. The uncertainty around the point-estimate of the difference in costs and effects is surrounded by a 95% elliptical confidence region. The discussion on whether the cost-effectiveness ratio is acceptable depends on the maximum that decision makers are willing to invest to obtain one unit of effect (e.g. to avoid one exacerbation). Because the value of this maximum acceptable ratio is unknown, the likelihood that tiotropium is cost effective at different values of the maximum acceptable ratio is plotted as a cost-effectiveness acceptability curve (CE acceptability curve) (van Hout et al., 1994).

Sensitivity analysis | To investigate the impact of assumptions made during the analysis and to test the robustness of results given variation in the data input, a number of sensitivity analyses were performed. The first sensitivity analysis included all resource use, instead of the respiratory-related resource use only. In the base case analysis trial-wide resource use was combined with Dutch unit costs. In the second sensitivity analysis resources used by Belgian patients were multiplied with Belgian unit costs and resources used by Dutch patients were multiplied with Dutch unit costs, after which the results were combined. In the third and fourth sensitivity analysis the calculation of costs and health outcomes was based on country-specific unit costs and on the subgroup of patients treated in that particular country. In a fifth sensitivity analysis, the price of ipratropium was set to the average price of the metered dose inhaler (€ 0.33 per day) and the price of the dry powder inhaler (DPI; € 0.97 per day), weighted by the actual use of these devices in the Netherlands (44% MDI versus 56% DPI). In the base case analysis an improvement of four units on the SGRQ total score was defined as a minimum clinically important difference. In a final set of sensitivity analyses (SA6 and SA7) the threshold value for a relevant improvement on the SGRQ was varied and set to six and eight units respectively.

■ 4.3 RESULTS

Patients | A total number of 535 patients were randomized; 356 in the tiotropium group and 179 in the ipratropium group. About 85% of the patients were enrolled in the Netherlands and 15% in Belgium. A total of 92 patients (18%) withdrew from the study, 54 (15%) in the tiotropium group and 38 (21%) in the ipratropium group. Main reasons for withdrawal were worsening of COPD (11 (3%) in the tiotropium and 11 (6%) in the ipratropium group) and other adverse events (23 (6%) and 8 (4%) respectively). A total of 519 patients completed at least one scheduled clinic visit after the baseline visit and were included in the cost-effectiveness analysis. Lung function parameters of these patients at baseline were slightly higher in the tiotropium group. Other baseline characteristics were comparable across the treatment groups (*table 4.2*).

	Tiotropium N=344	Ipratropium N=175
Age	64 (8)	65 (8)
Males: number (%)	289 (84)	151 (86)
Dutch: number (%)	294 (85)	151 (86)
Current smokers: number (%)	151 (43.9)	79 (45.1)
Smoking history in pack years	33.8 (17.8)	33.2 (16.7)
Duration of COPD in years	11.3 (10.0)	10.9 (9.7)
FEV_1 (litres)	1.21 (0.44)	1.13 (0.38)
FEV_1 (% of predicted)	40.6 (12.8)	38.0 (10.6)
FVC (litres)	2.68 (0.85)	2.52 (0.71)
SGRQ total score	45.5 (16.6)	43.7 (17.6)

Table 4.2: Patient characteristics per treatment group at baseline.
All data presented as mean (SD) unless otherwise stated; baseline characteristics are slightly different from those reported in (Vincken et al., 2002) because of the different number of patients included for the economic analysis.

Exacerbations | The mean (SE) number of exacerbations per patient was 0.74 (0.08) in the tiotropium group and 1.01 (0.10) in the ipratropium group; a difference of 0.27 (95% CI: 0.02; 0.52) or 27%. The percentage (SE) of patients with at least one exacerbation was 39.9% (2.9) in the tiotropium group and 53.5% (3.9) in the ipratropium group, a difference of 13.6% (95% CI: 4.1%; 23.1%). Approximately 17% of the exacerbations in the tiotropium group and 23% in the ipratropium group were associated with a hospitalisation (Pearson Chi-Square: p=0.374).

Quality of life | The percentage (SE) of patients with an improvement of at least four units on the SGRQ after 1 year was 51.2% (2.8) in the tiotropium group and 34.6% (3.8) in the ipratropium group; a difference of 16.6% (95% CI: 7.4; 25.9). The percentage of patients with a deterioration of at least 4 units was 26.0% in the tiotropium group and 33.7% in the ipratropium group, a difference of 7.7% (95% CI: -1.0; 17.0).

Pulmonary function | The percentage (SE) of patients with an improvement of at least 12% in FEV_1 over 1 year was 47.6% (2.8) in patients treated with tiotropium and 25.0% (3.6) in patients treated with ipratropium, a difference of 22.6% (95% CI: 13.8; 31.6).

Dyspnoea | About 30.5% of the patients in the tiotropium group and 16.2% of the patients in the ipratropium group experienced an improvement of at least one unit on the TDI focal score over one year (difference 14.3, 95% CI: 5.9; 22.7).

Resource use | The mean resource use per patient is presented in *table 4.3*. This table shows a consistent pattern of lower resource use in patients treated with tiotropium. The number of hospital admissions in the tiotropium group was reduced from 0.13 to 0.24, a difference of 45% (p=0.03). Approximately 11% of the patients in the tiotropium group and 19% in the ipratropium group had at least one hospital admission (p=0.03). The mean (SE) number of inpatient hospital days was reduced by 42%, from 2.98 (0.58) in the ipratropium group to 1.72

(0.37) in the tiotropium group (p=0.07). The mean (SE) number of unscheduled visits was reduced by 36%, from 3.18 (0.52) in the ipratropium group to 2.04 (0.16) in the tiotropium group (p=0.04). Only the mean (SE) number of inpatient days in the ICU was 0.08 (0.12) days higher in the tiotropium group, mainly due to one patient with an ICU stay of 24 days (p=0.37).

	Tiotropium N=344	Ipratropium N=175	Difference	95% CI
Hospital admissions	0.13 (0.02)	0.24 (0.05)	-0.11	-0.21; -0.01
Inpatient days				
General ward	1.62 (0.33)	2.96 (0.58)	-1.34	-2.64; -0.04
Intensive care unit	0.10 (0.09)	0.02 (0.02)	0.08	-0.10; 0.26
Total	1.72 (0.37)	2.98 (0.58)	-1.26	-2.60; 0.09
Unscheduled visits				
Pulmonologist	0.58 (0.06)	0.68 (0.10)	-0.10	-0.33; 0.13
GP	1.16 (0.10)	1.48 (0.19)	-0.32	-0.75; 0.11
Other HCP	0.25 (0.08)	0.88 (0.39)	-0.63	-1.42; 1.52
Emergency room	0.05 (0.01)	0.14 (0.03)	-0.09	-0.16; -0.02
Total	2.04 (0.16)	3.18 (0.52)	-1.14	-2.20; -0.08
Ambulance transports	0.05 (0.02)	0.16 (0.07)	-0.11	-0.25; 0.02
Puffs of salbutamol (rescue medication)	605 (42)	714 (68)	-109	-267; 47
Inactivity days*	23.98 (2.87)	29.19 (4.03)	-5.21	-14.92; 4.49

Table 4.3: Mean (SE) resource use per patient and year.
HCP: healthcare provider; Description in the CRF was: 'Number of days unable to perform the majority of usual daily activities'.*

Costs | Mean (SE) total costs were € 1721 (160) in the tiotropium group and € 1541 (163) in the ipratropium group, a difference of € 180 (95% CI: -268; 627; *table 4.4*). The higher costs of study medication for tiotropium (€ 453) were partly offset by savings in other types of healthcare resource use (€ -273, 95% CI: -721; 174), especially inpatient hospital days (€ -208, 95% CI: -591; 175). Costs of concomitant medication made up 30% of total costs and were almost the same in both treatment groups, € 526 (SE 20) in the tiotropium group and € 511 (SE 25) in the ipratropium group.

Cost-effectiveness | Because tiotropium was more effective and associated with higher costs, all incremental cost-effectiveness ratios were positive. The cost-effectiveness ratio was € 667 per exacerbation avoided and € 1084 per patient with a relevant improvement in disease-specific quality of life. The cost per patient with a relevant improvement in dyspnoea was € 1259 and the cost per patient with a relevant improvement in FEV_1 was € 796. The uncertainty around the ratios of the two main outcome measures is presented graphically on the CE plane (*figure 4.1a and 4.1b*). The three ellipses in each figure represent the 5, 50 and 95%

	Tiotropium N=344	Ipratropium N=175	Difference	95% CI
Inpatient days				
General ward	359 (73)	657 (127)	- 298	-586; -10
Intensive care unit	116 (100)	26 (21)	90	-110; 291
Total	475 (144)	683 (132)	- 208	-591; 175
Unscheduled visits				
Pulmonologist	30 (3)	35 (5)	- 5	-17; 7
GP	20 (2)	26 (3)	- 6	-13; 2
Other HCP	5 (2)	17 (8)	- 12	-27; 3
Emergency room	5 (1)	13 (3)	- 8	-15; -2
Total	60 (5)	91 (13)	- 31	-57; 5
Concomitant medication	526 (20)	511 (25)	15	-47; 78
Rescue medication (salbutamol)	16 (1)	19 (2)	- 3	-7; 1
Diagnostic/prognostic tests	59 (12)	76 (14)	- 17	-54; 17
Ambulance transport	12 (4)	41 (16)	- 29	-60; 4
Costs without study medication	1148 (160)	1421 (163)	-273	-721; 174
Study medication	573 (0)	120 (0)	453	
Costs including study medication	1721 (160)	1541 (163)	180	-268; 627

Table 4.4: Mean (SE) healthcare costs per patient and year based on Dutch prices in 2001 euro.
HCP: healthcare provider.

confidence areas of the difference in costs and effects. In the CE plane of the costs per exacerbation avoided (*figure 4.1a*), approximately 24% of the surface of the ellipses was situated in the lower-right quadrant, signifying lower costs and less exacerbations in the tiotropium group, whereas 74% was situated in the upper-right quadrant signifying a reduction in exacerbations against higher costs. The dotted line from the origin through the point estimate of the difference in costs (€ 180) and effects (0.27) crosses one exacerbation avoided exactly at € 667, the incremental cost-effectiveness ratio. Similarly, in the CE plane of the costs per patient with a relevant improvement in SGRQ total score (*figure 4.1b*), approximately 25% of the surface of the ellipse was situated in the lower-right quadrant and 75% in the upper-right quadrant. Both figures (4.1a and 4.1b) show that the uncertainty around the ratio was largely due to the uncertainty around the cost-difference.

The CE acceptability curves for the two incremental cost-effectiveness ratios are shown in *figure 4.2*. The vertical axis shows the percentage of time that tiotropium is cost effective, given the value of the maximum acceptable ratio on the horizontal axis. If, for instance, the maximum willingness to pay per exacerbation avoided is set at € 2000, then the percentage of time that tiotropium is cost effective (i.e. has a ratio below € 2000) is 80%. Similarly, if the maximum acceptable ratio of the costs per patient with a relevant improvement on the SGRQ

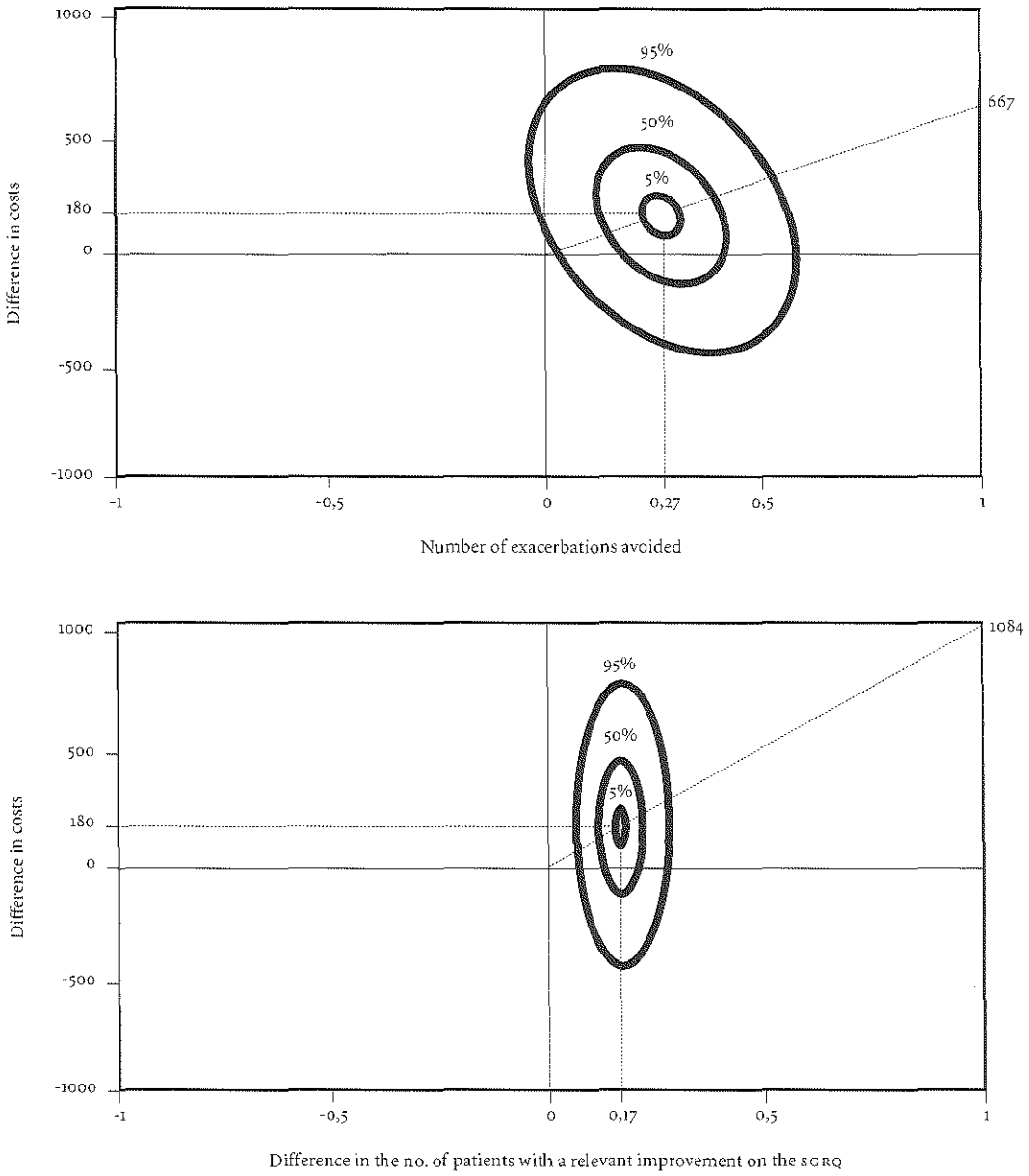


Figure 4.1: CE planes representing the 5, 50 and 95% confidence areas of a) the incremental healthcare costs per incremental exacerbation avoided and b) the incremental healthcare costs per patient with a relevant improvement on the SGRQ. The dashed horizontal line depicts the point estimate of the difference in costs between tiotropium and ipratropium (€ 180); the dashed vertical line depicts the point estimate of the difference in effects: 0.27 per exacerbation avoided and 0.17 per patient with a relevant improvement on the SGRQ; the dashed diagonal line, from the origin through the centre of the ellipses, takes the value of the cost-effectiveness ratios of € 667 and € 1084 respectively.

is set at € 2000, tiotropium is cost effective 72% of the time. The reading across the probability of 50% to the curves and down to the horizontal axis gives the point estimate of the incremental cost-effectiveness ratios (€ 667 and € 1084 respectively).

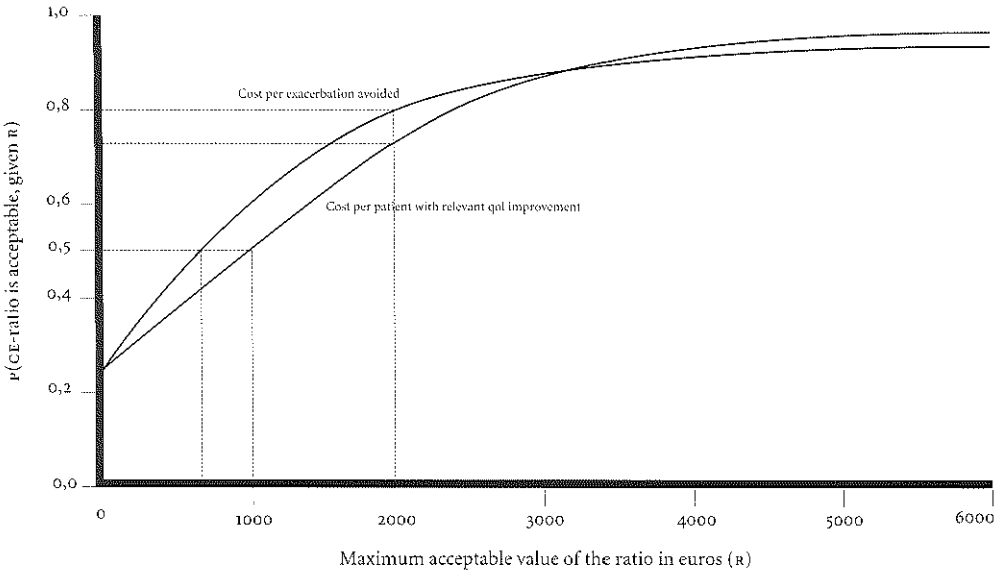


Figure 4.2: CE acceptability curves of the cost per exacerbation avoided and the cost per patient with a relevant improvement on the SGRQ.
The horizontal axis represents the maximum acceptable ratio of the cost per exacerbation avoided and of the cost per patient with a relevant (four units) improvement on the SGRQ total score respectively. If for instance the maximum acceptable ratio is set at € 2000, tiotropium will be acceptable 80% and 72% of the time respectively. The curves equal the point estimate of the incremental cost-effectiveness ratios at p=0.5. These are € 667 per exacerbation avoided and € 1084 per patient with a relevant improvement in quality of life.

	Mean (SE) difference in the cost per patient	Mean (SE) number of exacerbations avoided	Difference in proportion of patients improved on the SGRQ	ICER	
				Cost per exacerbation avoided	Cost per patient improved on the SGRQ
Base case	180 (228)	0.27 (0.13)	16.6% (0.05)	667	1084
SA 1	111 (277)	0.27 (0.13)	16.6% (0.05)	411	669
SA 2	221 (229)	0.27 (0.13)	16.6% (0.05)	819	1331
SA 3	203 (243)	0.25 (0.12)	13.9% (0.05)	812	1460
SA 4	159 (635)	0.43 (0.51)	33.2% (0.13)	370	479
SA 5	48 (228)	0.27 (0.13)	16.6% (0.05)	178	289
SA 6	180 (228)	0.27 (0.13)	13.3% (0.05)	667	1353
SA 7	180 (228)	0.27 (0.13)	11.1% (0.04)	667	1622

Table 4.5: Sensitivity analysis of the differences in costs and health outcomes between tiotropium and ipratropium.
ICER: incremental cost-effectiveness ratio; SA1: inclusion of all resource use; SA2: valuation based on country-specific prices; SA3: Dutch patients only; tiotropium n=294, ipratropium n=151; SA4: Belgian patients only; tiotropium n=50, ipratropium n=24; SA5: price ipratropium based on the average of the MDI and the DPI, weighted by the actual use of these devices in the Netherlands (MDI: 44%, DPI:56%); SA6: threshold value of a relevant improvement on the SGRQ set to six units; SA7: threshold value of a relevant improvement on the SGRQ set to eight units.

Sensitivity analysis | The results of the sensitivity analysis are presented in *table 4.5*. Inclusion of all healthcare resource use instead of respiratory-related resource use only (SA1), led to an increase in costs of approximately 40%, whereas the difference in costs between the two groups was reduced to € 111. Valuation based on country-specific unit costs (SA2) or inclusion of Dutch patients only (SA3) had a small impact on the difference in costs and health outcomes. If only the Belgian patients were selected (SA4), the difference in costs hardly changed, whereas the difference in the number of exacerbations increased from 0.27 to 0.43 and the difference in the proportion of patients with a relevant improvement on the SGRQ increased from 16.6% to 33.2%. However, the number of Belgian patients was small: 50 patients in the tiotropium group and 25 patients in the ipratropium group. Changes in the costs of ipratropium had the largest impact on the difference in costs (SA5). When the costs of ipratropium were based on the weighted average of the MDI and DPI price, the difference in costs was decreased to € 48, decreasing the cost-effectiveness ratio to € 178 per exacerbation avoided and € 289 per patient with a relevant improvement on the SGRQ. Increasing the threshold value of the SGRQ above which a change in health related quality of life was considered to be clinically important, decreased the proportions of patients improved with approximately 15% when the threshold value was set to six units (SA6) and with 30% when this value was set to eight units (SA7). The corresponding differences in the proportion of patients with a relevant improvement decreased to 13.3% and 11.1%, leading to higher cost-effectiveness ratios of € 1353 and € 1622 respectively.

4.4 DISCUSSION

This is the first cost-effectiveness study that directly compares the new, once-daily bronchodilator tiotropium to ipratropium. Compared to ipratropium, tiotropium led to a 27% reduction in the mean number of exacerbations and a 17% increase in the number of patients with a relevant improvement on the total score of the SGRQ. In addition, a significantly greater proportion of patients had a clinically relevant improvement in lung function and dyspnoea. These improvements in health outcomes were associated with increased costs of € 180 per patient per year. Hence, about 60% of the higher price of tiotropium (€ 453) was offset by a reduction in the costs of other healthcare resources (€ 273). These savings were primarily caused by a reduction in the number of hospital admissions and hospitalisation days, which were 45% and 42% lower respectively in patients receiving tiotropium than in patients receiving ipratropium. All other resource items, except concomitant medication, showed the same trend towards reduced costs in patients receiving tiotropium.

This economic evaluation was conducted alongside two RCTs. Hence, all resource use and health outcomes data were fully stochastic and collected prospectively over the one-year study period. This is an important strength of the current study as many economic evaluations use modelling often based on indirect data. Another strength of the study is the use of multiple imputation to deal with missing data of patients who dropped out before the scheduled end date after one year. This method imputes values that are sampled from patients who are comparable on demographic and baseline characteristics and on costs and effects in previous periods and makes full use of the costs and effects the withdrawals had during the

period they were still in the study. Above all, in contrast with other, more naïve, methods like case deletion, last value carried forward or mean imputation, multiple imputation takes account of the extra uncertainty that results from missing data, by imputing multiple values for each missing value (Rubin, 1987). In additional analyses presented in chapter 7, we will show that the estimates of the difference in costs between tiotropium and ipratropium obtained with multiple imputation were at least as conservative as those obtained with other methods to deal with the data of dropouts. Adopting a complete case analysis would have seriously underestimated the real costs in both treatment groups, as the more severely ill patients were more likely to dropout. In both treatment groups, the mean costs per day of the dropouts during the time they were in the study were approximately four times as high as the mean costs per day of patients who completed the study (see chapter 7).

This cost-effectiveness analysis was conducted from a societal perspective, which implies that all relevant costs should be taken into account. Hence, in addition to the direct health-care costs, we also studied the indirect costs, i.e. the costs associated with absence from work and inability to perform the usual daily activities. Compared to ipratropium, the number of days that patients were unable to perform their usual daily activities including paid work was 18% less in tiotropium, although this difference was not statistically significant. As there is still a lot of discussion on whether and how these days have to be valued (Brouwer et al., 1997; Gold et al., 1996; Johannesson, 1997), we choose not to include these indirect costs in the cost-effectiveness analysis. It is sometimes argued that the calculation of indirect costs is less relevant in a population of moderate and severe COPD patients, because only a small proportion of patients has a paid job. Indeed, in this study, only 9.6% had a paid job. Because the proportion of patients with a paid job at baseline differed between the treatment groups (9.3% in the tiotropium group and 10.3% in the ipratropium group), calculating the costs of lost working days would increase the risk of introducing a bias against ipratropium.

Among the disadvantages of an economic evaluation piggybacked to a clinical trial is the occurrence of protocol-driven costs. The costs of regular clinic visits were excluded because they were scheduled so frequently that they were not reflective of the treatment pattern in COPD. This may underestimate costs since these visits may have substituted visits that would have occurred if the trial had not taken place. On the other hand, because of the trial situation, patients may have felt less reservation to contact their physician sooner in case of minor complaints. The latter equally affects both treatment groups. The substitution effect however is more likely to occur in the ipratropium group as the condition of these patients was less well controlled. So, if there is a bias, it is more likely to be a bias against tiotropium. However, the contribution of unscheduled visits on total costs is small and it is unlikely that the difference in costs between treatment groups is largely affected by this bias. Concomitant medication is a more important contributor to total costs. Because investigators were instructed to keep the dose of concomitant medication constant throughout the trial (except in the event of an exacerbation), this study may have obscured changes in the use of concomitant medication. Hence, the costs of concomitant medication were almost the same in both treatment groups and, considering the improved health outcomes in the tiotropium group, may have led to an underestimation of the actual savings by tiotropium.

The sensitivity analysis showed that the difference in costs between tiotropium and ipratropium was most sensitive to the costs of the device by which ipratropium is administrated, varying from € 0.33 per day in the base case analysis based on the MDI price, to € 0.69 per day when the price of ipratropium was based on the weighted average of the MDI (44%) and the DPI (56%). The latter price most accurately reflects the costs of current treatment with ipratropium in the Netherlands. Because several papers have shown similar efficacy of the MDI and DPI (Cuvelier et al., 2002; Gimeno et al., 1988), this sensitivity analysis (SA5) suggests that the cost-effectiveness of tiotropium in daily practice in the Netherlands is probably better than demonstrated in this trial situation. The sensitivity of the results to the choice of the comparator should also be taken into account when reporting on the cost-effectiveness of tiotropium in other countries and is an important issue to consider with regard to the generalisability of cost-effectiveness analyses from one healthcare setting to another.

In other economic evaluations of lung diseases like lung transplantation (van Enckevort et al., 1998) or lung volume reduction surgery (Ramsey et al., 2003) health outcomes have sometimes been measured by a generic quality of life questionnaire that enabled the calculation of QALYs. As such a questionnaire was not administered in our study, we were not able to compare our results with the outcomes of these studies. However, the primary health outcomes that were used in our cost-effectiveness ratios are among the clinical outcome measures most relevant in COPD: exacerbations and quality of life (health status) (Pauwels et al., 2001). For reasons of comparison we have used the SGRQ, because it is the most frequently used questionnaire in COPD. There is one other economic evaluation by Jones et al. who calculated the costs per patient with a four-unit improvement on the SGRQ (Jones et al., 2003). In this study, salmeterol was compared with placebo over 16 weeks in 189 patients with COPD and an incremental cost-effectiveness ratio was found of £ 497 (€ 785). In another economic evaluation, Torrance et al. compared ciprofloxacin with usual antibacterial treatment for acute exacerbations of chronic bronchitis and reported incremental costs per acute exacerbation-day avoided of $ CAN 332 (€ 217) (Torrance et al., 1999). Considering the average duration of an exacerbation (16 days in this present study) this is considerably higher than the € 667 per exacerbation avoided that was found in the current analysis.

The threshold value of four units to identify patients with a minimum clinically relevant improvement on the SGRQ was determined in various studies conducted by the designer of the questionnaire (Jones, 2002) and nearly all studies reporting on the number of patients with a relevant improvement on the SGRQ have used this threshold. However, we have assessed the impact of changing the threshold value of the SGRQ on the cost-effectiveness ratio in the sensitivity analysis. This analysis showed that the cost-effectiveness ratio increased to € 1353 and € 1622 when the threshold value was set to 6 and 8 units respectively, but that the difference in the numbers of patients improving was still statistically significant at the 0.05 level.

Assessing the uncertainty around the cost-effectiveness ratios is especially important because many economic evaluations are piggybacked to clinical studies and sample size calculations are based on clinical rather than economic outcomes. Consequently, due to the large variation in costs between patients, the power of economic evaluations is usually not sufficient to detect statistically significant differences in all economic outcomes. The lack of power in

combination with the difficulties related to the interpretation of a ratio statistic, limit the use of classic statistical approaches commonly applied in clinical studies. It is therefore argued as by Briggs et al. *'the goal of economic evaluation should be the estimation of a parameter – incremental cost-effectiveness – with appropriate representation of uncertainty, rather than hypothesis testing'* (Briggs and O'Brien, 2001). The cost-effectiveness plane and the acceptability curve are two instruments that have been developed to facilitate a visual and straightforward interpretation of the uncertainty around the cost-effectiveness ratios. The cost-effectiveness planes in this study showed that most of the uncertainty around the ratios was associated with the difference in costs between treatment groups. The surface of the ellipses was almost entirely in the upper and lower-right quadrants. Clearly, as long as none of the treatments is dominant (that is when the ellipses fall entirely in the upper-left or lower-right quadrant), the decision whether to accept a new treatment depends on the maximum willingness to pay for a gain in health. The acceptability curves show the probability that tiotropium is acceptable, given this maximum acceptable ratio. In our study, these figures showed that if the willingness to pay equaled zero, the probability that tiotropium is cost effective is about 25%. In other words the probability that tiotropium is cost saving is about 25%. As the maximum acceptable ratio increases, the probability that tiotropium is cost effective increases. As the willingness to pay to avoid one exacerbation or to have one additional patient with a relevant improvement on the SGRQ is set at € 2000, the probability that tiotropium is acceptable is 80% and 72% respectively.

4.5 CONCLUSIONS

Tiotropium resulted in significant reductions of COPD exacerbations and significant improvements in quality of life, lung function and dyspnoea compared to ipratropium. The additional costs to achieve these favourable outcomes were € 180 per patient per year. The higher acquisition costs of tiotropium were offset by 60% through a decrease in other healthcare costs, especially costs of hospitalisations. This is a conservative estimate as tiotropium was compared to the cheapest way of administering ipratropium through the metered dose inhaler.

Acknowledgements
The study was supported by a grant from Boehringer Ingelheim GmbH.

CHAPTER 5

PROBABILISTIC MARKOV MODEL TO ASSESS THE COST-EFFECTIVENESS OF BRONCHODILATOR THERAPY IN COPD PATIENTS IN DIFFERENT COUNTRIES

5.1 INTRODUCTION

COPD is characterized by chronic airflow limitation that is usually slowly progressing and not fully reversible (Pauwels et al., 2003). The major environmental risk factor is smoking. The airflow limitation in COPD is associated with symptoms of chronic cough, sputum production and dyspnoea upon exertion, leading to significant impairments in exercise capacity and quality of life (Pauwels et al., 2003). It has also been shown that a decreased pulmonary function is associated with a higher frequency (Dewan et al., 2000) and more severe exacerbations (Andersson et al., 2002; Rodriguez-Roisin, 2000). In turn, exacerbations have also been shown to contribute to a more rapid decline in lung function (Donaldson et al., 2002), increased morbidity and a reduced quality of life (Seemungal et al., 2000).

The global burden of COPD is expected to increase substantially within the next decades. In developed countries, this increase is strongly associated with ageing of the population and increased use of tobacco in the past (Feenstra et al., 2001; Stang et al., 2000). Recent studies have shown that exacerbations contribute to approximately 35% to 45% of the total costs of COPD treatment in the Netherlands (see chapter 6), Sweden (Andersson et al., 2002) and Spain (Miravitlles et al., 2003), and that exacerbations associated with a hospitalisation contribute to approximately 90% of the total costs of exacerbations (chapter 6). Although COPD is increasingly recognised as a multifaceted disease requiring a multidisciplinary approach, with treatment goals focusing on smoking cessation, improving lung function, increasing exercise capacity, preventing exacerbations, and optimising nutrition, the cornerstone of COPD treatment remains bronchodilation and adequate treatment of exacerbations.

Recently, the inhaled anticholinergic bronchodilator tiotropium has been approved by health authorities and became available in many countries. With duration of action of at least 24 hours, tiotropium is dosed once daily. In a series of clinical trials tiotropium has been compared to 1) ipratropium, an anticholinergic with a recommended dosing of 4 times daily; 2) salmeterol, a long-acting inhaled beta-agonist with a recommended dosing of 2 times daily; and 3) placebo. Tiotropium has been shown to provide sustained bronchodilation, improvements in dyspnoea and health-related quality of life assessed with a disease-specific instrument and is associated with fewer exacerbations than ipratropium and placebo (Casaburi et al., 2002; Vincken et al., 2002). Improvements in lung function with tiotropium have been shown to be significantly better than with salmeterol (Brusasco et al., 2003). All clinical trials were multi-center studies conducted in a total of 19 countries.

In order to assess the cost-effectiveness of tiotropium in individual countries we developed a Markov model that integrates patient-level data from the aforementioned clinical trials. This probabilistic model allows to fully explore the uncertainty around the cost-effectiveness estimate by applying distributions to model parameters. The model is especially designed to be populated with country-specific treatment patterns and unit cost. This was done to inform local reimbursement authorities about the cost-effectiveness of tiotropium. In this paper, the results for the Netherlands and Canada are presented.

5.2 METHODS

The tiotropium trials | The safety and efficacy of tiotropium was studied in a number of multi-centre, randomised double-blind, double dummy, parallel group trials comparing tiotropium (18 µg once-daily) with either ipratropium (40 µg four times daily), salmeterol (50 µg twice daily) or placebo. The two ipratropium-controlled studies were conducted in the Netherlands and Belgium, the two salmeterol-controlled studies in North America, Australia, Europe and South Africa and the two placebo-controlled studies in the UK. All studies were conducted in patients with COPD who were required to be relatively stable, to have moderate to severe airflow obstruction with a $FEV_1 \leq 65\%$ (salmeterol-controlled trials $\leq 60\%$) of predicted normal and an $FEV_1 \leq 70\%$ of FVC. The duration of the ipratropium- and placebo-controlled trials was one year whereas the duration of the salmeterol-controlled trials was six months. Details of the trials including the results on lung function parameters, quality of life and dyspnoea as well as exacerbations and hospitalisations have been published elsewhere (Brusasco et al., 2003; Casaburi et al., 2002; Vincken et al., 2002).

Model structure | The Markov model was structured around disease states and exacerbations based on patient-level data derived from the trials described above (*figure 5.1*). Patients were classified into disease severity states based on the pulmonary function as measured by pre-bronchodilator FEV_1 as % of predicted normal, using the same severity classification as the updated GOLD criteria (2003): moderate COPD ($50\% < FEV_1$ % pred. $< 80\%$), severe COPD ($30\% < FEV_1$ % pred. $\leq 50\%$) and very severe COPD (FEV_1 % pred. $< 30\%$) (Pauwels et al., 2003). At the start of the model simulation, 25% of the patients were assumed to have moderate disease, 50% severe disease and 25% very severe disease. Because patients with mild COPD (FEV_1 % pred. $> 80\%$) were excluded from tiotropium's clinical trial program, this disease state was not included in the model. In addition, a death state was not included in this one-year model, because of the small numbers of deaths within the trial periods (2.3%, 1.1% and 0.2% in the ipratropium-, salmeterol- and placebo-controlled studies respectively). During each Markov cycle, patients in each treatment group were assigned a probability of transitioning from one disease state to another. Depending on treatment group and disease state, patients were also assigned a probability to experience a severe or non-severe exacerbation. In each clinical trial, an exacerbation was defined as a complex of respiratory symptoms (i.e. new onset or worsening of more than one symptom such as cough, sputum, dyspnoea or wheeze) lasting for at least three days. Exacerbation severity was based on physicians' assessments of the intensity of an adverse event, a classification that was used in all trials. Non-severe exacerbations were defined as 'an awareness of a sign or symptom which was easily tolerated' (mild intensity) or as 'discomfort enough to cause interference with usual activity' (moderate intensity). A severe exacerbation was defined as 'incapacitating or inability to do work or usual activity'.

In order to model the improvement in pulmonary function that was observed in all treatment groups during the first few days of the clinical trial, the length of the first cycle was set at eight days. The length of the subsequent cycles was one month. A period of one month was chosen to incorporate the full effect of an exacerbation in terms of resource use and quality

of life on the one hand, and to minimise the risk of experiencing more than one exacerbation during the same cycle (which was not possible in the model) on the other hand. Transitions between states were assumed to take place halfway through the cycle. The time span of the model was one year.

In the current analysis we compared tiotropium, salmeterol and ipratropium. Although data from the placebo-controlled trials were used to derive transition and exacerbation probabilities for the tiotropium arm, the placebo arm was not included in the model. This was done because the placebo arm in the clinical trials does not reflect usual care in practice, since concomitant use of anticholinergics and long-acting beta-agonists was not permitted. In total, data in our model were based on 1296 patients treated with tiotropium, 405 patients treated with salmeterol and 175 patients treated with ipratropium.

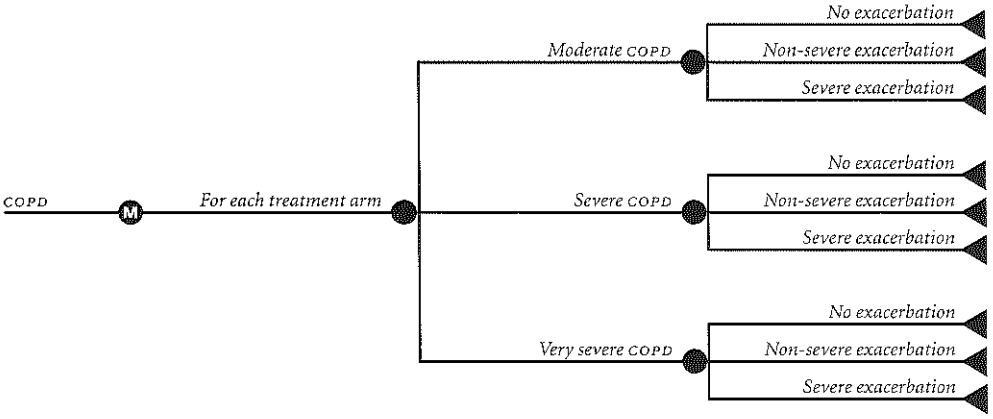


Figure 5.1: Graphical presentation of the Markov model.

Transitions between disease states | At each scheduled follow-up visit during the trial all patients were assigned to one of the three disease states, based on their pre-bronchodilation FEV_1 values (=trough FEV_1; the primary endpoint of the clinical trials). In all trials, we observed an initial improvement in pulmonary function between treatment initiation and the first follow-up visit. Subsequent to this initial improvement, the change in pulmonary function over time in each treatment group remained fairly constant. Consequently, for all treatment groups in all trials, two matrices of transition probabilities were determined. One transition matrix for the first cycle based on the difference between the frequency distributions of disease states at baseline and the first follow-up visit. A second transition matrix for the subsequent cycles, which is based on the difference between the frequency distributions of disease states at the first and last follow-up visit of the trials. These transition matrices were converted to identical time periods of eight days (first cycle) and one month (subsequent cycles) using a Taylor series expansion. In the base case analysis, transition probabilities for the tiotropium arm were estimated using the combined data from all six clinical trials. Transition probabilities for the comparator arms were then calculated, based on the relative difference to tiotropium as found in the individual trials. For example, the ipratropium-controlled trials showed that the probability to move from moderate to severe COPD was 2.7 times

greater for ipratropium than for tiotropium. Hence, the pooled probability to transition from moderate to severe COPD for all tiotropium patients across the three trials was multiplied with 2.7 to obtain the transition probability for ipratropium. The resulting sets of transition parameters are presented in *table 5.1*. In a final step, a Dirichlet distribution (Briggs et al., 2003) was assigned to these input parameters and a Monte Carlo simulation was performed in which values were randomly drawn from these distributions. Further details are provided in the appendix and the section Analysis below. To extend the time span of the model to one year, it was assumed that the monthly transitions for the salmeterol-controlled trials also applied to the period beyond the six-month trial period.

Exacerbations | The observed numbers of severe and non-severe exacerbations within each treatment arm and disease state at each trial visit were used to derive exacerbation probabilities. In the base case analysis, exacerbation probabilities in the tiotropium arm were based on combined data from all six trials, whereas the calculation of exacerbation probabilities for the comparator arms were based on the relative difference to tiotropium as found in the individual trials (*table 5.1*). Details of the calculations are provided in the appendix. It was assumed that patients could only have one exacerbation during each cycle. This assumption was justified by the clinical trial data in which the risk of having more than one exacerbation in one month was found to be very small.

Utilities | Utility values per disease state were based on empirical data from an observational study in patients with COPD classified into the GOLD stages (Borg et al., 2004). Mean (SE) EQ-5D (Euroqol questionnaire-5 dimensions) index scores for moderate, severe and very severe COPD were reported to be 0.755 (0.031), 0.748 (0.06) and 0.549 (0.104) respectively. During cycles in which patients experienced an exacerbation, it was assumed that the utility value was reduced by 15% in case of a non-severe exacerbation (Paterson et al., 2000) and by 50% in case of a severe exacerbation (Spencer and Jones, 2003).

Resource use and unit costs | Resource use was assigned to maintenance therapy according to disease severity state and to severe and non-severe exacerbations. With the exception of the acquisition cost for each study medication, the costs of maintenance therapy per disease severity state and the cost per severe and non-severe exacerbation were assumed to be the same for each treatment group.

Resource use for the Netherlands was obtained from an economic evaluation that was piggybacked to the ipratropium-controlled trials. The economic evaluation was based on data from 519 patients of whom 445 (86%) were Dutch. Almost all resource use associated with maintenance therapy and exacerbations could be derived from these data. The details of this economic evaluation and the collection of resource use data and the attached unit costs have been described in the previous chapter. Only those estimates that were likely to be influenced by the trial setting, like regular follow-up visits to a GP or respiratory physician, were derived from the Dutch guidelines for the treatment of COPD patients by general practitioners. Unit costs of (study) medications were based on list prices and included value added

	Disease state transition probabilities						Exacerbation probabilities	
	First cycle (8 days)			Subsequent cycles (per month)			Per month	
From To ⇒ ⇓	Moderate	Severe	Very severe	Moderate	Severe	Very severe	Exacerbation	Severe exacerbation†
Tiotropium								
Moderate	.907 (.018)	.092 (.018)	.001 (.002)	.957 (.010)	.040 (.010)	.003 (.003)	.051 (.004)	.097 (.024)
Severe	.259 (.017)	.716 (.018)	.025 (.006)	.023 (.006)	.954 (.008)	.023 (.006)	.075 (.003)	.136 (.018)
Very severe	.010 (.005)	.341 (.024)	.649 (.024)	.001 (.002)	.045 (.012)	.954 (.012)	.096 (.005)	.192 (.027)
Salmeterol								
Moderate	.900 (.033)	.100 (.033)	.001 (.001)	.928 (.024)	.066 (.024)	.005 (.007)	.057 (.013)	.030 (.031)
Severe	.201 (.028)	.766 (.030)	.033 (.013)	.023 (.010)	.918 (.019)	.059 (.017)	.089 (.011)	.138 (.033)
Very severe	.001 (.001)	.302 (.042)	.698 (.042)	.006 (.008)	.036 (.019)	.958 (.021)	.104 (.016)	.207 (.048)
Ipratropium								
Moderate	.739 (.090)	.257 (.089)	.004 (.013)	.923 (.049)	.073 (.048)	.003 (.011)	.080 (.020)	.267 (.114)
Severe	.102 (.028)	.842 (.034)	.056 (.022)	.013 (.011)	.950 (.021)	.037 (.018)	.097 (.013)	.188 (.041)
Very severe	.005 (.011)	.220 (.065)	.775 (.066)	.003 (.009)	.025 (.026)	.972 (.027)	.102 (.022)	.186 (.062)

Table 5.1: Mean (SE) disease state transition and exacerbation probabilities.*

** Parameters of the distributions are the numbers of patients transitioning between disease states and the number of months with/without exacerbations. For ease of interpretation this table presents the expected values (probabilities) and the associated SEs rather than these numbers. Probabilities can be recalculated into numbers using method-of-moments fitting (Briggs et al., 2002). For the transitions between disease states Dirichlet distributions are assigned, for the risk of exacerbations, beta-distributions are used; † Given an exacerbation, probability that it is severe.*

taxes and a mark-up of € 6.02 per prescription to cover pharmacist fees. The price of ipratropium was based on the average price of the MDI and the DPI, weighted by the actual use of these administration devices in the Netherlands in 2001 (44% MDI versus 56% DPI).

COPD-related resource utilisation for Canada was collected in a prospective, Canada wide, 52-week, multi-centre, observational study of patients with similar inclusion criteria as the tiotropium trials, conducted between June 2001 and September 2002. Approximately 76% (42) of the centres were run by GPs or family practitioners and 24% (13) were run by respirologists. In total, resource utilisation data were collected from 598 patients. For the use of maintenance medication, it was assumed that medications patients were taking at baseline were continued throughout the year with a compliance rate of 80%. Since, the study only recorded data regarding physician visits in the case of exacerbations, regular physician visits were collected from a survey among 69 physicians (both specialists and GPs) conducted in 2002. Because the definition of exacerbation severity as used in the tiotropium clinical trials was not feasible for an observational study in a naturalistic setting, all exacerbations associated with a hospital admission or emergency room visit were classified as severe. Unit costs of (study) medications were derived from the Ontario Drug Benefit Formulary (Comparative Drug Index) and included the authorised mark-up (10%) and the pharmacist current dispensing fees (€ 3.97, CAN $ 6.47). Minimum patient co-payment was subtracted from the cost of a 30-day-claim. Because the DPI device is not available in Canada, the price of ipratropium was based on the MDI price.

The resulting resource use and unit costs for Canada and the Netherlands are presented in *table 5.2*. Costs were determined from a healthcare perspective and included all costs relevant to the Dutch and Canadian healthcare budgets. All costs are reported in 2001 values expressed in euro (1 Canadian dollar = 0.62 euro (exchange rate April 2004)). Discounting was not applied in this one-year model.

Analysis | In order to assess the uncertainty around the point estimates of costs, effects and cost-effectiveness, the model was designed probabilistically adopting the methodology proposed by Briggs et al. (Briggs, 2000; Briggs et al., 2002). Uncertainty around transitions between disease states, exacerbations, utilities and resource use was considered simultaneously and input parameters were entered into the model as pre-specified distributions. As proposed by Price and Briggs, we adopted a Dirichlet distribution for transitions between disease states (Briggs et al., 2003), beta distributions for exacerbations and utilities and a gamma distribution for the estimation of resource use (Price and Briggs, 2002). Second-order Monte-Carlo simulations were undertaken in which values were randomly drawn from these distributions. The current analysis was based on 5000 simulations. Main outcomes of the model were the mean, SE (being the SD across the simulations) and 95% uncertainty intervals (UI) of the costs per patient per year, the number of exacerbations and quality adjusted life months. The presentation of paired differences between treatment groups (tiotropium versus salmeterol and salmeterol versus ipratropium) is based on the hierarchy of observed outcomes. The uncertainty around costs and effects was further explored by means of incremental CE planes and separate CE acceptability curves per treatment based on the net benefit approach (Fenwick et al., 2001).

Sensitivity analysis | In a first set of sensitivity analyses, the robustness of the model for alternative transition and exacerbation probabilities were investigated. In these sensitivity analyses (SA1 to SA3), transitions and exacerbations for tiotropium were based on observed values from the ipratropium-controlled (SA1), salmeterol-controlled (SA2) and placebo-controlled trials (SA3) separately, instead of the combined data from all six trials. Transitions and exacerbations for the comparator arms were calculated based on the relative difference to tiotropium as found in the individual trials. In two additional sensitivity analyses (SA4 and SA5), we investigated the separate contribution of exacerbation probabilities and disease state transitions to the outcomes of the model by assuming that either transition probabilities (SA4) or exacerbation probabilities (SA5) did not differ between treatment groups. In a third set of sensitivity analyses (SA6 to SA8), we varied the baseline distribution of patients over disease states. In these three analyses, it was assumed that at baseline, 100% of the patients were suffering from moderate disease (SA6), severe disease (SA7) and very severe disease (SA8), respectively. In sensitivity analysis 9, alternative utility values were assigned to disease states: moderate COPD 0.81 (SE: 0.02), severe COPD 0.72 (0.03) and very severe COPD 0.67 (0.05) (Kind et al., 2002). Because transitions in the model were derived from patient-level data in trials that included only stable patients who were not using long-term oxygen at study entry, we did not include the costs of oxygen therapy in our base case analysis. The effect of adding estimates of oxygen use (table 5.2) to the costs of maintenance therapy was investigated in a final sensitivity analysis (SA10).

Validation of the model | To validate the exacerbation outcomes of the model we compared these outcomes with the empirical analyses of the trial data. To perform this validation, the model was populated with the trial-specific baseline distribution of patients over disease states, and trial-specific transition and exacerbation probabilities. This validation can only be performed pair-wise, as each trial compared only two treatments. For the comparison between tiotropium and salmeterol the time span of the model was set to six months to match the duration of the trial. The exacerbation rates obtained with the model were compared with the empirical exacerbation rate based on an analysis in which multiple imputation was used to account for incomplete data due to dropout (Rubin and Schenker, 1991). For the ipratropium-controlled trials, these rates have been presented in the previous chapter, and for the salmeterol-controlled trials, these rates are reported for the first time in this paper.

5.3 RESULTS

Health outcomes | *Table 5.3* summarises the main outcomes of the Markov model. Mean (SE) estimates of the number of exacerbations varied from 0.85 (0.03) in the tiotropium group to 1.02 (0.10) in the salmeterol group and 1.14 (0.13) in the ipratropium group. The difference (95% UI) between tiotropium and salmeterol was 0.17 (- 0.02; 0.37) and the difference between salmeterol and ipratropium was 0.12 (- 0.17; 0.44). The proportion of exacerbations that was severe varied from approximately 15% in the tiotropium and salmeterol group to 20% in the ipratropium group. Differences in quality adjusted life months were small and

	The Netherlands				Canada			
Resource use of disease states (per annum)	Unit cost	Moderate	Severe	Very severe	Unit cost	Moderate	Severe	Very severe
Outpatient visit RP	48.00	—	2.00 (0.16)	4.00 (0.32)	44.30	—	0.80 (0.10)	2.80 (0.35)
Outpatient visit GP	16.00	2.00 (0.16)	—	—	16.37	1.00 (0.07)	1.20 (0.11)	1.20 (0.26)
Spirometry	14.00	2.00 (0.16)	2.00 (0.16)	4.00 (0.32)	10.08	1.00 (0.07)	1.00 (0.06)	1.00 (0.09)
Influenza vaccination	15.00	0.75 (0.09)	0.75 (0.09)	0.75 (0.09)	17.03	0.80 (0.07)	0.80 (0.06)	0.80 (0.09)
Beta-adrenergics†	0.43	30.85 (10.18)	27.95 (5.37)	13.55 (5.63)	0.32	69.14 (8.94)	50.11 (6.67)	261.57 (8.93)
Theophylline†	0.30	20.27 (7.50)	51.39 (7.52)	61.57 (11.49)	0.37	16.17 (4.79)	46.09 (6.46))	64.80 (11.38)
Inhaled steroids†	1.04	256.20 (22.34)	289.57 (32.42)	281.11 (35.12)	1.15	175.90 (10.45)	187.61 (8.11)	199.53 (12.96)
Other medications†	0.65	79.78 (17.06)	106.24 (17.00)	133.37 (18.83)	0.67	86.80 (9.62)	139.34 (8.91)	145.08 (13.83)
Oxygen therapy†§	12.64	—	14.60 (2.76)	73 (12.52)	18.61	2.96 (2.10)	27.21 (5.14)	67.31 (11.54)
Resource use of exacerbations (per exacerbation)		Unit cost	Non-severe	Severe		Unit cost	Non-severe	Severe
ICU-days†		1113.00	—	0.86 (0.67)		633.00	—	0.38 (0.13)
Non-ICU days†		223.00	1.01 (0.21)	11.08 (1.32)		427.87	—	5.85 (0.47)
Emergency room visits		91.00	0.03 (0.01)	0.25 (0.07)		80.47	—	0.95 (0.02)
Outpatient visit RP		77.00	0.34 (0.03)	0.82 (0.22)		44.30	0.14 (0.04)	0.05 (0.02)
Outpatient visit GP		21.00	0.66 (0.06)	0.70 (0.16)		16.37	0.88 (0.04)	0.12 (0.03)
Visit other HCP		18.00	0.27 (0.14)	0.50 (0.39)		14.64	0.02 (0.01)	0.18 (0.03)
Antibiotics†		2.68/7.07‡	7.94 (2.00)	11.75 (2.20)		2.69	5.84 (0.44)	4.86 (0.35)
Systemic steroids†		0.91	7.94 (0.94)	24.08 (4.33)		0.58	3.4 (0.24)	5.18 (0.39)
Other medications†		0.90	15.83 (2.48)	55.02 (12.57)		0.50	0.50 (0.08)	0.84 (0.16)
Oxygen†		12.64	—	0.21 (0.02)		18.61	0.17 (0.07)	3.62 (0.41)
Costs of study medication		Tiotropium	Ipratropium	Salmeterol		Tiotropium	Ipratropium	Salmeterol
Unit cost per day		1.57	0.69	1.24		1.51	0.54	1.21

Table 5.2: Mean (SE) resource use of disease states and exacerbations by country with the associated unit costs (in 2001 N).*

*RP: respiratory physician; HCP: healthcare provider; * Gamma distributions were assigned to model the uncertainty around the resource use estimates; † resource use expressed in days, unit cost represent the costs per day; ‡ unit costs of antibiotics in non-severe exacerbations € 2.68 and in severe exacerbations € 7.07 per day; § Costs of oxygen therapy not included in the base case analysis, but only in sensitivity analysis 10.*

associated with wide UIs. The mean (SE) number of quality adjusted life months varied from 8.42 (0.40) in the tiotropium group to 8.17 (0.46) in the salmeterol group and 8.11 (0.49) in the ipratropium group.

Costs | Estimates of the mean (95% UI) one-year cost per patient in the Netherlands varied from € 1760 (1563; 2011) in the tiotropium group to € 1802 (1515; 2195) in the salmeterol group and € 1930 (1503; 2525) in the ipratropium group. The corresponding estimates of the mean total costs for Canada were considerably lower and varied from € 1309 (1222; 1408) for tiotropium, € 1306 (1142; 1516) for salmeterol and € 1307 (1050; 1637) for ipratropium. The costs of the study medication tiotropium made up approximately 33% of the total costs in the Netherlands and 42% of the total costs in Canada. Salmeterol accounted for 25% and 24% of total costs in the Netherlands and Canada respectively, whereas the costs of ipratropium contributed to approximately 14% of total costs in both countries. The largest difference in costs between the two countries was observed in the costs of exacerbations. In the Netherlands these costs varied from € 670 (38% of the total costs) in the tiotropium group to € 812 (45%) in the salmeterol group and € 1131 (59%) in the ipratropium group. The corresponding estimates for Canada were € 387 (30%), 466 (36%) and € 712 (54%) respectively.

	Tiotropium		Salmeterol		Ipratropium	
Number of Exacerbations						
Non-severe	0.73 (0.68; 0.78)		0.87 (0.71; 1.05)		0.91 (0.72; 1.14)	
Severe	0.12 (0.10; 0.15)		0.15 (0.09; 0.21)		0.23 (0.15; 0.33)	
Total	0.85 (0.80; 0.91)		1.02 (0.84; 1.22)		1.14 (0.92; 1.40)	
Quality adjusted life months	8.42 (7.59; 9.20)		8.17 (7.24; 9.06)		8.11 (7.08; 9.04)	
Mean (SE) costs (in 2001 €)	Netherlands	Canada	Netherlands	Canada	Netherlands	Canada
Exacerbations						
• Hospitalizations	583 (112)	340 (43)	707 (166)	410 (88)	1004 (257)	646 (144)
• Other exa-related costs	87 (7)	47 (3)	105 (13)	56 (6)	127 (17)	66 (8)
Subtotal exacerbations	670 (113)	387 (43)	812 (173)	466 (92)	1131 (266)	712 (144)
Maintenance therapy	517 (22)	371 (16)	537 (23)	398 (19)	547 (25)	398 (23)
Study medication	573 (- -)	551 (- -)	453 (- -)	442 (- -)	252 (- -)	197 (- -)
Total costs	1760 (116)	1309 (47)	1802 (175)	1306 (96)	1930 (267)	1307 (150)
95% UI of total cost	1563; 2011	1222; 1408	1515; 2195	1142; 1516	1503; 2525	1050; 1637
			Tiotropium versus salmeterol		Salmeterol versus ipratropium	
Difference in			Netherlands	Canada	Netherlands	Canada
Costs (in 2001 €)			-42	3	-128	-1
			(-484; 353)	(-227; 203)	(-795; 457)	(-376; 323)
Exacerbations avoided			0.17 (-0.02; 0.37)		0.12 (-0.17; 0.44)	
Quality adjusted life months			0.25 (-0.90; 1.47)		0.06 (-1.26; 1.42)	

Table 5.3: Results of the Markov simulation.
Mean and 95% UI unless otherwise stated.

Cost-effectiveness | *Figure 5.2* presents the uncertainty around the costs and effects on the CE plane. In order to reduce the number of figures, only the CE planes comparing tiotropium to salmeterol are presented. Each dot represents one of the 5000 model simulations. The CE planes show that the uncertainty about costs in the Netherlands was somewhat larger than in Canada. Dots were almost evenly distributed over the upper- and lower quadrants, showing the near cost neutrality between tiotropium and salmeterol. The CE planes also show that there were no substantial differences between treatment groups with regard to quality adjusted life months. In contrast, the difference in exacerbations clearly was in favour of tiotropium. The proportion of iterations in the right quadrants for this outcome was approximately 95%.

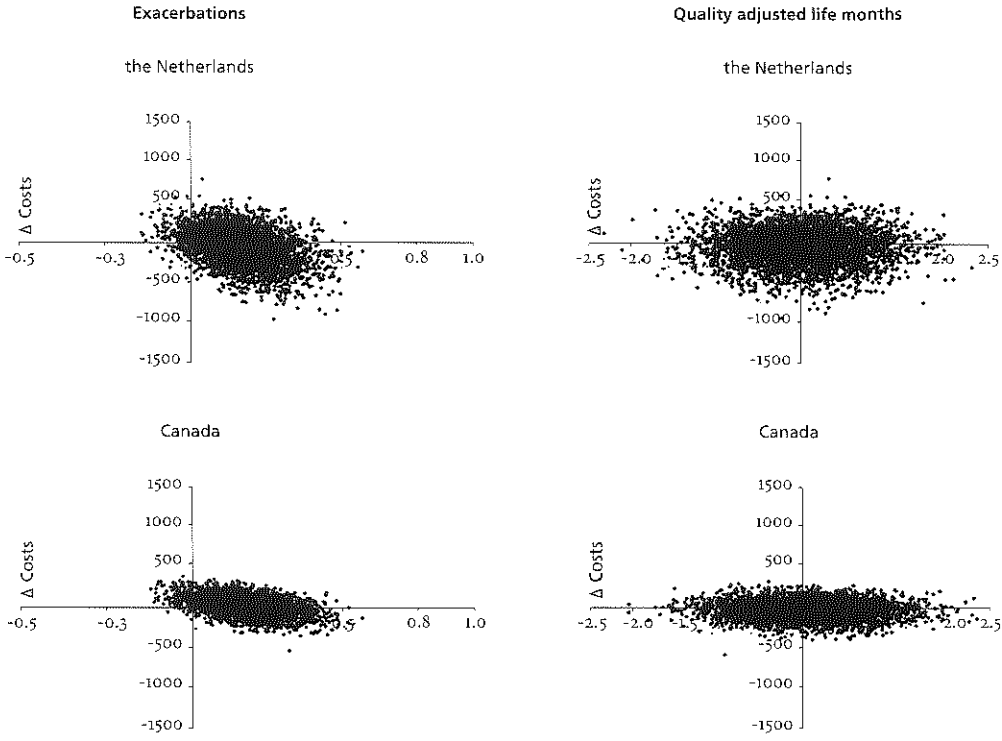


Figure 5.2: Cost-effectiveness planes of the difference in costs and effects of tiotropium versus salmeterol. Each dot represents 1 of 5000 model simulations; the horizontal axis represents the difference in the number of exacerbations and quality adjusted life months respectively.

In the Netherlands, the acceptability curves (*figure 5.3*) representing quality adjusted life months show that the probability tiotropium is cost effective was almost independent of the value of the ceiling ratio, reflecting the small differences for this outcome in the one-year model. A much larger impact of the value of the ceiling ratio was observed in the acceptability curves regarding exacerbations avoided. The probability for tiotropium to be cost effective in the Netherlands gradually increased from 43% when the ceiling ratio was set to € 0 to 60% when the ceiling ratio was set to € 500. In Canada, tiotropium had the highest probability of

being cost effective when the ceiling ratio for avoiding an exacerbation was above € 160 and the ceiling ratio for gaining one quality adjusted life month was at least € 120. For lower values, ipratropium had the highest probability of being cost effective but was not associated with the highest expected net benefit (Fenwick et al., 2001). Salmeterol had the highest expected net benefit for values of the ceiling ratio until approximately € 10, while tiotropium had the highest expected net benefit for all values higher than € 10. Hence, the CE acceptability frontier (not shown), followed the salmeterol curve for values of the ceiling ratio below € 10 and followed the tiotropium curve for all values higher than € 10, signifying that in Canada, tiotropium was the preferred treatment in terms of cost per exacerbation avoided, except for ceiling ratios below € 10. For an explanation about acceptability curves and the acceptability frontier in case of multiple treatments and skewed distributions we refer to Fenwick et al. (Fenwick et al., 2001).

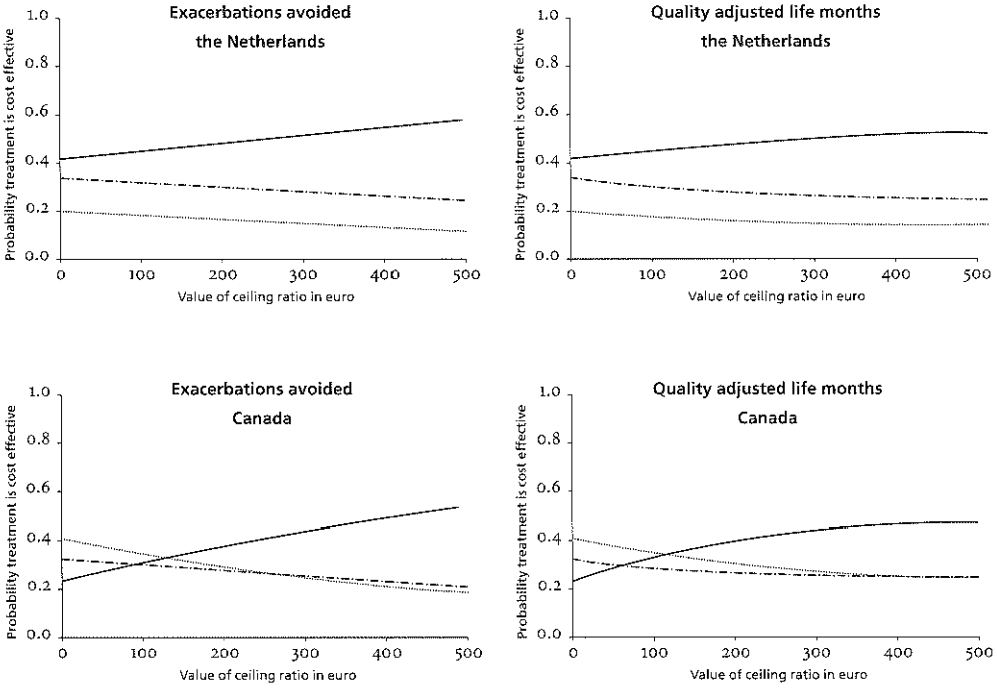


Figure 5.3: CE acceptability curves of exacerbations avoided and quality adjusted life months for the Netherlands and Canada. ——— : tiotropium; – – – – : salmeterol; ..... : ipratropium. In the Netherlands the cost-effectiveness acceptability frontier follows the tiotropium curves. In Canada, the cost-effectiveness acceptability frontier follows the salmeterol curves for values of the ceiling ratio below approximately € 10. For all values higher than € 10 the frontier follows the tiotropium curves, signifying the higher expected net benefit (Fenwick et al., 2001).

Sensitivity analysis | The results of the sensitivity analyses are presented in *table 5.4*. For each sensitivity analysis, the table shows the values of the ceiling ratio at which each treatment has the maximum expected net benefit. Tiotropium showed the maximum expected net benefit for all values of the ceiling ratio when probabilities were set relative to the salmeterol-controlled trials (SA2), while the value of the ceiling ratio for which tiotropium was asso-

ciated with the maximum expected net benefit increased when probabilities were set relative to the ipratropium-controlled (SA1) and placebo-controlled trials (SA3). Applying similar transition probabilities to treatment groups (SA4) had a large impact on quality adjusted life months. In the Netherlands, salmeterol was associated with maximum expected net benefit for values below € 1080, while in Canada ipratropium had maximum expected net benefit for values below € 1180. The impact on exacerbations was much less and tiotropium was associated with maximum expected net benefit for all values of the ceiling ratio above € 180 in the Netherlands and above € 400 in Canada. Sensitivity analysis 5 shows that for exacerbations ipratropium was associated with the maximum expected net benefit for all values of the ceiling ratio below € 8500 in the Netherlands and below € 11000 in Canada. In terms of quality adjusted life months these values were € 1120 and € 1340 respectively. Changing the division of patients over disease severity states (SA6 to 8) showed divergent results. When all patients had moderate disease at baseline, salmeterol showed the maximum expected net benefit in the Netherlands for values of the ceiling ratio below € 440, whereas ipratropium had the highest expected net benefit for low values of the ceiling ratio for both outcomes when all patients had very severe disease at baseline. Alternative utility values (SA9) did not change the cost-effectiveness frontier as compared to the base case analysis, while adding the costs of oxygen therapy to the costs of maintenance therapy (SA10) favoured tiotropium in both settings.

Validation of the model | When the model was populated with the baseline distribution of patients over disease states, and exacerbation and transition probabilities from the ipratropium-controlled trials, the mean (SE) numbers of exacerbations were exactly the same as observed in the trials and as presented in the previous chapter (tiotropium 0.74 (0.08), ipratropium 1.01 (0.12)). Only the SE of ipratropium was slightly larger (model 0.12, trials 0.10). A comparison with the salmeterol-controlled trials based on a 6-month period showed that the mean estimates of the model (tiotropium 0.69 (0.07); salmeterol (0.82 (0.09)) were somewhat higher than observed in the trials (tiotropium 0.63 (0.06); salmeterol 0.76 (0.06), but that the estimated difference between treatment groups was exactly the same, 0.13 (0.08).

5.4 DISCUSSION

In this study, we showed how patient-level clinical data could be used to construct a one-year model to compare the cost-effectiveness of three bronchodilators in different countries. The model demonstrated that tiotropium was associated with a reduction of 17% in the number of exacerbations when compared to salmeterol. When salmeterol was compared to ipratropium, the number of exacerbations was reduced by 11%. No substantial differences in quality-adjusted life were found between treatment groups in this one-year model. Overall, costs were considerably higher in the Netherlands than in Canada, mainly because of the higher costs associated with exacerbations. In the Netherlands, tiotropium was associated with small cost-reductions, while in Canada costs were almost the same in all treatment groups. The CE acceptability frontier of exacerbations showed that tiotropium was associated with the maximum expected net benefit for all values of the ceiling ratio above € 0 (the Netherlands) and € 10 (Canada) in the base case analysis.

	Exacerbations avoided		Quality adjusted life months	
	the Netherlands	Canada	the Netherlands	Canada
Base case	Tiotropium : ≥ 0	Salmeterol : 0–9 Tiotropium : ≥ 10	Tiotropium : ≥ 0	Salmeterol : 0–9 Tiotropium : ≥ 10
SA 1	Salmeterol : 0–59 Tiotropium : ≥ 60	Ipratropium: 0–539 Tiotropium : ≥ 540	Salmeterol : 0–44 Tiotropium : ≥ 45	Ipratropium: 0–499 Tiotropium : ≥ 500
SA 2	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0
SA 3	Salmeterol : 0–49 Tiotropium : ≥ 50	Salmeterol : 0–219 Tiotropium : ≥ 220	Salmeterol : 0–24 Tiotropium : ≥ 25	Salmeterol : 0–119 Tiotropium : ≥ 120
SA 4	Salmeterol : 0–179 Tiotropium : ≥ 180	Ipratropium: 0–379 Salmeterol : 380–399 Tiotropium : ≥ 400	Salmeterol : 0–1079 Tiotropium : ≥ 1080	Ipratropium: 0–1179 Salmeterol : 1180-2719 Tiotropium : ≥ 2720
SA 5	Ipratropium: 0–8499 Tiotropium : ≥ 8500	Ipratropium: 0–10999 Tiotropium : ≥ 11000	Ipratropium: 0–1119 Tiotropium : ≥ 1120	Ipratropium: 0–1339 Tiotropium : ≥ 1340
SA 6	Salmeterol : 0–439 Tiotropium : ≥ 440	Tiotropium : ≥ 0	Salmeterol : 0–399 Tiotropium : ≥ 400	Tiotropium : ≥ 0
SA 7	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0
SA 8	Ipratropium: 0–339 Tiotropium : ≥ 340	Ipratropium: 0–779 Tiotropium : ≥ 780	Ipratropium: 0–159 Tiotropium : ≥ 160	Ipratropium: 0–359 Tiotropium : ≥ 360
SA 9	Tiotropium : ≥ 0	Salmeterol : 0–9 Tiotropium : ≥ 10	Tiotropium : ≥ 0	Salmeterol : 0–9 Tiotropium : ≥ 10
SA 10	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0	Tiotropium : ≥ 0

Table 5.4: Results of the sensitivity analyses showing the values of the ceiling ratio at which each treatment has the maximum expected net benefit (i.e. description of the cost-effectiviness frontiers; in 2001 €). SA 1 = transition and exacerbation probabilities relative to those observed in the trials comparing tiotropium to ipratropium; SA 2: transition and exacerbation probabilities relative to those observed in the trials comparing tiotropium to salmeterol; SA 3: transition and exacerbation probabilities relative to those observed in the trials comparing tiotropium to placebo; SA 4: similar transition probabilities in all treatment groups (probabilities equal those of ipratropium in the base case analysis); SA 5: similar exacerbation probabilities in all treatment groups (probabilities equal those of ipratropium in the base case analysis); SA 6 to SA8: 100% of the patients at baseline in moderate, severe and very severe disease respectively; SA 9: alternative utility weights per disease state, mean (SE): moderate 0.81 (0.02), severe 0.72 (0.03), very severe 0.67 (0.05) (Kind et al., 2002); SA 10: use of oxygen therapy added to the costs of maintenance therapy.

Our model was specifically developed to facilitate the process of adaptation of pharmacoeconomic data to the local setting. Indeed, the difference in results between the Netherlands and Canada reflect differences in treatment patterns between these countries. The finding that tiotropium was somewhat more cost effective in the Netherlands than in Canada was largely driven by the observation that the hospitalisation cost per exacerbation in the Netherlands were approximately 25% higher than in Canada, as a result of longer length of stay. Hence, a reduction in the number of exacerbation-related hospital admissions in patients treated with tiotropium leads to considerably higher savings in the Netherlands than in Canada, especially when considering the fact that the daily acquisition cost of tiotropium hardly differ between the two countries. Other differences between the two countries were the lower use of antibiotics and systemic steroids during exacerbations and the higher number of emergency room visits in Canada. The higher costs of maintenance therapy in the Netherlands are largely due to the higher use of inhaled steroids in all disease severity states.

An important characteristic of the model is that all model inputs related to the effectiveness of treatment are based on patient-level trial data of the tiotropium clinical trial program. This minimises the impact of different inclusion and exclusion criteria as well as trial design features commonly complicating across study comparisons. In addition, this approach offers the possibility to test the internal consistency of the model by comparing the model outcomes with the results of the clinical trials. It was shown that the model closely resembled the difference in the numbers of exacerbations that were observed in the original trials. The ability to compare the outcomes of the model with the original trial data makes the model transparent and may thereby increase the acceptance of this model by local reimbursement authorities.

Limitations of this approach are that other available data about the efficacy of the existing treatments, ipratropium and salmeterol, are not considered and that the outcomes of the model are based on the specific design of the tiotropium studies. More research is needed to formally integrate all available evidence on the effects of bronchodilator therapy in COPD on exacerbations in meta-analyses, and then modelling their cost-effectiveness to provide further information for medical decision-making. However, this is complicated by the lack of a uniform definition of exacerbations (Pauwels et al., 2004).

In accordance with clinical guidelines, lung function parameters were used to classify patients into disease severity states (Pauwels et al., 2003). Two recently published models in COPD also used this disease classification (Borg et al., 2004; Sin et al., 2004). Because COPD is increasingly recognised as being a multifaceted disease that not only impairs lung function, but also has systematic consequences, in the future, it may become common practice to determine disease severity based on a combination of variables. For instance, Celli et al. constructed a multidimensional grading system, based on FEV_1, six-min walk test result, dyspnoea assessment, and body mass index (Celli et al., 2004). They showed that this so-called BODE index is better than FEV_1 at predicting mortality. Because this index was only recently proposed, the classification has not been used in the clinical trials underlying the current model.

In the GOLD guidelines, the classification into disease states is based on post-bronchodilator FEV_1, whereas in the model disease classification was based on pre-bronchodilator values (trough FEV_1) because post-bronchodilator values at baseline (i.e. before the start of study medication) were not available. The trials showed that other lung function parameters like peak FEV_1 and the area under the curve of 0-3 hours post-bronchodilation as well as the forced vital capacity were also consistently better in the tiotropium group (Brusasco et al., 2003; Vincken et al., 2002). Because we have used pre-bronchodilator values for all treatments, where lung function measurements were done 24 hours after the last dose of tiotropium, 12 hours after the last dose of salmeterol and six hours after the last dose of ipratropium (i.e. at the end of each recommended dosing interval), there is no reason to belief that the use of pre-bronchodilator values has favoured tiotropium over other treatments. In addition, sensitivity analyses 4 and 5 have shown that exacerbation probabilities were the main driver of the cost-effectiveness in terms of exacerbations avoided and that the impact of differences in disease state transitions in this one-year model was limited.

Our model is a short-term model that is not intended to reflect the lifetime disease progression of COPD. This explains why we did not model the impact of mortality and smoking.

In this study, outcomes were measured in terms of exacerbations avoided and quality adjusted life months (EQ-5D). The outcome parameter exacerbations closely reflects the primary aim of the currently available medical treatment options in COPD, which is relieving symptoms and preventing exacerbations of the disease (Pauwels et al., 2003). Together with FEV_1 it is the outcome measure most often used in clinical studies in COPD of this duration. Other economic evaluations in asthma and COPD also adopted exacerbations or related outcome measures like exacerbation-free months or symptom-free days, as a primary outcome measure (Sculpher and Price, 2003; Torrance et al., 1999). The EQ-5D was not administered in the tiotropium trials and utilities according to disease states and exacerbations had to be derived from literature. Data about utility values in patients with COPD was limited and we only found two studies reporting EQ-5D values according to disease severity (Borg et al., 2004; Kind et al., 2002). These data and other studies reporting on disease-specific quality of life suggest that the relationship between disease severity and quality of life is not very strong and that reductions in quality of life become most apparent in patients with an FEV_1 % predicted below 50%. Until to date, pharmaceutical treatments have not been able to demonstrate an important effect on quality adjusted life years in patients with COPD and hence, we were not surprised to find no effect in this one-year model either. Varying the utility weights (SA9) did not change these findings.

As there was no head-to-head comparison of all three treatments in the same trial, data from the trials were combined. This was facilitated by the similar study protocols of all tiotropium trials. For probabilities in the tiotropium arm we simply pooled the trials together. To obtain the data for ipratropium and salmeterol, we applied the relative difference between tiotropium and the two comparators as observed in the individual trials to the pooled data of tiotropium. There are other options, such as taking the absolute difference or simply using the transition and exacerbation probabilities as they were observed in the ipratropium and salmeterol arms of the trials. The latter option was rejected, because of differences in the exacerbation rate between the trials. When probabilities were based on the salmeterol-controlled trials (SA2) the number of exacerbations was almost twice as high as in the analysis in which probabilities were based on the ipratropium-controlled trials (SA1). Additional analyses showed that this difference in exacerbation rates between the trials was not related to patient-characteristics or the difference in the duration of the trial. Hence, using the combined probabilities for tiotropium and using the relative difference of tiotropium to the other treatments most accurately reflects the differences between treatments that were actually observed.

5.5 CONCLUSIONS

This probabilistic model-based economic evaluation demonstrates how clinical trial data can be combined and integrated with country-specific information about resource utilisation and unit cost in order to assess the cost-effectiveness of bronchodilators in COPD patients. Quality-adjusted life months did not substantially differ between treatment groups. In terms of exacerbations, tiotropium was associated with maximum expected net benefit for plausible values of the ceiling ratio. In sensitivity analyses, this outcome was most sensitive to changes in exacerbation rates.

APPENDIX 5.1: THE CALCULATION OF TRANSITION AND EXACERBATION PROBABILITIES

Transition probabilities | Probabilities to transition between disease severity states were based on observed data from the clinical trials. The calculation consisted of the following steps:

1. Prebronchodilator FEV_1 was measured at baseline and regularly thereafter during scheduled follow-up visits. Based on these measurements patients were classified into disease states at each visit.

2. The difference between the frequency distribution of patients over disease states at baseline and the first visit was used to calculate the transition probabilities for the first cycle. The difference between the frequency distribution at the first and last visit was used to calculate transition probabilities for the remaining cycles.

3. The time between baseline and the first visit was eight days in the ipratropium- and placebo-controlled trials and 15 days in the salmeterol-controlled trials. In addition, the time between the first visit and last visit in the ipratropium-, salmeterol- and placebo-controlled trials was 356, 161 and 336 days respectively. Because the length of the first cycle in the model was set to eight days and the length of the remaining cycles was set to one month, probabilities had to be recalculated into probabilities covering an eight days (first cycle) and one-month (subsequent cycles) period.

4. The recalculation of probabilities in case of multiple transitions from a single state is not straightforward and involves a difficult choice among several possible n^{th} root solutions of the transition matrix (Miller and Homan, 1994). In the current study, we obtained a solution for the transition matrix using a Taylor series expansion as described in the last part of this appendix. This step resulted in transition matrices containing eight-days (first cycle) and one-month (subsequent cycles) probabilities. Separate matrices were available for each treatment group by trial (i.e. 3 sets of transition matrices for tiotropium based on the ipratropium-, salmeterol- and placebo-controlled trials, a set of transition matrices for ipratropium and a set of transition matrices for salmeterol).

5a. In order to combine the data from the trials the transition probabilities for patients treated with tiotropium were based on the average of three matrices weighted by the number of tiotropium patients. For instance, the probability to transition from moderate to severe COPD for patients treated with tiotropium during the first eight days was 0.107 in the ipratropium-controlled trials, 0.095 in the salmeterol-controlled trials and 0.080 in the placebo-controlled trials. The numbers of patients treated with tiotropium in these trials were 344, 402 and 550 respectively. Hence, the combined probability to remain in the moderate state during the first cycle was calculated as: $(344 \times 0.107+402 \times 0.095+550 \times 0.080) / (344+402+550) = 0.092$.

5b. The transition probabilities for patients treated with ipratropium and salmeterol were based on the relative differences in transition probabilities as found in the trials. These relative differences are multiplied with the corresponding probabilities calculated at step 5a. For instance, in the ipratropium-controlled trials it was found that the probability for a patient with moderate COPD in the tiotropium and ipratropium group to transition to severe COPD during the first eight days was 0.107 and 0.291 respectively. Hence, the relative probability for ipratropium compared to tiotropium was $0.291 / 0.107 = 2.72$. Finally, the probability for

patients treated with ipratropium to transition from moderate to severe COPD was calculated as 2.72 times 0.092 is 0.257.

6. Standard errors of the probabilities determined in step 5 were calculated as $(p \times (1-p)/n)^{1/2}$, where p is the probability to transition between two disease states and n the original number of patients in the disease state at the start of the interval. For instance, the probability to transition from moderate to severe COPD for patients treated with ipratropium was calculated as 0.257 and the number of patients with moderate disease in the ipratropium group at baseline was 23. Hence, the SE was calculated as $(0.257 \times (1-0.257)/23)^{1/2} = 0.091$.

7. The resulting sets of transition parameters are presented in *table 5.1*. A Dirichlet distribution (Briggs et al., 2003) was assigned to these input parameters and a Monte Carlo simulation was performed in which values were randomly drawn from these distributions. Parameters of the Dirichlet distribution are the numbers of patients transitioning between disease states. For ease of interpretation this table presents the expected values (probabilities) and the associated standard errors rather than these numbers. Probabilities can be recalculated into numbers using method-of-moments fitting (Briggs et al., 2002).

Exacerbation probabilities | The calculation consisted of the following steps:

1. The number of months patients remained in each disease state and the number of exacerbations experienced while being in a particular disease state were collected from the clinical trial data.

2a. The number of exacerbations and months per disease state for tiotropium were summed over all three trials. For instance, the number of exacerbations experienced by patients in the tiotropium group, while being in a moderate disease state were 68.6, 44.1 and 93.4 in the ipratropium-, salmeterol- and placebo-controlled trials respectively. The total number of months patients remained in the moderate disease state were 1469, 689 and 1856 respectively. Hence, the total number of exacerbations for patients treated with tiotropium in the moderate disease state was 206.1 and the total number of months was 4014, an overall probability of 0.051.

2b. To calculate the input parameters for ipratropium and salmeterol we first calculated (for each disease state separately) the trial-specific exacerbation probability for the comparator relative to tiotropium. This relative probability was then multiplied with the overall probability for tiotropium (as calculated at 2a). For instance, the trial-specific exacerbation probabilities for patients in the moderate disease state treated with tiotropium and ipratropium were 0.0467 (68.6 exacerbations in 1469 months) and 0.0733 (23.4 exacerbations in 319 months), respectively. The relative probability of ipratropium compared to tiotropium was 0.0733/0.0467 is 1.568. This relative probability is multiplied with the overall probability for tiotropium of 0.051 is 0.080.

2c. To estimate the SE of the probabilities calculated at 2b, we performed simple simulations based on 5000 iterations. Input parameters were the number of exacerbations and months for tiotropium and the comparator (ipratropium or salmeterol) as observed in the trials and the corresponding numbers for the combined tiotropium data (as calculated at step 2a). In each iteration, probabilities were randomly drawn from beta distributions with these parameters. Hence, each iteration resulted in a probability for the tiotropium arm (trial-specific:

A) the comparator-arm (B) and the combined tiotropium arm (C). The probability of the comparator relative to tiotropium (D) was calculated as B divided by A, and the new probability for the comparator (E) as C times D. The SE is now calculated as the standard deviation of E of the 5000 iterations. A separate simulation was performed for every SE.

Taylor series expansion for the calculation of period-specific transition probabilities | In order to calculate the transition probability per model cycle, we invoke the assumptions of a Markov chain and assume that there is a short-term improvement in disease status within the first eight days immediately following the initiation of treatment. Thereafter, disease progression is assumed to be constant over time. Mathematically, these assumptions can be represented by the equations:

$N_1 = P_1^T N_0$

for the first period, and:

$N_k = P_k^T N_{k-1}$

for the subsequent periods, where N_0 is a vector representing the initial distribution of patients over disease state, N_k is a vector depicting the distribution in period *k*, P_1 is the transition matrix for the first period (transposed as denoted by the superscript 'T') and P_k the transition matrix for the following periods. These equations can be combined because:

$N_1 = P_1^T N_0$

$N_2 = P_2^T N_1 = P_2^T P_1^T N_0$

$N_3 = P_2^T N_2 = P_2^T P_2^T P_1^T N_0 = (P_2^T)^2 P_1^T N_0$

Etc.

Thus $N_k = (P_2^T)^{k-1} P_1^T N_0$ for k = 1, 2, 3, ...

Under these assumptions, we can calculate the period-specific transition probability matrices to be used in the model. These are based on a Taylor Series expansion.

With Taylor Series in one variable, functions f(x) may be represented by a power series of the form:

$f(x) = f(a) + f^{(1)}(a)(x-a)/1! + f^{(2)}(a)(x-a)^2/2! + f^{(3)}(a)(x-a)^3/3! + \ldots + f^{(n)}(a)(x-a)^n/n! + \ldots$

where *a* is a point of interest to the investigator, $f^{(k)}(a)$ is the *k*'th derivative of f(x) evaluated at *a* and k! is the factorial function (k! = k(k-1)(k-2)...(3)(2)(1)). Not unusually, the point *a* of interest is 0. In that case, the Taylor Series reduces to a Maclaurin Series, namely:

$f(x) = f(0) + f^{(1)}(0)\, x/1! + f^{(2)}(0)\, x^2/2! + f^{(3)}(0)\, x^3/3! + \ldots + f^{(n)}(0)\, x^n/n! + \ldots$

The series relevant to this study is the *binomial series* which relates to the function $(1 + x)^b$ (where b is some number, not necessarily either positive or integer). The Maclaurin expansion for this function is:

$(1 + x)^b = 1 + bx/1! + b(b-1)x^2/2! + b(b-1)(b-2)x^3/3! + \ldots$

The corresponding function for $(1 - x)^b$ is:

$(1 - x)^b = 1 - bx/1! + b(b-1)x^2/2! - b(b-1)(b-2)x^3/3! + \ldots$

This expression does not converge for all values of x. For a univariate function, the series converges on the range $-1 \leq x \leq 1$ if $b > 0$ but is not an integer.

Such an expansion may also hold for functions of matrices. In this study, we observe a transition matrix for a given period e.g. 6 months. For the purposes of modelling, this may

not be directly useful as we may wish to work at a shorter time period. In other words, we are interested in functions such as $A^{1/n}$ for some value of n. For example, suppose we have a matrix P_{2mo}, which represents the transition probabilities associated with a 2 months period, and that we wish to estimate the transition probability matrix associated with a 1 month period, assuming the Markov property. If we denote that probability matrix by P_{1mo} then by the Markov assumption:

$P_{2mo} = P_{1mo} P_{1mo} = (P_{1mo})^2$

In other words,

$P_{1mo} = P_{2mo}^{1/2}$

If we observe P_{2mo} (but not P_{1mo}), we can estimate P_{1mo} by:

$(I - (I - P_{2mo}))^{1/2} = I - 1/2(I - P_{2mo})/1! + 1/2(1/2-1)(I - P_{2mo})^2/2! - 1/2(1/2-1)(1/2-2)(I - P_{2mo})^3/3! + \ldots$

since $P_{1mo} = P_{2mo}^{1/2} = (I - (I - P_{2mo}))^{1/2}$ where I is the identity matrix. The question remains as to under what conditions this expansion will converge. It is known that the largest eigenvalue of a transition matrix is equal to 1. Therefore, the *spectral radius* of a transition matrix will be equal to 1 and the series will converge. Thus, we can estimate a binomial function of a matrix, provided it converges. Of course, the expansion has to be truncated at some finite level. This, and other small errors associated with the calculation process may lead to small negative elements in the solution matrix. If this situation arises, the simplest solution is to set the negative element to 0 and adjust the positive elements accordingly.

Acknowledgements

The study was supported by a grant from Boehringer Ingelheim GmbH. We gratefully acknowledge Joseph McDonnell and Maiwenn Al for assistance with the calculations. We also thank the anonymous reviewers for their valuable comments to earlier versions of this manuscript.

CHAPTER 6

RECOURCE USE AND RISK FACTORS IN HIGH-COST EXACERBATIONS OF COPD

6.1 INTRODUCTION

COPD is a highly prevalent, progressive degenerative respiratory disorder and a major cause of disability and premature death. COPD morbidity and mortality are increasing world-wide, due to increasing numbers of smokers, especially among women, and the ageing of the population (Murray and Lopez, 1997). Consequently, an alarming world-wide increase in the burden and costs of COPD is projected for the foreseeable future (Feenstra et al., 2001; Jacobson et al., 2000; Rutten-van Mölken et al., 1999).

Many patients with COPD experience recurrent exacerbations defined as episodes of worsening breathlessness and/or wheeze, often accompanied by greater volume or purulence of sputum and increased cough (Anthonisen et al., 1987; Pauwels et al., 2001). Exacerbations contribute to a more rapid decline in lung function (Donaldson et al., 2002), morbidity and poorer quality of life (Seemungal et al., 2000), as well as to increased healthcare costs. They are the main reason for COPD-related hospital admission, often after failed initial exacerbation therapy in the community (Miravitlles et al., 2002). Estimates of the costs of exacerbations vary from DM 297 (1998; € 152) per exacerbation in a German study (Rychlik et al., 2001), SEK 3136 (1999; € 344) in a Swedish study (Andersson et al., 2002) to FF 3289 (1994, € 501) in a French study (Pechevis et al., 1996). The Swedish and German studies show the costs to rise considerably with the severity of the exacerbation. All of these studies report that inpatient hospital care is the major cost driver, responsible for 67%, 60% and 44% of the exacerbation-related costs in the Swedish, French and German study, respectively. The difference in cost estimates between these studies is remarkable and probably due to a variety of factors including differences in definitions, patient populations, treatment patterns and methods of data collection. Moreover, the classification of exacerbation severity is based on the resource use associated with exacerbations, which of course reinforces the association between severity and costs.

To reduce the costs of exacerbations, we need to know more about risk factors for hospital admission, as this is the major cost driver. Such knowledge would enable us to better target new treatments and to minimise healthcare costs in patients with COPD. The goal of the present study was to determine the costs of exacerbations by exacerbation severity and identify risk factors associated with high-cost exacerbations in patients whose COPD was considered stable at entry into the study. The classification of the severity of exacerbations was based on physicians' assessments and a very detailed record was kept of all relevant types of healthcare utilisation associated with exacerbations.

6.2 METHODS

Design trials | This analysis was based on data from the prospective cost-effectiveness analysis described in chapter 4. This economic evaluation was linked to two randomised controlled double-blind trials comparing 18 µg tiotropium inhalation capsules administered once daily via the HandiHaler® device with ipratropium 2 puffs of 20 µg administered four times daily via the MDI. The results of the economic evaluation have been presented in chapter 4, the design and results of the trials have been published by Vincken et al. (Vincken et al., 2002). In brief, patients with a diagnosis of COPD and a relatively stable airway obstruction defined

as an $FEV_1 \leq 65\%$ of predicted normal and $FEV_1 \leq 70\%$ of FVC were randomised to either tiotropium (n=344) or ipratropium (n=175) in a ratio of 2:1. Patients were aged over 40 and had a smoking history of at least 10 pack-years. Patients with a diagnosis of asthma were excluded. One trial was performed in the Netherlands and the other in the Netherlands and Belgium. Because the design of these trials was identical the data were combined. After a two-week run-in period, patients were seen at regularly scheduled follow-up visits 1, 4, 7, 10, 13, 19, 26, 32, 29, 45 and 52 weeks after randomisation, during which healthcare resource use, health outcomes and adverse events, including COPD exacerbations were recorded. The trials were approved by the medical ethics committees of the participating hospitals, and all patients gave written informed consent.

Exacerbations | All COPD exacerbations were recorded as adverse events in the CRF. A COPD exacerbation was defined as a complex of respiratory symptoms (i.e. new onset or worsening of more than one symptom such as cough, sputum, dyspnoea or wheeze) lasting for at least three days. Exacerbations were classified in the CRF as either mild, moderate or severe based on ratings by the physician-investigator. A mild exacerbation was defined as 'awareness of a sign or symptom which is easily tolerated', a moderate exacerbation as 'an exacerbation causing discomfort enough to cause interference with usual activity' and a severe exacerbation was defined as 'an exacerbation that was incapacitating or causing inability to do work or usual activity'.

Resource use and costs | A very detailed record was kept of all relevant types of healthcare utilisation associated with exacerbations, including: hospital admissions and length of stay (ICU and non-ICU days), ER visits, visits to respiratory physicians, GPs and other healthcare providers that were not scheduled in the clinical trial protocol, ambulance transportations, tests, rescue medication (salbutamol MDI, 1 puff=100 µg) and other respiratory medications including antibiotics. When a COPD exacerbation was associated with a hospital admission this was recorded by the physician-investigator in the CRF. At the end of the trial, all hospital admissions were verified using hospital chart audit. When hospitalised patients withdrew from the study before they were discharged, the total length of hospital stay was taken into account, including the days after withdrawal. To calculate the medication costs during exacerbations, we only included those new respiratory medications that patients started to take during the exacerbation and the dose increases of respiratory medication that patients were already taking before the onset date of the exacerbation. Costs of new medications and dose increases were calculated until the end date of the prescription, with a maximum duration of three weeks after the end date of the exacerbation in case the new medication or new dose was continued. All changes in dose during the exacerbation and the three weeks thereafter were taken into account.

Costs were calculated by multiplying the resource use related to exacerbations with unit costs for the Netherlands in 2001 euro. All healthcare costs were included in the analysis and calculated from the societal perspective. This implies that all healthcare costs were taken into account, regardless of whether they were borne by government, private or public insurers,

or the patients and their families. In addition, unit costs of the major cost drivers were based on estimates of real resource use and not reimbursements. Caregiver and productivity costs were not included.

Statistical analyses | Differences in baseline characteristics between patients with and without exacerbations were tested using t-tests for continuous variables and chi-square tests for categorical variables. Bootstrapping was used to determine 95% CIs around estimates of the duration, resource use and costs of exacerbations by severity. The univariate association between baseline-variables and time to hospital admission were investigated using log-rank tests. In case of continuous variables, the median was used to create two groups. Except for the BMI which was dichotomised using the underweight criterion of 18.5 (National Obesity education initiative, 1998). The independent contribution of each potential risk factor for hospitalisation was investigated using a Cox proportional hazards analysis. Covariates selected for the Cox analysis included: BMI (0: ≥ 18.5; 1: < 18.5), smoking status (0: ex-smoker; 1: current smoker) and smoking pack-years, number of concomitant diagnoses, number of unscheduled physician visits prior to trial, use of inhaled corticosteroids (0: no steroid use, 1: steroid use) and number of concomitant medications, baseline dyspnoea index (BDI), FEV_1 % predicted normal and the total score of the SGRQ. To verify whether the associations between FEV_1 % predicted and hospitalisation and the SGRQ total score and hospitalisation became stronger when not only the baseline values of these parameters were considered, we also performed an analysis in which FEV_1 % predicted and the SGRQ total score were entered as time dependent variables (i.e. an analysis based on four strata of 3 months in which the FEV_1 and SGRQ measurements of the beginning of each stratum were used). The Cox regressions were performed with SPSS® 11.0.1, using a backward stepwise elimination procedure, selecting covariates with a p-value ≤ 0.1. Variable independency was checked by inspection of the correlation matrix. The assumption of proportional hazards was checked using a log-minus-log survival plot.

6.3 RESULTS

Patients | Five hundred and thirty five patients were randomised of whom 519 had completed at least one visit after randomisation. The baseline characteristics of these 519 patients are shown in *table 6.1*, comparing patients with and without exacerbations. The mean (SD) age of patients was 64 years and 440 (85%) patients were male. Compared to patients who had no exacerbations during the observation period, patients who experienced at least one exacerbation had significantly more concomitant diagnoses (2.6 versus 2.2; p = 0.007), higher use of concomitant medications (2.9 versus 2.4; p =0.007), lower FEV_1 % predicted (37.5 versus 41.2; p = 0.001), worse health status (SGRQ total score: 48.7 versus 42.3; p < 0.001) and more dyspnoea at baseline (BDI: 6.7 versus 7.5; p = 0.001).

	Without exacerbations N=312	With exacerbations N=207
Age	63.9 (8.3)	63.9 (8.1)
Males: no. (%)	268 (86)	172 (83)
Current smokers: no. (%)	134 (43)	96 (46)
Smoking history in pack years	33.4 (17.5)	33.9 (17.4)
Duration of COPD in years	11.1 (10.0)	12.1 (9.8)
No. of concomitant diagnoses	2.2 (1.4)	2.6 (1.5)‡
No. of physician visits 3 months prior to trial	0.51 (2.9)	0.49 (2.6)
Body Mass Index (BMI)	24.8 (3.7)	25.1 (4.0)
No. of concomitant medications	2.4 (1.7)	2.9 (2.0)‡
Use of inhaled steroids (y/n): no. (%)	244 (78)	164 (79)
Prebronchodilator FEV_1 (liters)	1.25 (0.44)	1.09 (0.38)‡
Prebronchodilator FEV_1 (% of pred.)	41.2 (12.3)	37.5 (11.6)‡
SGRQ total score	42.3 (16.9)	48.7 (16.4)‡
Baseline Dyspnoea Index (BDI)	7.5 (2.4)	6.7 (2.6)‡

Table 6.1: Characteristics of the study population at baseline (N=519).
Mean (SD) unless stated otherwise; †‡ P-value of the difference between patients with and without exacerbations: † < 0.05, ‡ < 0.01.

Exacerbations | The 519 patients included in the current analysis experienced 364 exacerbations. The mean (95% CI) number of exacerbations was 0.70 (0.60; 0.81) per patient and the mean number of exacerbation-days was 11.32 (9.34; 13.30) per patient. After correcting for differences in the duration that patients remained in the study, the mean number of exacerbations and exacerbation-days per patient-year were 0.80 and 12.97, respectively. About 40% (n = 207) of the 519 patients experienced at least one exacerbation. The maximum number of exacerbations in one patient was nine. The mean (95% CI) duration of an exacerbation was 16.1 (14.8; 17.6) days and the median duration was 12 days. About 10% (n = 36) of the exacerbations was severe, 47% (n = 173) moderate and 43% (n = 155) mild. The mean (median) duration of a severe exacerbation was 25.3 (20) days (95% CI: 19.8; 31.3), which was significantly longer than the mean (median) duration of moderate and mild exacerbations which were 16.6 (13) days (95% CI: 14.7; 18.9) and 13.5 (11) days (95% CI: 12.1; 15.1) respectively.

Resource use during exacerbations | Healthcare resource use data were complete for 350 of the 364 exacerbations (*table 6.2*). Of the severe exacerbations, 78% (n = 28) was associated with at least one hospital admission and 25% (n = 9) with at least one ER visit. Of the moderate exacerbations, 16% (n = 26) was associated with a hospital admission and 5% (n = 8) with a visit to the ER. Only in one case, a patient was hospitalised when experiencing an exacerbation rated as mild by the clinician. In case of a hospitalisation, the mean (median) length of stay was 15.4 (14) (95% CI: 12.5; 19.2) days for a severe exacerbation and 11.8 (10) (95% CI: 9.8; 14.1) for a moderate exacerbation. The mean (median) number of unscheduled visits was 2.3 (2) (95% CI: 1.5; 3.4) in severe exacerbations and 1.6 (1) (95% CI: 1.2; 2.3) in moderate exacerbations. All resource use other than GP-visits was highest in severe exacerbations.

		Severity of exacerbation		
	All (N=350)	Severe (N=36; 10.3%)	Moderate (N=164; 46.9%)	Mild (N=150; 42.8%)
Days general/pulmonary ward	2.05 (5.26)	11.08 (7.89)	1.87 (4.85)	0.08 (0.98)
Days intensive care unit	0.09 (1.30)	0.86 (4.03)	—	—
Visits to respiratory physician	0.39 (0.73)	0.82 (1.32)	0.37 (0.61)	0.31 (0.61)
Visits to GP	0.66 (1.03)	0.70 (0.94)	0.78 (1.08)	0.52 (0.99)
Visits to other healthcare provider	0.29 (2.50)	0.50 (2.33)	0.44 (3.42)	0.09 (0.67)
Visits to ER	0.05 (0.22)	0.25 (0.44)	0.05 (0.22)	0.01 (0.08)
Puffs of salbutamol rescue medication	72 (117)	143 (168)	73 (137)	53 (62)
Ambulance services	0.06 (0.31)	0.31 (0.75)	0.07 (0.27)	—

Table 6.2: Mean (SD) resource use per exacerbation.

Cost of exacerbations | *Table 6.3* presents the unit costs of the major resource use items and the mean costs per exacerbation by severity. The mean (95% CI) cost of an exacerbation was € 720 (515; 1003). The mean costs of severe exacerbations were € 4007 (2551; 6366), which was approximately 7 times higher than the mean costs of moderate exacerbations (€ 579, 95% CI: 407; 769) and 47 times as high as the costs of mild exacerbations (€ 86, 95% CI: 60; 130). The median costs of mild, moderate and severe exacerbations were € 49, € 86 and € 2824 respectively. About 86% of the costs of severe exacerbations resulted from inpatient hospital days and 6% from diagnostic tests. Despite the relatively low percentage of hospitalisations in moderate exacerbations (16%), hospitalisation costs accounted for 71% of the costs of moderate exacerbations. In mild exacerbations, concomitant medications were the main cost driver and accounted for 37% of the total costs whereas unscheduled visits accounted for 33% of the total costs. Costs of concomitant medications varied considerably with exacerbation severity and ranged from € 158 (95% CI: 90; 249) for a severe exacerbation to € 32 (95% CI: 27; 37) for a mild exacerbation. *Figure 6.1* shows the costs of medications during exacerbations. About 58% of these costs were due to antibiotics and about 22% due to systemic corticosteroids.

Figure 6.2 shows the relationship between the proportion of exacerbations and the proportion of total costs of exacerbations. The curve has a steep slope at the beginning, indicating that a small number of exacerbations accounted for a large part of the total costs. The 16% of exacerbations that were associated with a hospitalisation accounted for approximately 90% of the total costs of exacerbations. About half of these exacerbations were rated severe and the other half was rated moderate. The total costs of all exacerbations accounted for approximately 34% of the total respiratory-related healthcare costs that were calculated in the prospective economic evaluation in chapter 4.

Factors associated with time to hospital admission | *Table 6.4* shows the univariate association between patient characteristics and the Kaplan-Meier estimate of the cumulative proportion of patients with hospitalisation. A low BMI, a high number of concomitant diagnoses, a high number of respiratory medications, a low BDI score and assignment to the ipratropium

	Unit costs per day/visit	All (N=350)	Severity of exacerbation		
			Severe (N=63)	Moderate (N=176)	Mild (N=111)
General/pulmonary ward	222	454 (1167)	2456 (1748)	413 (1074)	18 (217)
Intensive care unit	1146	101 (1493)	987 (4617)	—	—
Visits respiratory physician	52	20 (38)	43 (69)	19 (32)	16 (32)
Visits to GP	18	12 (18)	12 (16)	14 (19)	9 (17)
Visits to ohcp	19	6 (48)	10 (44)	8 (65)	2 (12)
Visits to ER	98	5 (22)	25 (43)	5 (21)	1 (8)
Rescue medication		2 (3)	4 (5)	2 (4)	1 (2)
Respiratory medications		52 (95)	158 (244)	46 (52)	32 (32)
Diagnostic tests		53 (167)	236 (340)	54 (157)	7 (20)
Ambulance transportations		16 (79)	77 (189)	17 (69)	—
Total healthcare costs		720 (2354)	4007 (5922)	579 (1227)	86 (233)

Table 6.3: Mean (SD) costs per exacerbation in 2001 euro.
HCP: healthcare provider.

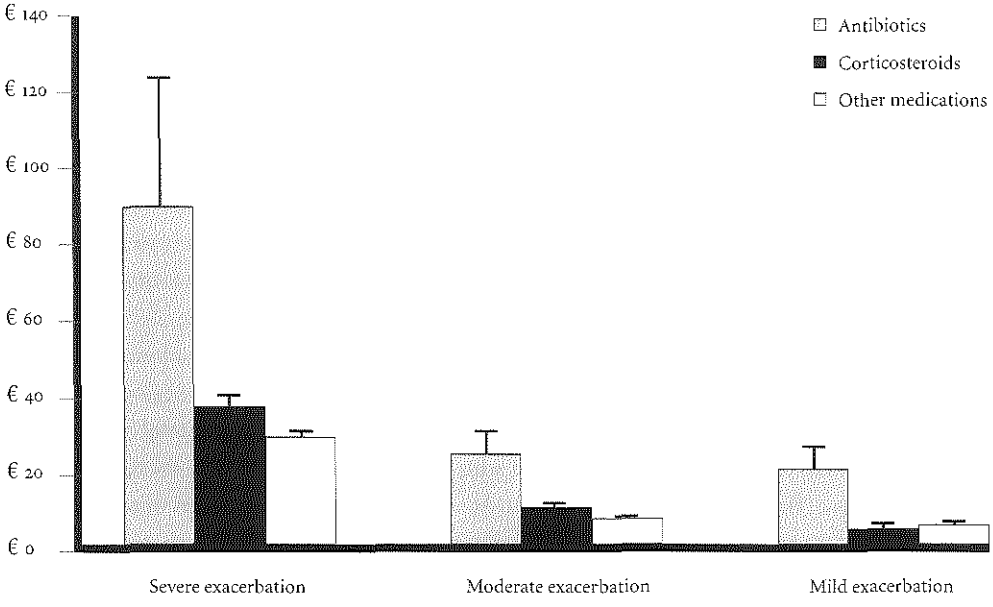


Figure 6.1: Costs of medication per exacerbation in 2001 euro.

treatment arm were significantly associated with an increased risk of hospitalisation. The largest differences in the proportions of patients with hospitalisation were found between patients with a history of more or less than two concomitant diagnoses (0.19 versus 0.06) and between patients with a BMI below or above 18.5 (0.37 versus 0.10).

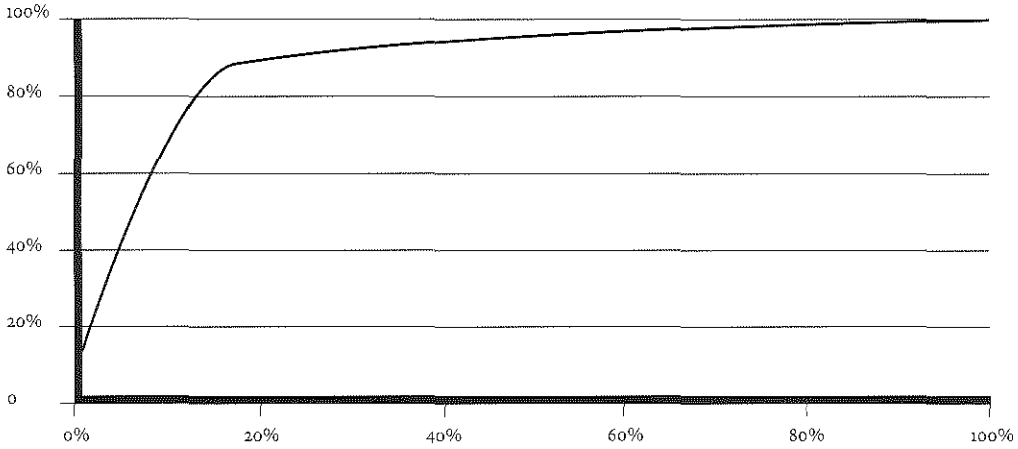


Figure 6.2: Cumulative costing curve of the costs per exacerbation.

In the Cox proportional hazards analysis BMI, treatment group, the number of concomitant diagnoses at baseline and BDI were independent risk factors of hospital admission (*table 6.5*). The risk of hospitalisation for patients with a BMI below 18.5 was 3.6 times higher than for patients without underweight. A one-unit decrease (=increased dyspnoea) in the BDI increased the risk of hospitalisation by 18.2% and each additional diagnosis increased the hazard rate by 40.4%. Patients treated with tiotropium experienced less than half the hospitalisation risk of patients treated with ipratropium. The correlation matrix showed that the Pearson correlation coefficients between the 11 variables that entered the model at the start were always less than 0.37, except for the BDI and the SGRQ. These two variables were highly correlated (Pearson R: 0.651; $p < 0.001$). When the baseline values of both were entered, the BDI proved to be a more powerful predictor. When the BDI was left out, the SGRQ demonstrated an increased risk of hospitalisation of 8% for each four units increase (=worsening) in the SGRQ total score. Entering FEV_1% predicted and the SGRQ as time-dependent variables did not improve the model, nor did adding the SGRQ domains instead of the SGRQ total score. Excluding treatment arm as a covariate did not lead to a statistical significant contribution of other risk factors, while the other significant covariates in the model were hardly affected.

6.4 DISCUSSION

About 40% of the patients whose COPD was judged to be stable at entry into the current study experienced at least 1 exacerbation and the mean number of exacerbations per patient-year was 0.80. The mean duration of an exacerbation was 16 days. There was a wide variation in the costs of exacerbations, ranging from € 4007 for a severe exacerbation to € 579 and € 86 for moderate and mild exacerbations. About 16% of the exacerbations was associated with a hospital admission and these exacerbations were responsible for about 90% of the total costs of exacerbations. A BMI below 18.5, a higher number of concomitant diagnoses and increased dyspnoea (a low BDI score) at baseline were significantly associated with an increased risk of hospitalisation. In addition, treatment with tiotropium reduced the risk of hospitalisation by 57% compared to ipratropium.

Predictor variable	Subgroups*	Cumulative proportion of patients with hospitalisation†	P-value (log-rank test)
Smoking status	ex smoker	0.09	
	smoker	0.14	0.094
Smoking pack years	< 30	0.08	
	> 30	0.16	0.058
Body Mass Index (BMI)	< 18.5	0.37	
	> 18.5	0.10	< 0.001
No. of concomitant diagnoses	< 2	0.06	
	> 2	0.19	< 0.001
No. of concomitant medications	< 2	0.07	
	> 2	0.16	0.019
Use of inhaled steroids (y/n)	no steroid use	0.11	
	steroid use	0.11	0.690
No. of physician visits 3 mo prior to trial	< 1	0.11	
	> 1	0.07	0.760
FEV_1 (% of predicted normal)	< 39	0.13	
	> 39	0.10	0.070
SGRQ total score	< 44	0.08	
	> 44	0.15	0.066
Baseline Dyspnoea Index (BDI)	< 7	0.17	
	> 7	0.06	0.004
Treatment group assignment	ipratropium	0.14	
	tiotropium	0.10	0.048

Table 6.4: Univariate analysis of factors associated with time to hospitalisation.
** In case of continuous variables the median was used to create two groups; BMI was split into two groups using the underweight criterion of 18.5 (National Obesity education initiative, 1998); † cumulative proportions based on Kaplan-Meier curves.*

Predictor variable*†	Increment	Relative risk	95% CI
Body Mass Index (BMI)	0 = > 18.5 1 = < 18.5	3.62	1.50; 8.71
No. of concomitant diagnoses	per additional diagnosis	1.40	1.15; 1.72
Baseline Dyspnoea Index (BDI)	per unit decrease	1.18	1.04; 1.34
Treatment group assignment	0 = ipratropium 1 = tiotropium	0.43	0.23; 0.78

Table 6.5: Final Cox proportional hazards analysis of the time to hospitalisation for COPD exacerbation.
** Backward stepwise eliminination procedure ($p < 0.10$); † variables entered into the full model: smoking status, smoking pack years, BMI, no. of concomitant diagnoses, no. of concomitant medications, use of inhaled corticosteroids, no. of physician visits 3 months prior to the trial, FEV_1 % predicted, SGRQ total score, BDI, treatment group.*

The definition of exacerbation severity was based on physicians' assessments of the extent to which an exacerbation interfered with the ability to perform usual activities. Previous studies have used healthcare resource use to define severity. In such a classification, severe exacerbations are often defined as those associated with a hospital admission, moderate exacerbations as those associated with an outpatient visit with or without the prescription of an oral steroid or an antibiotic, and mild exacerbations as those that are primarily self-managed. We did not adopt this definition because treatment patterns are likely to vary across settings. An exacerbation that is severe in one country might not be rated as severe in another country because the countries may use different criteria to hospitalise a COPD patient. However, when we applied this definition to our study, we observed a further increase of the difference in costs between severe exacerbations (€ 4117) on the one hand and moderate (€ 123) and mild exacerbations (€ 29) on the other hand, signifying the large impact of hospitalisations on the cost per exacerbation.

The Cox proportional hazard analysis showed that, apart from treatment group, a low BMI, a higher number of concomitant diagnoses and dyspnoea measured with the BDI were significantly associated with hospitalisation. The BDI was a better predictor than FEV_1, which is generally considered to be the most important marker of COPD severity. Even when BDI was removed from the model, FEV_1 was not identified as an independent risk factor for hospital admission. This finding is in line with other studies, which reported that dyspnoea was predictive of (severe) re-exacerbations (Seemungal et al., 2000) and mortality (Nishimura et al., 2002). Dyspnoea might *'reflect more comprehensive information than airway obstruction'* (American Thoracic Society, 1999) and might be a marker of COPD severity that is at least as important as FEV_1. Especially when dyspnoea is measured with an instrument like the BDI, which also covers the functional aspects of breathlessness as assessed by the patient. This probably explains part of the interchangeability between the BDI and the SGRQ. The SGRQ has been shown to predict hospital admission before (Osman et al., 1997). When BDI was removed from the model, the SGRQ became a significant risk factor of hospitalisation, whereas the coefficients of the other significant covariates were hardly affected. When both were entered, the BDI appeared to be more powerful. Even when the SGRQ was entered as a time-dependent variable – to study whether quality of life in the period before the hospitalisation is a better predictor than quality of life at baseline – the BDI continued to show a stronger association with hospital admission than the SGRQ.

Strengths of the current study were the prospective and detailed collection of data about exacerbations and resource use. It should be noticed, however, that these data were collected in a clinical trial setting. As in most clinical trials, patients were monitored closer than in daily practice. There is always a risk that the regularly scheduled trial visits substitute visits that would have occurred if the trial had not taken place. On the other hand, patients in a trial are often more strongly encouraged to contact the physician when their condition deteriorates. As described in chapter 4, the trial was performed in a population of stable COPD patients with less comorbidity than in the average COPD population. Even in this population, we found that exacerbations accounted for approximately 34% of the total respiratory-related healthcare costs. This 34% compares well to a Swedish observational study that applied very

few selection criteria. In that study it was found that approximately 35% – 45% of the costs of COPD treatment were due to exacerbations (Andersson et al., 2002). These percentages are very different from a number of database-studies reporting costs in the US (Hilleman et al., 2000; Sullivan et al., 2000). Although these studies did not explicitly relate costs to exacerbations, it was estimated that approximately 70% of the costs of treating patients with COPD were due to hospitalisations. This higher percentage is probably related to the higher unit costs of an inpatient hospital day in the US, but may also be due to differences in treatment patterns and study design. The data at least suggest that there are large differences in the costs of treating COPD across countries, which may have large impacts on the cost-effectiveness of (new) treatments in COPD. Any treatment that successfully prevents severe exacerbations and costly hospitalisations is likely to be most cost-effective in countries with high costs of hospitalisation.

6.5 CONCLUSIONS

Many patients whose COPD is judged to be stable experience exacerbations. According to our definition of severity, roughly 40% of the exacerbations was mild, 50% was moderate and 10% was severe. Costs of severe exacerbations were approximately 7 and 47 times as high as the costs of moderate and mild exacerbations respectively. Exacerbations that were associated with a hospitalisation accounted for 90% of the total costs of exacerbations. A low BMI, a history of concomitant diseases and increased dyspnoea are factors that are likely to identify patients who are at increased risk of generating high costs as a result of hospitalisation.

Acknowledgements

The study was supported by a grant from Boehringer Ingelheim GmbH. We gratefully acknowledge Wim Hop of the department of Epidemiology and Biostatistics, Erasmus MC Rotterdam, for assistance with the statistical analyses.

CHAPTER 7

METHODS TO ANALYSE INCOMPLETE COST DATA OF PATIENTS WHO DROPOUT IN A CLINICAL TRIAL SETTING

7.1 INTRODUCTION

Missing data because of patients withdrawing from a study before reaching the scheduled end date cause a well-known problem in the analysis of longitudinal data. If patients are to be followed for a fixed time period, for instance one year, and a patient decides to withdraw from the study after nine months, data for the last three months may be lacking. Although it is better to avoid withdrawal (dropout) or to continue the collection of data after dropout, in practice this often proves to be impossible and the data of patients who withdraw cannot be analysed using standard methods for analysis. Missing data because of dropout is a common problem, but it has received little attention in the context of economic evaluations.

Compared with most clinical and quality of life data, cost data have some special characteristics affecting their analysis. Firstly, the variable of interest in cost-analysis is the cumulative cost over a certain time-period, whereas clinical and quality of life data are usually measured at several points in time. Secondly, costs tend to vary widely over time within one patient. A patient, who is being hospitalised during time interval t, will have very high costs over this interval, while costs over t-1 and t+1 can be very low or even zero. Thirdly, resource use data are usually characterised by severe skewness of the underlying probability distribution; a few patients with very high costs and the majority of patients with limited or even zero costs. The combined occurrence of these three properties is characteristic for cost data and it is unclear whether existing methods for dealing with censored data are suitable to deal with these characteristics.

Despite the fact that premature study withdrawal is likely to occur in almost every longitudinal economic evaluation, only few studies actually report the rate of dropout and the methods that have been used to analyse the data from these patients (Barber and Thompson, 1998; Rutten-van Mölken et al., 1994). Whether or not dropout is a serious problem in a particular study depends on the rate of dropout, the difference in this rate between treatment groups and the underlying causes for dropout. Particularly in those situations in which a relationship is found between disease severity and the rate of dropout, or when the rate of dropout differs between treatment groups, the impact of the method used to deal with the data of these patients may be large.

The methods that are available to deal with the missing data of patients who dropout can be distinguished into so-called naïve and principled methods (Schafer, 1997). Naïve methods aim to provide an estimate of the mean costs by omitting patients (complete cases analysis) or by imputing one single value for each missing observation. Naïve methods do not adjust the variance for the missing observations and, consequently, the CIs or p-values of the cost analysis are likely to be biased. This is even true in cases where the dropout pattern is completely at random. Examples of naïve methods are extrapolation, last value carried forward (LVCF), or regression-based methods such as predicted mean. Principled methods do not only provide an estimate of the mean cost but also aim to provide an unbiased estimate of the variance by taking account of the missing observations (Schafer, 1997). Examples of principled methods are the expectation maximisation (EM) algorithm (Little and Rubin, 1987), multiple imputation (Rubin, 1987), general linear mixed models (Zeger and Liang, 1986), and techniques based on survival analysis (Lin et al., 1997). To date, the use of principled methods to

deal with cost data of dropouts in economic evaluations has been very limited and mainly focused on the use of survival based methods. This method provides an unbiased estimator of the average costs in case of independent censoring, but this is not necessarily the case when dropout is related to the condition of the patient (Lin et al., 1997).

In the prospective economic evaluation of tiotropium versus ipratropium in chapter 4, we were confronted with patients who dropped out during the study before their scheduled end date. The overall dropout rate was modest (17%), but worsening of a patient's health was the reason for more than 70% of the dropouts. Hence, there was a large difference in the mean costs per day of the dropouts compared with the costs of the patients who completed the study. To deal with the data of dropouts in this study we applied multiple imputation and compared the results with four naïve methods; complete cases analysis, linear extrapolation, hot decking and predicted mean. The objectives of this paper were to demonstrate the impact of dropouts on the outcomes of an economic evaluation, and to compare the mean and variation in costs obtained with each of the five methods to deal with missing observations.

7.2 METHODS

The trials | The economic evaluation was performed alongside two clinical trials performed in the Netherlands and Belgium in patients with COPD comparing the new long-acting bronchodilator tiotropium (18 µg once daily) with the short-acting bronchodilator ipratropium (40 µg four times daily). COPD is a chronic progressive disease of the respiratory system, characterised by chronic bronchitis and/or emphysema. Patients with COPD have a decreased pulmonary function and usually show symptoms of cough, sputum production and dyspnoea upon exertion (Pauwels et al., 2001). Treatment of patients with COPD is directed to relieve symptoms and to prevent exacerbations of the disease. To be included in the trials, patients with a diagnosis of COPD were required to have a moderate to severe airway obstruction ($FEV_1 \leq 65\%$ predicted normal), to be at least 40 years of age, and to have a smoking history of more than ten pack-years. Patients with a history of asthma or any significant disease other than COPD were excluded. Patients were randomised to receive either tiotropium or ipratropium in a ratio of 2:1. The clinical studies showed that FEV_1 improved above baseline by 120 ML after one year for patients receiving tiotropium, whereas it declined by 30 ML for patients receiving ipratropium. Tiotropium was also found to be more effective in improving dyspnoea, exacerbations and health-related quality of life.

Primary outcomes of the economic evaluation were the incremental cost-effectiveness ratios over one year: cost per exacerbation avoided and cost per patient with a relevant improvement on the SGRQ. The SGRQ is a disease-specific quality of life questionnaire, with scores ranging from 0 (no impairment) to 100 (full impairment) (Jones et al., 1991). Direct healthcare costs were measured from a societal perspective and included inpatient hospital days on wards and ICUs, outpatient visits to pulmonologists, GPs and other healthcare providers, ER visits, ambulance transportations and costs of study drugs and concomitant medications. Only respiratory-related resource use was included. Health outcomes and resource use were collected during scheduled follow-up visits 1, 4, 7, 10, 13, 19, 26, 32, 39, 45 and 52 weeks after randomisation. At each visit resource use data since the previous visit were collected.

Costs were calculated by multiplying the resource use per patient with fixed country-specific unit costs expressed in 2001 euro. The economic evaluation is described in chapter 4. For a detailed description of the trial we refer to other publications (Vincken et al., 2002). The costs reported in this paper differ slightly from the costs reported in chapter 4, because resource use for all patients in chapter 4 was based on Dutch unit costs. In the current analysis the resource use of Belgian patients (15% of the patients) is valued using Belgian unit costs.

Problem description | A total number of 535 patients were randomised into the trials, 356 into the tiotropium and 179 into the ipratropium group. *Figure 7.1* shows the proportions of patients in both treatment groups remaining in the study at each visit. Five hundred and nineteen patients, 344 (97%) in the tiotropium group and 175 (98%) in the ipratropium group, completed at least the first clinic visit one week after randomisation and were included in the analyses. After six months, the numbers of patients in these groups were 323 (91%) and 155 (87%) respectively. Three hundred and two patients (85%) in the tiotropium group and 141 patients (79%) in the ipratropium group completed the entire one-year trial (Pearson Chi-square p = 0.08). The mean (SD) time on treatment of patients who dropped out of the study was 113 (120) days in the tiotropium group and 107 (96) in the ipratropium group (Student's t-test p = 0.81). *Table 7.1* shows the reasons for dropout in the two treatment groups. All patients who completed at least the first visit after one week were included in the economic evaluation.

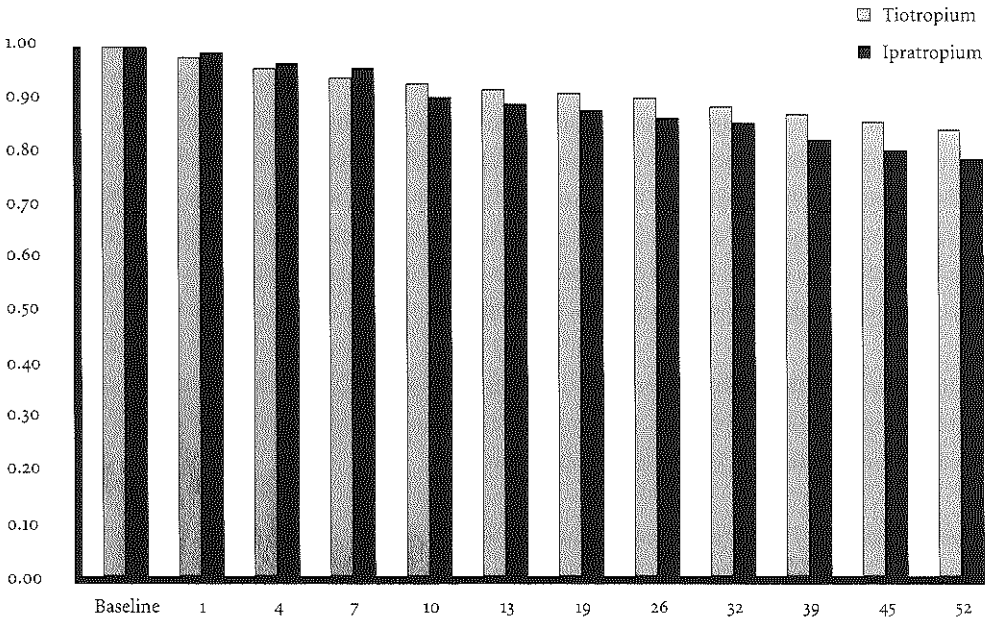


Figure 7.1: Proportion of patients remaining in the study at each visit.

Reason for dropout	Tiotropium (N=356)	Ipratropium (N=179)
Worsening of COPD	11 (3.1)	11 (6.1)
Worsening of other disease	2 (0.6)	4 (2.2)
Other adverse events	23 (6.5)	8 (4.5)
Withdrawn consent	5 (1.4)	4 (2.2)
Lost to follow-up	2 (0.6)	0 (0.0)
Lack of efficacy	3 (0.8)	3 (1.7)
Non-compliant	3 (0.8)	3 (1.7)
Other reasons	5 (1.4)	5 (2.8)
Total	54 (15.2)	38 (21.2)

Table 7.1: Number (%) of dropouts by reason and treatment group.

There were considerable differences between patients who completed the trial and those who dropped out before the scheduled end date. Mean (SD) age of the completers was 63.6 (8.1) years versus 65.4 (8.5) years for the dropouts (Students' t-test p = 0.05). Patients who dropped out also had a lower mean (SD) FEV_1 percentage predicted normal at baseline (completers 40.3% (12.1); dropouts 37.0% (12.5); Students' t-test p = 0.02) and lower quality of life as measured by the SGRQ (completers 44.2 (16.8); dropouts 48.7 (17.1); Students' t-test p = 0.02). In addition, a large difference was found between the costs of patients who completed the study versus the costs of patients who dropped-out before the scheduled end date. The mean (SD) costs per day of the completers were € 2.64 (6.56) in the tiotropium group and € 2.66 (4.06) in the ipratropium group. The mean costs per day of the dropouts during the time they remained in the study were approximately four times higher; € 10.32 (21.75) in the tiotropium group and € 10.98 (16.87) in the ipratropium group. A bootstrap analysis showed that the differences between the completers and the dropouts were statistically significant (mean (95% CI) difference in the tiotropium group: 7.87 (1.82; 14.87); ipratropium group: 8.32 (3.14; 14.95). The combined occurrence of a difference in dropout rate between the two treatment groups and the large difference in the costs per day between the completers and the withdrawals was the main reason to further investigate the impact of dropouts on the mean and variation in costs.

Methods to deal with the data of dropouts | Five methods were applied to deal with the cost data of dropouts (Little and Rubin, 1987): complete cases analysis, linear extrapolation, predicted mean, hot decking, and multiple imputation.

Complete cases | The complete cases analysis excluded the data of all patients who dropped-out the study before the scheduled end date. Hence, the analysis was based on patients with complete follow-up data only.

Linear extrapolation | In linear extrapolation, the costs of the censored patients were extrapolated to 1 year by dividing the observed costs of a patient by the number of days that particular patient remained in the study and multiplying the result by 365.

Predicted mean | Predicted mean (or regression imputation) uses ordinary least squares multiple regression analysis to impute the most likely value for each missing cost observation (Little and Rubin, 1987). Separate regression equations were estimated for both treatment groups and for all subsequent visits using data from patients who had an observation in that particular period. Independent variables included demographic and baseline variables and costs and health outcomes from previous periods. The relationship between costs per visit and all independent variables was determined through regression and variables were selected if their p-value was < 0.05 at least once, or < 0.10 at two or more time-periods. If the correlation between the selected independent variables was above 0.40, only the variable with the highest frequency of significant correlations with costs was selected. Similar sets of independent variables were used in the two treatment groups. *Table 7.2* shows the variables that were selected as independent variables. For each patient with a missing cost observation the value predicted by the regression equation was imputed.

Baseline variables	Variables from the previous period
Weight	Healthcare costs
Evening peak expiratory flow rate (PEFR)	The number of COPD exacerbations
Smoking status (current or former smoker)	Disease-specific quality of life (SGRQ total score)
Number of pack-years smoked	Transitional Dyspnoea Index Score
Weekly number of puffs of rescue medication prior to trial	Pulmonary function (FEV_1)
Use of short-acting bronchodilator prior to trial (yes/no)	
Number of unscheduled visits to physician prior to trial	
Number of unscheduled visits to other HCP prior to trial	

Table 7.2: Independent variables selected for predicted mean and multiple imputation.
HCP: healthcare provider.

Hot decking | Hot decking involves the selection of a limited number of categorical variables by which patients are sorted in so-called imputation classes (Little and Rubin, 1987). The intention is to create imputation classes which each contain a homogeneous group of patients. To form imputation classes in the current analysis, two variables that showed a strong relationship with costs were selected. A variable indicating whether or not a patient had an exacerbation during the previous period, and a variable related to costs during the previous period. The latter was recoded into a categorical variable such that patients were divided in three equal-sized groups. Hence, hot decking was performed using six (two times three) imputation classes. The procedure was performed separately for both treatment groups and for all subsequent visits. Within an imputation class, for each patient with a missing cost observation, a value was randomly drawn from the patients who were in the same imputation class and who had complete follow-up data for that visit.

Figure 7.2 shows a simplified example of the hot deck procedure. In this example, 20 patients were divided into four groups (imputation classes) based on their costs and exacerbations in the previous period: low costs without an exacerbation, low costs with an exacer-

bation, high costs without an exacerbation and high costs with an exacerbation. Four patients had a missing observation. For each of these patients, a value was randomly drawn with replacement from the other patients in the same imputation class. Imputing these values for the missing observations resulted in a complete data set.

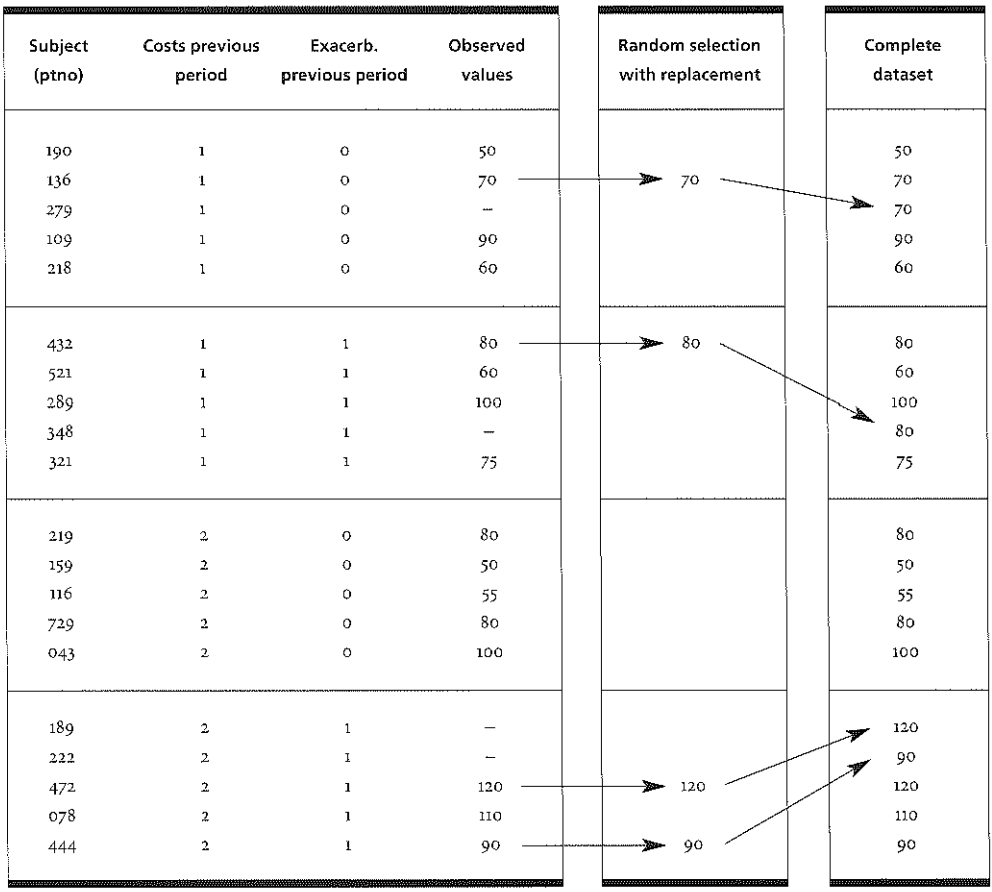


Figure 7.2: Simplified example of the hot deck procedure.

Multiple imputation | Multiple imputation is a technique that, instead of imputing one value for each missing observation, replaces each missing observation with a set of m (in this case ten) plausible values (Lavori et al., 1995; Rubin, 1987; Rubin and Schenker, 1991). Hence, the method resulted in ten complete data sets for which the overall mean and variance were estimated. The mean was simply calculated as the mean of the ten data sets. The variance within data sets was combined with the variance between data sets to take account of the extra uncertainty that resulted from missing values, using (Schafer, 1997):

$$\text{Total variance} = \text{Var}_{in} + (1 + m^{-1}) \times \text{Var}_{bet}$$

where Var_{in} is the mean of the variances within the data sets divided by the sample size, Var_{bet} is the variance between the data sets, and m is the number of imputed data sets.

Several methods are available to impute values within each data set (Schafer, 1997). In the current analysis we applied the propensity score method that consisted of four steps. Firstly, a logistic regression was performed to predict the probability that a cost observation was missing. The outcome is called a propensity score. Note that, unlike the predicted mean method, the outcome of the regression reflects the probability a value was missing rather than the value to impute. Next, patients were sorted according to their propensity score and divided into five equal-sized groups. During the third step, a random sample with replacement was drawn from the observed values within each quintile. This sample is called the posterior predictive distribution. Finally, for each missing observation, a value was taken at random from this posterior predictive distribution. *Figure 7.3* gives a graphical presentation of these four steps (O'Callaghan, 1999). This procedure is repeated ten times for both treatment groups and for all subsequent visits. The procedure to select covariates for the logistic regression was similar to the one described above in relation to the predicted mean method.

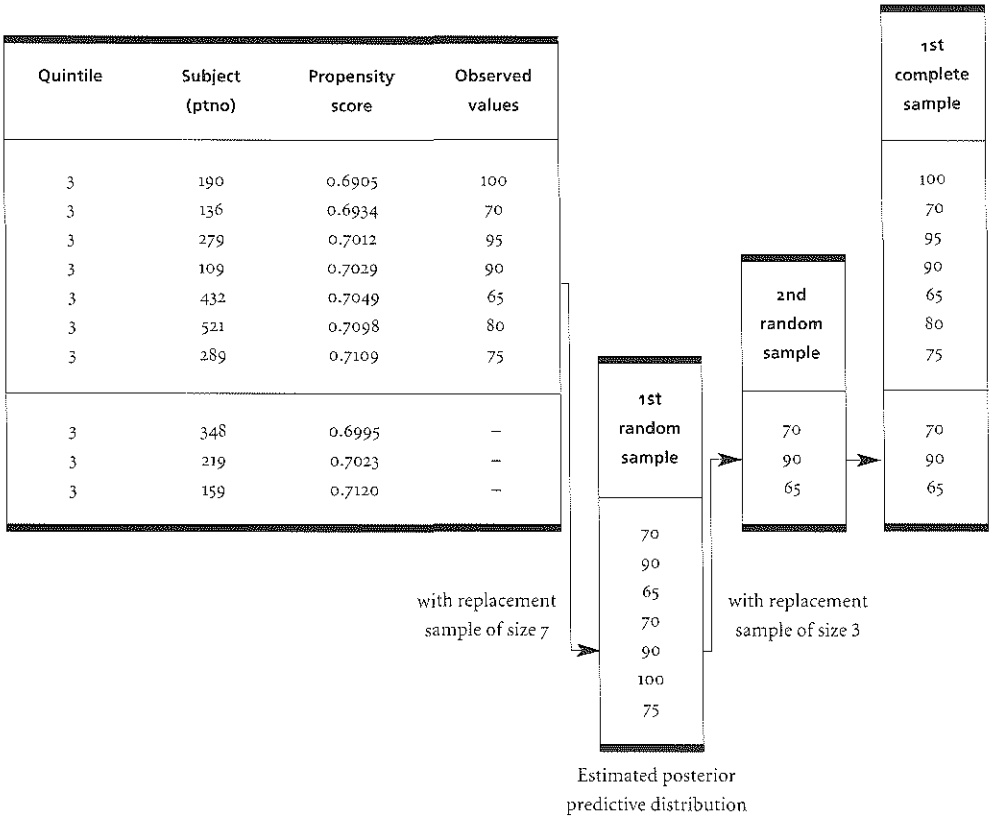


Figure 7.3: Graphical presentation of the propensity score method to impute values using the multiple imputation procedure (reproduced from: O'Callaghan, 1999, with permission).

'As Observed' costs | In addition to the five methods described above, the results section reports the costs 'as observed', which were calculated as the sum of the observed costs of all patients (completers and dropouts) divided by the sum of days patients remained in the study and multiplied by 365. This calculation resulted in an estimate of the mean cost per patient without an estimate of the variation around this mean. Hence, it was not a proper method to statistically analyse data of censored patients. The results of this method were reported only for reasons of comparison.

Analysis | The results are expressed as the mean costs and standard errors per patient year. Data were analysed using SPSS version 10.0.7. Multiple imputation and hot decking were performed using Solas 2.0.

7.3 RESULTS

Mean costs | Because 16 patients (12 in the tiotropium group and 4 in the ipratropium group) discontinued the trial before the first follow-up visit, 519 patients were included in the cost-effectiveness analysis of whom 344 in the tiotropium group and 175 in the ipratropium group. *Table 7.3* shows the estimates of the mean (SE) annual costs per patient within the two treatment groups after applying each method. Costs in the tiotropium group varied from € 955 after complete cases analysis to € 1298 after linear extrapolation. The corresponding estimates in the ipratropium group were € 970 and € 1561 respectively. The difference between the imputation methods on the one hand and the complete cases and as observed analysis on the other hand was greater in the ipratropium group than in the tiotropium group. This reflected the greater proportion of dropouts because of worsening of COPD in the ipratropium group. Estimates of the mean costs after hot decking, predicted mean and multiple imputation in the tiotropium group were almost similar and varied from € 1110 after predicted mean to € 1150 after multiple imputation. In the ipratropium group they varied from € 1415 after multiple imputation to € 1512 after predicted mean. Higher estimates of the mean costs were associated with increased standard errors. SEs were found to be lowest after the complete cases analysis and highest after linear extrapolation.

	Tiotropium		Ipratropium		Difference	
	Mean	SE	Mean	SE	Mean	95% CI
As observed	1081	NA	1252	NA	-171	NA
Complete cases analysis	955	137	970	125	-15	-379 ; 349
Linear extrapolation	1298	198	1561	244	-263	-878 ; 353
Predicted mean	1110	136	1512	204	-402	-883 ; 79
Hot decking	1126	133	1485	163	-359	-771 ; 54
Multiple imputation	1150	160	1415	161	-265	-709 ; 180

Table 7.3: Mean (SE) costs per treatment group after applying different methods to deal with the data of dropouts (in 2001 euro). NA: not applicable.

Mean difference in costs | The difference in costs between treatment groups varied from € 15 after complete cases analysis to € 402 after predicted mean analysis. Despite the high costs after linear extrapolation in both groups, the corresponding difference in costs was only € 263, similar to the difference obtained by multiple imputation and considerably lower than that obtained by hot decking and predicted mean. The as observed analysis showed that a considerable share of the estimated annual difference in the costs between the treatment groups was already observed before applying any of the imputation methods. *Figure 7.4* gives a graphical presentation of the differences in mean costs and the corresponding 95% CIs. The figure shows that the difference in costs was not statistically significant with any of the methods. The widest interval was obtained after linear extrapolation and the smallest after complete cases analysis.

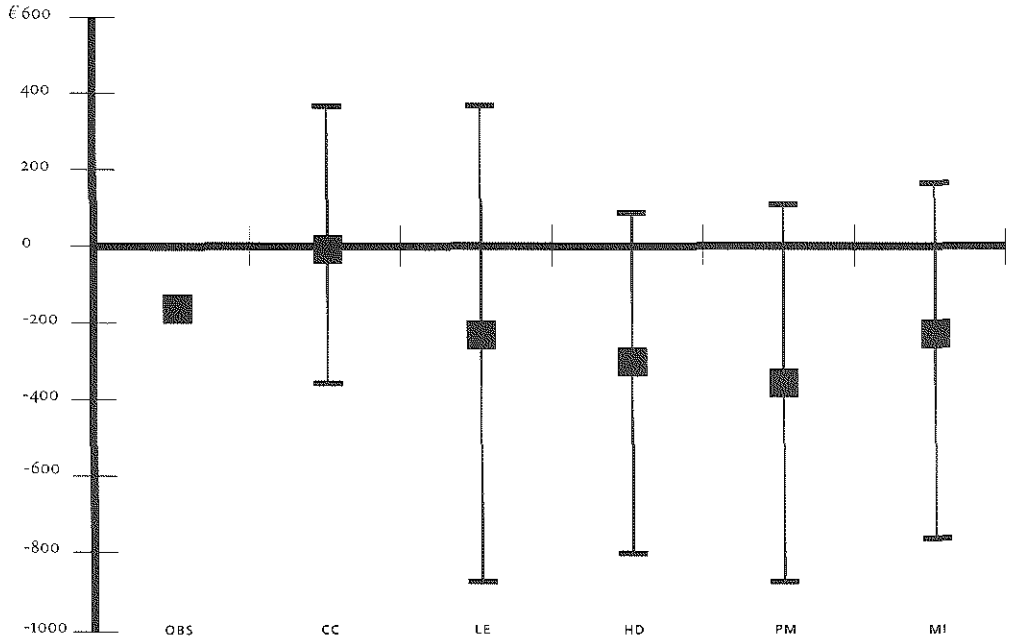


Figure 7.4: Mean and 95% CI of the difference in costs between treatment groups in 2001 euro.
OBS=AS observed; CC=complete cases analysis; LE=linear extrapolation; PM=predicted mean; MI=multiple imputation.

Costs of the dropouts | *Figure 7.5* shows the mean costs of the dropouts only. The figure shows that in the tiotropium group, a large difference was found between the mean costs of the dropouts after linear extrapolation compared with the other methods of imputation. Estimates of the mean costs ranged from € 2229 to € 3768. In the ipratropium group, the difference between linear extrapolation and any of the other methods was much smaller. Estimates of the mean costs ranged from € 3260 after multiple imputation to € 4011 after linear extrapolation. Costs of the dropouts in the ipratropium group were consistently higher than in the tiotropium group. Especially after hot decking and predicted mean, the difference in costs between treatment groups was found to be large, € 1260 and € 1532 respectively.

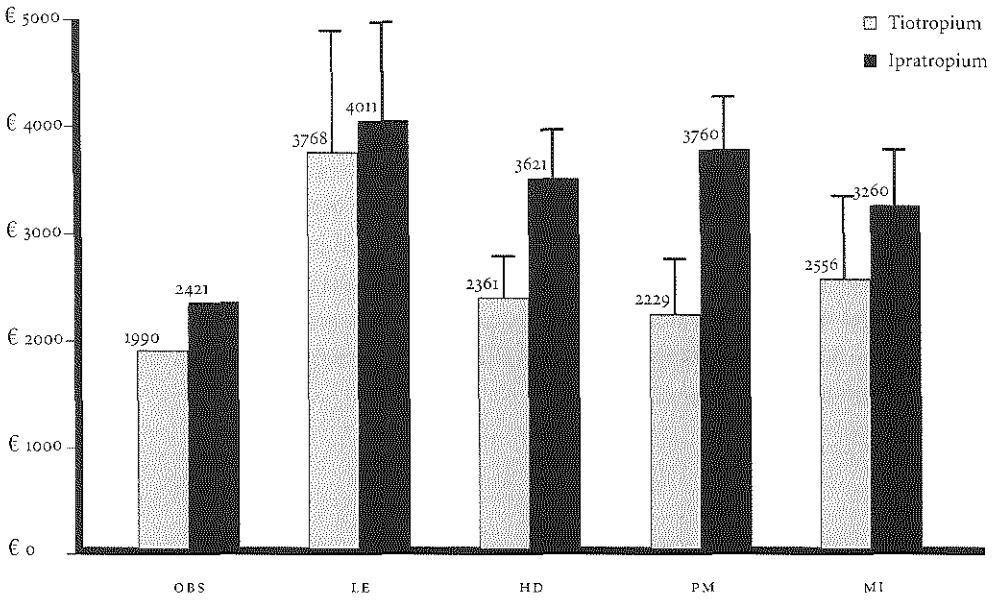


Figure 7.5: Mean one-year costs (SE) of patients who dropped out the study before the scheduled end date in 2001 euro. OBS: as observed; LE: linear extrapolation; HD: hot decking; PM: predicted mean; MI: multiple imputation.

7.4 DISCUSSION

In this study, we applied five different methods to obtain estimates of the mean annual costs in a situation where 15% of the patients in one group and 21% of the patients in the other group withdrew from the study before the scheduled end date. Complete cases analysis did not incorporate the costs of dropouts for the time they remained in the study and, because of the higher costs of the withdrawals before dropout and the difference in dropout rate between treatment groups, underestimated the mean and difference in costs. Estimates of the difference in costs obtained with the other methods varied from € 171 in the as observed analysis to € 402 with the predicted mean method. In relation to the mean total costs in this study, such a difference between the estimates will usually be considered to be relevant and may give rise to different interpretations of the results. Linear extrapolation showed the largest CI of the difference in costs (from –878 to 353). CIs after predicted mean (-883; 79) and hot decking (-771; 54) were much smaller and came close to statistical significance.

Because the 'true costs' of the dropouts in this study were unknown, the study was unable to demonstrate which method resulted in the best estimate of the mean and difference in costs. For such an analysis, simulation studies in which dropout is artificially created in a complete data set are needed. The mean and variance can then be estimated using a number of methods to deal with the data of dropouts, and these can be compared with the true mean and variance. We are currently conducting such simulation studies. Nevertheless, the current study challenges us to speculate on the performance of the five methods. We believe it is safe to say that the as observed analysis provides a minimum estimate of the true cost, since it includes all actually observed costs before withdrawal, and we do not expect the costs after withdrawal to reduce to the level of the costs of the completers because the costs are associated with

worsening disease. Linear extrapolation perhaps overestimates the mean costs, because it extrapolates the cumulative costs, which may be strongly influenced by one or two periods with very high costs. The appropriateness of linear extrapolation may also be affected by trends in the cost data over time. This seemed not to be the case in the current study, as the mean costs per day in the second half year were only 11% higher than in the first six months (paired samples t-test: 0.115). In any case, because linear extrapolation was based on costs during previous periods only, the method led to higher estimates of the mean costs than hot decking, predicted mean and multiple imputation, which also made use of other variables to predict costs.

Of the five methods applied in the current study, only multiple imputation is a so-called 'principled method', as opposed to the other 'naïve methods'. Therefore, from a theoretical perspective, the standard errors and confidence intervals obtained by multiple imputation were most likely to give the best representation of the actual variation in costs. The relative increase in variance because of the missing data of dropouts was approximately 30%. Despite this correction for the missing data, the estimated confidence intervals after multiple imputation were not very different from the intervals obtained with the other methods and, for instance, considerably smaller than after linear extrapolation. This effect can be explained from the higher costs that were imputed in case of linear extrapolation, predicted mean and hot decking, especially in the ipratropium group. Even a few extreme imputed values can already have a large impact on the SEs.

An important reason to apply multiple imputation in the current study is the ability of the method to impute individual resource use items instead of total costs only. In case of missing observations of a patient at a particular visit all resource use items to be imputed are drawn from the same patient with observations at that visit. Hence, the correlation structure between the resource items is being maintained. This ability of multiple imputation is an important feature of the method with regard to cost data. In almost all economic evaluations, not only total costs but also the individual resource use items like hospitalisations and outpatient visits are being reported. Moreover, the data from economic evaluations are frequently used in additional analyses like, for instance, modelling studies. These types of analyses can only be performed when per patient estimates of resource use are available. The ability of other principled methods to impute individual resource use items while maintaining the correlation structure between the items remains to be investigated.

■ 7.5 CONCLUSIONS

This study has drawn attention to the problem of missing data due to dropout. The method of dealing with these data can have a large impact on the outcomes of an economic evaluation. Information about the rate of dropout and the way data of dropouts is treated is of vital importance in assessing the results of economic evaluations and should always be reported. Multiple imputation is a principled method for dealing with the data of dropouts that provides estimates of the individual resource use items. Simulation studies are needed to determine to what extent this and other methods to deal with the data of dropouts are able to provide accurate estimates of the true mean and variance.

CHAPTER 8

THE ANALYSIS OF INCOMPLETE COST DATA DUE TO DROPOUT

8.1 INTRODUCTION

Censoring constitutes a common problem in the analysis of costs in longitudinal economic evaluations. Censoring may occur when survival and costs until death are the primary endpoints of a study and the observation period is not long enough to observe all deaths. Patients with longer survival intervals are more likely to have censored cost data than patients with shorter survival intervals. Another type of censoring may occur in studies that aim to make inference about cost over a fixed period of time, for instance one year, but in which some individuals withdraw from the study before this time has elapsed. We will refer to this latter problem as 'incomplete data due to dropout'. Especially in those situations in which the reasons to dropout are influenced by compliance, disease progression or other factors relating to patients' health, dropout may not assumed to be completely at random and the analysis of the data set may be seriously hampered.

The problem of censored cost data has been addressed in various publications. In 1997, Lin et al. introduced the product-limit estimator to analyse incomplete cost data induced by censored survival times (Lin et al., 1997). In this approach, average total costs are estimated by the sum of the Kaplan-Meier estimator for the probability of surviving to the start of each interval multiplied by an appropriate estimator for the average cost over the interval conditional on surviving to the start of the interval (Lin et al., 1997). Variants of this approach have been suggested by Bang and Tsiatis, Carides et al. and Willan et al (Bang and Tsiatis, 2000; Carides et al., 2000; O'Hagan and Stevens, 2004; Willan et al., 2003). These approaches were shown to provide unbiased estimates of total costs when survival time is subjected to censoring, but did not aim to analyse incomplete cost data due to dropout that is not completely at random.

The attention that has been paid to the specific problem of incomplete data due to dropout in relation to the analysis of costs in economic evaluations is surprisingly small. Despite the fact that dropout is likely to occur in almost every longitudinal economic evaluation, only few studies actually report the rate of dropout and the methods that have been used to analyse the data from these patients (Barber and Thompson, 1998; Briggs et al., 2003). Recently, Briggs et al. described the problems related to the analysis of incomplete resource use data and showed that simple imputation methods like complete or available cases analysis are inefficient and likely to be biased. They also describe multiple imputation methods and apply several variants of multiple imputation to two data sets with incomplete resource use data (Briggs et al., 2003). As described in the previous chapter, we applied multiple imputation in the prospective cost-effectiveness analysis of tiotropium versus ipratropium and compared the outcomes of the cost analysis with a number of single imputation methods. The study showed that the estimates of the difference in costs between treatment groups were affected considerably by the imputation methods. But, like in the study of Briggs et al, because the true complete sample estimators were unknown, this study was unable to assess the different methods in terms of bias and SEs.

In the current study we specifically address the analysis of incomplete cost data due to dropout. Other types of censoring, forced dropout because of death, and intermittent missing data are not considered. The aim of this study is to investigate how standard methods for

dealing with incomplete data perform when applied to cost data with various distributions and various types of dropout. Because of the 'simulation design' of the study, the outcomes of subsequent analyses can be compared with the known complete data set parameters and be used to give guidance to the choice for specific methods in a given situation.

8.2 METHODS

Design | The study consisted of simulations in which methods for the analysis of incomplete data were applied to artificially created samples with cost data. The simulations constituted the creation of samples, the creation of dropout in these samples and the analysis of the incomplete samples. In these simulations we varied the distribution of cost data and the underlying cause of dropout. Because of the 'simulation design' of the study, the results obtained with each method in terms of mean costs and SE could be compared with the known parameters of the complete data sets.

Methods for the analysis of incomplete data | Methods for the analysis of incomplete data can be divided into so-called naïve and principled methods. Naïve methods that were applied in this study included: complete cases analysis, mean imputation, linear extrapolation, LVCF, predicted regression and hot decking (Little and Rubin, 1987). Principled refers to methods that *'account for the missing values, and the uncertainty they introduce, at each step of the analysis in a formal way'* (Schafer, 1997). In the current study, the following principled methods were selected: the product-limit estimator of Lin et al., the EM algorithm and multiple imputation.

Naïve methods | The complete cases analysis excludes the data of all patients who withdraw from the study before the scheduled end date. Mean imputation imputes the mean of the observed values for each missing observation. In linear extrapolation, costs of the patients who dropout are extrapolated by dividing the observed costs of a patient by the number of days that particular patient remained in the study and multiplying the result by the study duration in days. LVCF imputes for each missing value the last observed value of the particular patient. Predicted mean (or regression imputation) uses ordinary least squares multiple regression analysis to impute the most likely value for each missing cost observation. Hot decking involves the selection of a limited number of categorical variables by which patients are sorted in so-called imputation classes (Little and Rubin, 1987). The intention is to create imputation classes which each contain a homogeneous group of patients. In the current study, at each time interval costs and quality of life of the previous period were both categorised into two equal-sized groups to create a total number of four imputation classes. For each patient with a missing cost observation, a value was randomly drawn from the patients who were in the same imputation class and who had complete follow-up data for that visit.

Product-limit estimator | The product-limit estimator of Lin et al. aims at estimating total costs per patient in a population where some patients may die, and where part of the observations are missing due to dropout (Lin et al., 1997). Costs are estimated by dividing the study

period into intervals. For each interval, the probability of survival until that interval and the average costs per patient during that interval, conditional on being alive at the start of the interval, are calculated. These two estimates are multiplied and summed over all intervals. This is the estimate of the average costs per patient over the study period. For the calculation of the average costs per patient in an interval, either all patients still in the study at the beginning of the interval or all patients with a complete observation for that time interval can be used. Because observed data will usually cover an entire interval we have chosen the latter approach. When no patients die during the study period, the probability of survival for each interval is one, simplifying the above procedure. The method of Lin et al. assumes that patient withdrawal is both independent of the risk of dying and independent of costs. For the calculation of the SE of the product-limit estimator we refer to the original publication (Lin et al., 1997). The input for the calculation of the product-limit estimator and associated SE can be obtained in SAS® using proc 'lifetest'.

EM algorithm | Expectation Maximisation is an iterative method to analyse data with missing values (Dempster et al., 1977). Let Y_{obs} denote the observed data, Y_{mis} the missing data and θ the parameter of interest (costs during interval t). The EM algorithm obtains an initial estimate (e.g. mean of complete cases) of θ based on Y_{obs} that is used to 'fill in' Y_{mis}. θ is then re-estimated using Y_{obs} and the filled-in Y_{mis}, and is used to re-estimate the Y_{mis}. This step is repeated until the parameter value converges. The EM algorithm assumes data to be distributed multivariate normal. The EM algorithm is facilitated by SPSS® and SAS® and results in estimates of the mean and the covariance matrix of the multivariate normal distribution. An important drawback of this method is that no principled estimate of the SE is provided. To get around this problem, we adopted two approaches. In the first, naïve, approach (referred to as the singular EM algorithm), the estimated variance is divided by the sample size. In the second approach, a bootstrap procedure was performed before application of the EM algorithm. In this approach, 1000 samples of n patients were drawn with replacement from the original sample with size n, after which each bootstrap sample was analysed by the EM algorithm. The SD of the means of the 1000 bootstrap replicates was used as an estimate of the SE of the mean cost estimate.

Multiple imputation | Multiple imputation is a technique in which each missing value is replaced by $m > 1$ simulated values (Lavori et al., 1995; Rubin, 1987; Rubin and Schenker, 1991). The m sets of imputations reflect uncertainty about the true values of the missing data. After the multiple imputations are created, m plausible versions of the complete data exist, each of which are analysed by standard complete-data methods. The results of the m analyses are then combined to produce a single result that includes uncertainty due to missing data (Rubin, 1996; Schafer, 1997). For the fractions of missing information in the current study, $m = 10$ was found to be sufficiently large to stabilise the outcomes in terms of the SE for all analyses (Schafer, 1997). The overall mean costs are simply calculated as the mean of the mean costs in each data set. The overall associated variance is found by combining the variance within data sets with the variance between data sets (Schafer, 1997). MI can be performed using

proc MI in SAS® (available in version 8.02 and higher). In this procedure, three methods are available for the imputations: propensity score method, regression and Monte Carlo Markov Chain (MCMC). The propensity method is a non-parametric approach, whereas regression and MCMC require the underlying distribution to be multivariate normal. Variables in all multiple imputation models included age at baseline and quality of life and costs at each time interval.

The propensity score method consists of several steps. First, a logistic regression is performed to predict the probability that a cost observation is missing. The outcome is called a propensity score. Next, observations are sorted according to their propensity score and divided into n (usually 5) equal-sized groups. During the third step, a random sample with replacement is drawn from the observed values within each group. This sample is called the posterior predictive distribution. Finally, for each missing observation, a value is drawn at random from this posterior predictive distribution (Rubin, 1987).

In the regression method, a regression model is fitted for each variable with missing values, with variables at baseline and from previous time intervals as covariates (Rubin, 1987). The fitted model includes estimates for the regression parameters and the associated covariance matrix. For each imputation, new values for the parameters and variance are drawn from the posterior predictive distribution of the parameters. These parameter values are used to form a predictive regression model for the missing value.

The MCMC procedure is quite similar to that in the EM algorithm (Schafer, 1997). However, the EM algorithm is deterministic and converges to a point estimate of the parameter, whereas MCMC algorithms are stochastic and converge to a probability distribution. In this case, the MCMC method converges to the posterior predictive function from which values are drawn to impute in the data set.

Simulations | The simulations constitute the creation of the complete samples, the creation of dropout in the complete samples and the analysis of the incomplete samples.

Creation of complete samples | Each sample consisted of 200 patients and ten time intervals. Each patient was assigned an age and a baseline quality of life value. In addition, patients were assigned a quality of life and cost value for each time interval. In all analyses, quality of life was assumed to be distributed multivariate normal, with higher values representing better quality of life. Costs were assigned various distributions. In the first set of samples, costs in each time interval were distributed multivariate normal. In the second set of samples, costs were distributed multivariate lognormal. In the third set of samples, costs consisted of multiple components: a cost based on a multivariate lognormal distribution and an additional cost representing the treatment of medical events. Assignment of the costs of events depended on whether patients experienced an event and whether they were hospitalised. All patients with an event without hospitalisation got assigned an additional cost that was based on a normal distribution, while patients with an event who were hospitalised got assigned an additional cost based on a lognormal distribution. Details of the creation of samples, and the distributions of and correlations between parameters are presented in appendix 8.1. We will

refer to these three sets of samples as costs that are distributed 1) multivariate normal; 2) multivariate lognormal; and 3) multivariate lognormal enlarged with costs of events.

Creation of dropout in the complete samples | Dropout was created according to three different mechanisms: dropout completely at random (DCAR), dropout at random (DAR) and informative dropout (ID). This terminology is equivalent to the more general terms missing completely at random (MCAR), missing at random (MAR) and informative missing (IM) introduced by Rubin et al. (Lavori et al., 1995). DCAR refers to the situation in which the occurrence of patients to dropout is independent from the observed and unobserved data. DAR refers to the situation in which dropout is related to data that are observed in the periods before dropout. For instance, if higher costs and/or lower quality of life are observed during time interval t, this patient may be more likely to dropout during t+1. The third type of dropout (ID) occurs if the dropout during time interval t+1 depends on the unobserved data during t+1. Typically, this occurs if the high costs and/or worsening of quality of life do not occur until after the patient has dropped out of the study.

To create DCAR, at each time interval a number of patients were randomly selected from patients still observed and from that point onwards all observed data (costs and quality of life) of these patients were deleted. The number of patients was chosen such that the proportion of patients with observed data gradually declined from 100% during time interval t=1 to approximately 70% during time interval t=10. Hence, the rate of dropout was 30% and the proportion of the total number of time intervals for which cost data were observed was approximately 84%.

The procedure to create DAR was largely similar to the procedure described for DCAR, the only difference being that patients who dropped out were no longer selected randomly. In DAR, the probability to dropout during time interval t was positively associated with costs during t-1, an increase in costs between t-2 and t-1, a decrease in quality of life between t-2 and t-1, age at baseline and negatively associated with quality of life at t-1. The association between dropout and the observed variables was verified by examination of the correlation between dropout at t and costs and quality of life at t-1. Details of the creation of DAR are provided in appendix 8.2.

The mechanism to create ID closely resembled the mechanism that was used to create DAR. Under ID, the likelihood to dropout was associated with increased costs and worse quality of life after dropout rather than before dropout. Hence, to create ID, the probability to dropout during time interval t was a function of costs during t, the difference in costs between t-1 and t, quality of life at t, the difference in quality of life at t-1 and t and age at baseline.

Analysis and outcome parameters | In the final step of each simulation, the incomplete samples were analysed using all the selected methods to deal with incomplete data. Each simulation consisted of 3000 iterations. This number appeared to be sufficient to stabilise the results, while computer time was still acceptable for even the most computationally intensive methods. One iteration involved the creation of a complete sample, the creation of dropout and

application of the selected methods. The results of the 3000 iterations for each analysis were combined and compared with the 'true' costs of the complete data set. These 'true' costs were obtained by drawing a very large number (i.e. 50000) of complete samples. In accordance with previous simulation studies, four outcome measures were used to assess the results of the analyses (Lin et al., 1997):

- Absolute and relative bias: the difference (%) in mean costs between the complete data set and the analysed samples.
- Sampling standard error (SSE) for the estimator, being the standard deviation (SD) of the mean costs of the 3000 iterations.
- Sampling average of the standard error estimator (SEE), being the mean of the standard errors of the 3000 iterations. In addition, the value of SEE relative to SSE (SEE/SSE) has been provided.
- Sampling coverage probability of the 95% confidence interval (CP), being the proportion of iterations of which the 95% confidence interval includes the 'true' mean costs.

With a sufficient number of iterations, for each analysis, the SSE can be considered as the 'true' SE. For any analysis, as to account for the additional uncertainty introduced by the incomplete data, this SSE should be larger than the SE of the complete data set. In addition, the extend to which the SEE approaches the SSE can be considered as a measure as to whether the analysis provides an adequate estimate of the SE of the data set. SEs within each sample were based on normal approximation. Verification through bootstrapping showed that SEs obtained with both approaches were almost the same.

Variations in sample size and dropout rate | The primary simulations were based on samples of size 200 and a dropout rate of approximately 30%. A sample size of 200 was chosen as this may reflect a realistic number of patients in many clinical trial-based economic evaluations and that is large enough to allow normal approximation of the SE within a sample. A dropout rate of 30% was chosen because we were primarily interested in the performance of the different methods under various distributions of costs and different dropout patterns, and we did not want the performance of methods to be disturbed only because of extreme dropout rates. To investigate the impact of dropout rate and sample size, these parameters were varied in three additional sets of simulations. In the first two additional simulations the dropout rate was set to approximately 18% and 60% respectively, while in the third additional set of simulations a sample size of 400 was used. Because of limited space, the results of these additional analyses are only presented for two situations: DAR and a lognormal distribution of costs and DAR with costs distributed lognormal enlarged with costs of events. Results of these additional analyses are presented in terms of relative bias and the value of SEE relative to SSE (SEE/SSE). An overview of all simulations included in the manuscript is presented in *table 8.1*.

■ 8.3 RESULTS

Dropout completely at random | *Table 8.2* shows the results of the simulations in case of DCAR. The first part of this table shows that when costs are distributed normal, all methods provided unbiased estimates of the mean costs. Bias did not exceed 0.03% in any analysis. As

	Cost distribution		
	multivariate normal	multivariate lognormal	multivar. lognorm. with event costs
Primary analyses[1]			
DCAR (30%; n=200)	table 8.2	table 8.2	table 8.2
DAR (30%; n=200)	table 8.3	table 8.3	table 8.3
ID (30%; n=200)	table 8.4	table 8.4	table 8.4
Additional analyses[1]			
DAR (18%; n=200)		fig 8.1	fig 8.2
DAR (60%; n=200)		fig 8.1	fig 8.2
DAR (30%; n=400)		fig 8.1	fig 8.2

Table 8.1: Overview of simulations included in the manuscript.
Figures between brackets indicate dropout rate and sample size respectively.

shown by the difference between the SSE and SEE, group means, predicted regression and hot decking underestimated the SE. The results of the simulations based on a lognormal distribution of costs were comparable to those based on a normal distribution. Bias was always below 0.2%, while group means, predicted mean and hot decking still underestimated the SE. Only in the simulations in which costs were distributed multivariate lognormal enlarged with costs of events, linear extrapolation and last valued carried forward resulted in a bias of 2.1% and 1.5% respectively. The other methods remained unbiased. Of these methods, the product-limit estimator provided the best estimate of the SE (SEE/SSE: 0.999). Deviations in the SEs obtained with multiple imputation were relatively low. The MI propensity score method slightly underestimated the SE (SEE/SSE: 0.976), while SEs obtained with MI regression (SEE/SSE: 1.026) and MI MCMC (SEE/SSE: 1.019) were slightly overestimated. The simulations also show the additional effect of applying the bootstrap EM algorithm. The singular EM algorithm underestimated the SE (SEE/SSE: 0.931), while the SEE obtained with the bootstrap EM algorithm closely resembled the SSE (SEE/SSE: 1.010).

Dropout at random | *Table 8.3* shows the simulations in case of DAR. When costs were distributed multivariate normal, only predicted regression, the EM algorithm, MI regression and MI MCMC provided unbiased estimates of the mean costs. Bias according to the other methods varied from 0.9% with the MI propensity score method to 8.5% with complete cases. Of the unbiased methods, the two MI approaches (SEE/SSE regression: 1.020; MCMC: 1.007) and the bootstrap EM algorithm (SEE/SSE: 0.993) provided almost perfect estimates of the SE, while SEs obtained with predicted regression (SEE/SSE: 0.884) and the singular EM algorithm (SEE/SSE: 0.925) were underestimated. When costs were based on a lognormal distribution, predicted regression, the EM algorithm, MI regression and MI MCMC were still unbiased. However, the value of the SEE/SSE ratio decreased for nearly all methods. Only the bootstrap EM algorithm still provided an almost perfect estimate of the SE (SEE/SSE: 0.996). In the simulations in which costs were distributed lognormal enlarged with costs of events, all methods

Distribution of costs:	normal					lognormal					lognormal with event costs				
	mean		SSE	SEE	CP	mean		SSE	SEE	CP	mean		SSE	SEE	CP
Complete sample	7028		136	136	—	8738		239	239	—	14302		944	944	—
	bias (%)					bias (%)					bias(%)				
Complete cases	0	(0.00)	166	165	0.95	-2	(0.02)	290	287	0.94	-1	(0.01)	1144	1139	0.94
Group means	1	(0.01)	147	123	0.90	0	(0.00)	258	216	0.89	0	(0.00)	1024	846	0.89
Linear extrapolation	-1	(0.01)	148	149	0.95	16	(0.18)	274	275	0.95	-297	(2.08)	1015	1021	0.91
Last value carried forward	0	(0.00)	149	150	0.95	9	(0.10)	283	282	0.95	-215	(1.50)	1105	1094	0.91
Predicted regression	1	(0.01)	144	130	0.92	0	(0.00)	256	225	0.91	3	(0.02)	1015	893	0.91
Hot decking	2	(0.03)	150	131	0.91	1	(0.01)	266	230	0.91	-5	(0.03)	1066	902	0.90
Product-limit estimator	1	(0.01)	147	146	0.95	0	(0.00)	258	257	0.95	0	(0.00)	1024	1023	0.94
EM algorithm, singular	1	(0.01)	144	136	0.94	0	(0.00)	256	238	0.93	3	(0.02)	1015	945	0.92
EM algorithm, bootstrap	-2	(0.03)	147	144	0.94	-6	(0.07)	259	256	0.95	9	(0.06)	1027	1037	0.94
MI propensity score	1	(0.01)	146	141	0.94	0	(0.00)	258	250	0.94	24	(0.17)	1029	1004	0.94
MI regression	1	(0.01)	145	147	0.95	1	(0.01)	258	264	0.95	5	(0.03)	1025	1052	0.95
MI Markov Chain Monte Carlo	1	(0.01)	145	146	0.95	1	(0.01)	257	261	0.95	4	(0.03)	1021	1040	0.95

Table 8.2: Summary statistics of 3000 simulations with sample size 200 and dropout rate of 30%; dropout completely at random.

SSE: sampling standard error, being the standard deviation of the mean costs of the 3000 iterations; SEE: standard error estimator, being the mean of the standard errors of the 3000 iterations; CP: 95% coverage probability, being the proportion of iterations of which the 95% confidence interval includes the 'true' mean costs.

showed increased bias when compared to the simulations based on a normal and lognormal distribution of costs. Bias obtained with predicted regression, the EM algorithm, MI regression and MI MCMC was around 2%, while the bias obtained with other methods ranged from 3.8% with MI propensity score to 28.3% with complete cases. The SEs obtained with MI regression (SEE/SSE: 0.859) and MI MCMC (SEE/SSE: 0.875) were considerably underestimated. The bootstrap EM algorithm still provided the best estimate of the SE (SEE/SSE: 1.047).

Informative dropout | All methods underestimated the mean costs in case of ID (*table 8.4*). In the simulations based on a multivariate normal distribution of costs, the best estimates of the mean costs were obtained with LVCF (bias 0.83%, SEE/SSE: 0.974) and linear extrapolation (bias 1.95%, SEE/SSE: 0.974). Bias according to all other methods exceeded 3.5% and strongly increased when the distribution of costs deviated further from normal. When cost data were distributed lognormal, bias obtained with LVCF and linear extrapolation increased to 1.85% and 3.4% respectively. In case of a lognormal distribution enlarged with costs of events these percentages were about 13.5%.

Variations in sample size and the proportion of dropouts | *Figures 8.1* and *8.2* show the results of the additional analyses in which we varied the rate of dropout and sample size in simulations with DAR. Figure 8.1 refers to the situation in which costs were distributed lognormal and figure 8.2 refers to the situation in which costs were distributed lognormal enlarged with costs of events. The vertical axes represent the % bias (figure 8.1a and 8.2a) and the ratio of SEE and SSE (figure 8.1b and 8.2b) respectively. The different methods are set out at the horizontal axis. For each method, there are four bars. The first bar presents the results of the primary analysis based on a dropout rate of 30% and a sample size of 200. The second and third bars are based on samples of size 200 and dropout rates of 18% and 60% respectively. The fourth bar for each method is based on simulations with a dropout rate of 30% and samples of size 400. The size of the bias and the value of the SEE/SSE ratio vary almost linearly with the rate of dropout. These figures also show that there is almost no effect of increasing the sample size from 200 to 400. Compared to the primary analysis, the percentage of bias and the value of the SEE relative to the SSE remained almost unchanged. Methods that performed best in the primary simulations were still the preferred methods in these additional simulations.

8.4 DISCUSSION

To our knowledge, this is the first study in which various methods for the analysis of incomplete cost data were assessed in a simulation study. This design allowed us to compare the results of subsequent analyses with the known parameters of the complete data set and to formally assess the performance of the different methods. It was shown that the distribution of cost data and the underlying pattern of dropout have a large impact on the performance of methods for the analysis of incomplete data. Almost all methods provided unbiased estimates of the mean in case of DCAR, but only the principled methods provided adequate estimates of the SE. The bootstrap EM algorithm, MI regression and MI MCMC provided the best

Distribution of costs:	normal					lognormal					lognormal with event costs				
	mean		SSE	SEE	CP	mean		SSE	SEE	CP	mean		SSE	SEE	CP
Complete sample	7028		136	136	—	8738		239	239	—	14302		944	944	—
	bias (%)					bias(%)					bias(%)				
Complete cases	-596	(8.48)	155	148	0.02	-1043	(11.94)	231	221	0.01	-4040	(28.25)	706	710	0.00
Group means	-316	(4.50)	143	113	0.27	-459	(5.25)	232	186	0.35	-2050	(14.33)	826	665	0.21
Linear extrapolation	81	(1.15)	147	154	0.93	530	(6.07)	324	357	0.74	827	(5.78)	1463	1414	0.94
Last value carried forward	298	(4.24)	147	163	0.55	1265	(14.48)	362	419	0.06	3528	(24.67)	1969	1922	0.59
Predicted regression	-4	(0.06)	147	130	0.92	-3	(0.03)	277	217	0.87	-292	(2.04)	1248	867	0.79
Hot decking	-182	(2.59)	148	123	0.66	-304	(3.48)	251	204	0.64	-1605	(11.22)	933	736	0.42
Product-limit estimator	-316	(4.50)	143	135	0.37	-459	(5.25)	232	219	0.45	-2050	(14.33)	826	793	0.31
EM algorithm, singular	-4	(0.06)	147	136	0.93	-3	(0.03)	277	229	0.89	-294	(2.06)	1244	910	0.81
EM algorithm, bootstrap	-4	(0.06)	150	149	0.95	-3	(0.03)	276	275	0.95	-268	(1.87)	1245	1304	0.92
MI propensity score	-66	(0.94)	149	146	0.92	-111	(1.27)	274	267	0.89	-545	(3.81)	1216	1183	0.86
MI regression	-3	(0.04)	148	151	0.95	-3	(0.03)	278	264	0.93	-286	(2.00)	1263	1085	0.87
MI Markov Chain Monte Carlo	-4	(0.06)	149	150	0.96	-1	(0.01)	281	261	0.93	-294	(2.06)	1258	1101	0.88

Table 8.3: Summary statistics of 3000 simulations with sample size 200 and dropout rate of 30%; dropout at random.

SSE: sampling standard error, being the standard deviation of the mean costs of the 3000 iterations; SEE: standard error estimator, being the mean of the standard errors of the 3000 iterations; CP: 95% coverage probability; being the proportion of iterations of which the 95% confidence interval includes the 'true' mean costs.

Distribution of costs:	normal					lognormal					lognormal with event costs				
	mean		SSE	SEE	CP	mean		SSE	SEE	CP	mean		SSE	SEE	CP
Complete sample	7028		136	136	—	8738		239	239	—	14302		944	944	—
	bias (%)					bias (%)					bias (%)				
Complete cases	-573	(8.15)	156	149	0.04	-1012	(11.58)	234	223	0.02	-3995	(27.93)	705	716	0.00
Group means	-404	(5.75)	144	112	0.11	-769	(8.80)	223	176	0.04	-3092	(21.62)	706	597	0.01
Linear extrapolation	-137	(1.95)	150	145	0.84	-298	(3.41)	296	283	0.75	-1955	(13.67)	994	970	0.45
Last value carried forward	-58	(0.83)	152	148	0.92	-162	(1.85)	317	305	0.86	-1923	(13.45)	1067	1035	0.49
Predicted regression	-250	(3.56)	146	121	0.48	-609	(6.97)	238	188	0.18	-2474	(17.30)	855	666	0.13
Hot decking	-328	(4.67)	147	120	0.28	-693	(7.93)	235	188	0.11	-2866	(20.04)	776	645	0.05
Product-limit estimator	-404	(5.75)	144	133	0.17	-769	(8.80)	223	206	0.07	-3092	(21.62)	706	707	0.03
EM algorithm, singular	-250	(3.56)	146	127	0.51	-609	(6.97)	238	198	0.21	-2474	(17.30)	855	701	0.15
EM algorithm, bootstrap	-252	(3.59)	147	151	0.72	-609	(6.97)	240	256	0.35	-2478	(17.33)	857	883	0.28
MI propensity score	-283	(4.03)	147	134	0.46	-653	(7.47)	236	213	0.20	-2591	(18.12)	846	818	0.18
MI regression	-250	(3.56)	148	139	0.57	-608	(6.96)	240	221	0.26	-2474	(17.30)	862	801	0.20
MI Markov Chain Monte Carlo	-250	(3.56)	147	139	0.57	-608	(6.96)	239	220	0.25	-2476	(17.31)	858	794	0.19

Table 8.4: Summary statistics of 3000 simulations with sample size 200 and dropout rate of 30%; informative dropout.

SSE: sampling standard error, being the standard deviation of the mean costs of the 3000 iterations; SEE: standard error estimator, being the mean of the standard errors of the 3000 iterations; CP: 95% coverage probability, being the proportion of iterations of which the 95% confidence interval includes the 'true' mean costs.

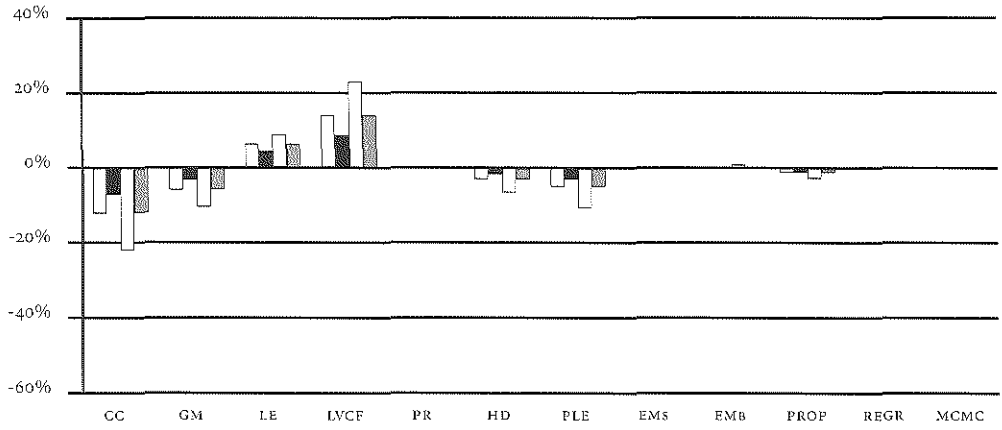
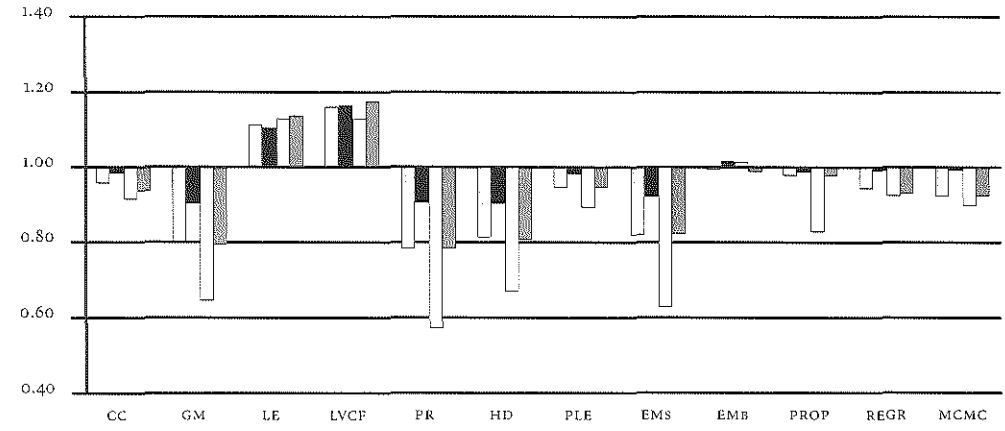


Figure 8.1: Variation in sample size and the proportion of dropouts in simulations based on DAR and lognormal distributed costs. 8.2a: Percentage bias.



8.1b: Value of SEE relative to SSE (SEE/SSE).
For each method, there are four bars: first (light-grey) bar (=primary analysis): n=200, dropout rate 30%; second (black) bar: n=200, dropout rate 18%; third (white) bar: n=200, dropout rate 60%; fourth (dark-grey) bar: n=400, dropout rate=30%. Because percentage of bias in case of PR, EMS, EMB, REGR, MCMC approach zero, they are too small to observe in figure 1a. SSE: sampling standard error; SEE: standard error estimator; CC: complete cases; GM: group means; LE: linear extrapolation; LVCF: last value carried forward; PR: predicted regression; HD: hot decking; PLE: product-limit estimator; EMS: singular EM algorithm; EMB: bootstrap EM algorithm; PROP: MI propensity method; REGR: MI regression; MCMC: MI Monte Carlo Markov Chain.

estimates of the mean and SE in case of DAR. These methods were able to deal with skewed data and only became biased when applied to costs that were distributed multivariate lognormal enlarged with costs of events. None of the methods was able to deal adequately with ID. Changing sample size or dropout rate did not substantially change the performance of the different methods.

The simulation design of the study is important to consider in relation to the interpretation of the results. It may be questioned to what extent the artificially created samples and dropout patterns reflect what will usually be found in real data sets. In real-life, the occurrence

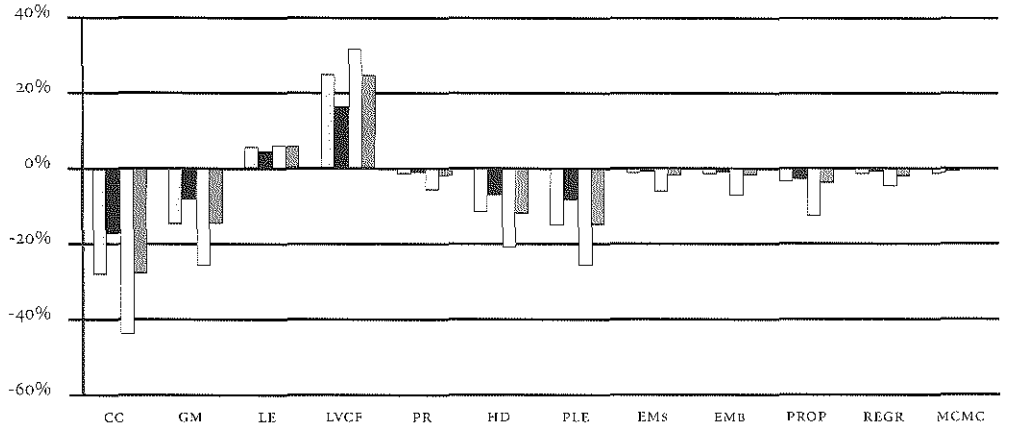
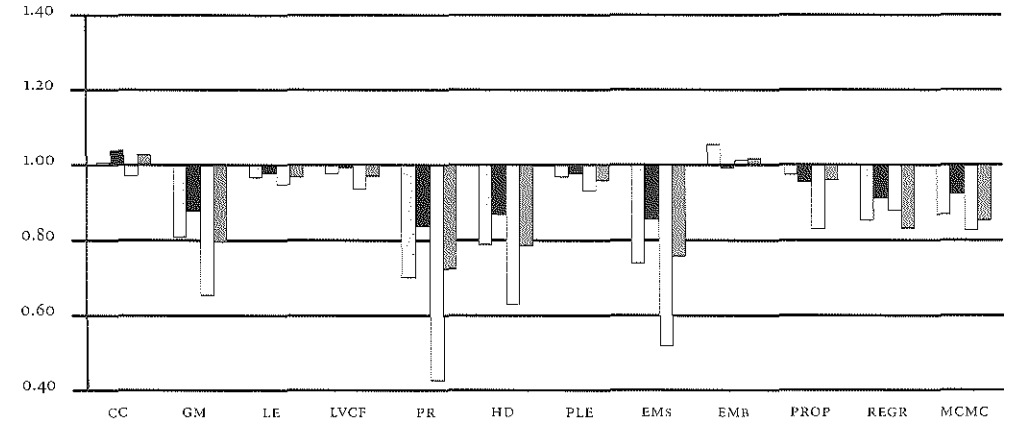


Figure 8.2: Variation in sample size and the proportion of dropouts in simulations based on DAR *and costs distributed lognormal enlarged with costs of events.*
8.2a: Percentage bias.



8.2b: Value of SEE *relative to* SSE *(*SEE*/*SSE*).*
For each method, there are four bars: first (light-grey) bar (=primary analysis): n=200, dropout rate 30%; second (black) bar: n=200, dropout rate 18%; third (white) bar: n=200, dropout rate 60%; fourth (dark-grey) bar: n=400, dropout rate=30%. SSE*: sampling standard error;* SEE*: standard error estimator;* CC*: complete cases;* GM*: group means;* LE*: linear extrapolation;* LVCF*: last value carried forward;* PR*: predicted regression;* HD*: hot decking;* PLE*: product-limit estimator;* EMS*: singular* EM *algorithm;* EMB*: bootstrap* EM *algorithm;* PROP*:* MI *propensity method;* REGR*:* MI *regression;* MCMC*:* MI *Monte Carlo Markov Chain.*

of DCAR may be rare and disease progression and the patient's condition do often play a role in the decision to withdraw from a study (see, for instance (Briggs et al., 2003) and chapter 7). The likelihood of DCAR can be assessed by, for instance, comparing dropout rates between treatment groups and by comparing the costs of the completers with the cost history of the non-completers. The likelihood of DCAR may also be assessed formally (Curran et al., 1998; Diggle, 1989). In real-life data sets, it is not possible to determine whether data are DAR or ID, because information about the condition of patients after dropout is usually not available. This may constitute a serious problem, as our analysis showed that none of the standard methods applied in this study was able to deal adequately with ID. To increase the plausibility

of data being DAR, it has been recommended to include as many as possible observed characteristics of each patient in the analysis (Gelman et al., 1995).

In order to resemble real-life cost distributions, we took many efforts as to include typical cost characteristics in our simulations. Costs were to some extent related to the condition (quality of life) of the patients and to cost data from previous periods. More important, we gradually increased the complexity of the distribution of costs. The first set of simulations was based on multivariate normal distributed data. This may not represent a common scenario and the main reason to include this scenario was to assess how different methods performed with such a distribution, facilitating the interpretation of the simulations in case of more complex distributions. The second set of simulations, based on lognormal (skewed) distributions, is much more likely to represent common real-life cost data. For instance, Briggs et al. analysed the distribution of costs of five data sets. All cost distributions were skewed and a log transformation was found to be the best approach to normalise the data in the majority of data sets (Briggs and Gray, 1998). In the third (and most complex) set of simulations costs consisted of multiple components: a cost based on a multivariate lognormal distribution and an additional cost representing the treatment of medical events. In our experience, this mixed distribution represents a real world scenario that may be common in many chronic diseases. For instance, in chapter 6 a distinction was made into costs of COPD maintenance therapy, non-hospital costs of exacerbations and hospital costs of exacerbations. Additional analyses of these COPD data showed that costs of maintenance treatment, non-hospital costs of exacerbations and hospital costs of exacerbations all approached a lognormal distribution signifying that complex real-life cost data can be simulated by combining data sampled from various parametric distributions. When applied to such complex distributions in case of DAR, none of the methods was still able to provide unbiased estimates of the mean and SE. Apparently, it is not skewness, but mixed or bimodal distributions and large variation in costs within a patient over time that constitute the real obstacles for the adequate analysis of incomplete cost data. In data sets with such complex mixed distributions and with substantial and different dropout rates between treatment groups the applicability of standard methods to analyse the incomplete data should be questioned.

Our study clearly shows that when the mean and the associated uncertainty are considered the performance of all naïve methods is poor. Unbiased estimates may incidentally be obtained, but none of the naïve methods provided results that were consistently unbiased over analyses. This for instance was the case for LVCF. LVCF outperformed all other methods in case of ID, but was among the worst performers in case of DCAR and DAR. The small bias in case of ID may specifically have been due to the mechanisms to create costs and dropouts in our samples. The poor performance of the naïve methods is especially important because of the suspicion that these methods are most commonly used to overcome the problem of missing data in economic evaluations (Briggs et al., 2003). Even small (and opposite) biases in the estimates of total costs may lead to a considerable bias in the estimates of the difference in costs between treatment groups. Hence, one should be cautious to apply naïve methods, and these can only be used in situations in which the assumption of DCAR is justified and the dropout rate is low and comparable between treatment groups.

An important criterion for the selection of principled methods was whether a particular method could be considered as a 'standard method'. The criterion of 'standard method' implies that a method can be applied (available procedure or easily programmable) using standard software packages like SPSS®, SAS® or STATA®. Several papers present tailor made methods, pertinent only to the type of data under study (see for instance (Cook, 1997; Hogan and Laird, 1997; Matsuyama and Ohashi, 1997)). Another category of methods that were excluded were the generalised linear mixed models (GLMM) and it variants generalised estimating equation (GEE) and pattern-mixture models (Beacon and Thompson, 1996; Little and Wang, 1996; Touloumi et al., 2001). Mixed models are generalisations of the general linear model (i.e. multiple linear regression) to repeated measurements of the dependent variable by taking account of the correlation between measurements from the same subject and are known to handle data that are MAR (Little, 1995). However, these methods require that missing values must be confined to the single response variable as they cannot deal with missing values in multiple response variables or in additional covariates or predictors (Schafer and Yucel, 2002). In this simulation study we addressed the specific problem of dropout. The occurrence of dropout at a certain time-interval implied that from that time interval onwards, all quality of life and cost observations in the data sets were made missing. Hence, the application of mixed models in this study is not straightforward and it may need further study to determine whether mixed models can be applied to such data sets. Advanced solutions to this problem, as discussed by Schafer and Yucel (Schafer and Yucel, 2002), clearly are beyond the scope of this study in which only standard methods for the analysis of incomplete data were considered. Finally, in later extensions of the product-limit estimator Lin introduced a regression-based approach that allows the inclusion of covariates potentially enabling the analysis of DAR as long as the included covariates can predict the probability that the data becomes missing (Lin, 2000). However, it is not straightforward how this method should be adjusted to incorporate costs and quality of life data of previous periods, as covariates and the appropriateness of the method to be used under DAR should further be explored.

In this study, we only analysed total costs per patient per time interval, and did not consider individual resource use items. Using resource use data may further complicate the analysis considerably. The distribution of each resource use item separately is likely to be much more extreme than the distribution of total costs. Resource use frequently consists of count data and is often characterised by many 'zeros', i.e. with only a few patients with resource use. Attempts in this study to apply predicted regression and the multiple imputation approaches to the base and event costs separately (in simulations based on lognormal distributed costs enlarged with costs of events) indicate that none of these methods were able to deal adequately with the event costs. Considering the good performance of the multiple imputation approaches, it is interesting to further explore multiple imputation in other software packages that enable to specify a distribution for each variable in the analysis (van Buuren and Oudshoorn, 1999).

8.5 CONCLUSIONS

Incomplete data due to dropout constitute a significant problem to economic evaluations that has received only little attention to date. The distribution of the data and the underlying cause of dropout are the most important factors to consider with regard to the analysis of these data. The bootstrap EM algorithm, MI regression and MI MCMC are the preferred methods for the analysis of incomplete cost data. These methods are able to deal with cost data in case of random and completely random dropout and are robust to the 'multivariate normal assumption'. This implies that these methods still provide almost unbiased estimates of the mean and SE when costs are severely skewed. In case of DCAR, also the product-limit estimator can be applied. In data sets characterised by high costs of events and a large variation in costs within patients over time in combination with substantial dropout that is not completely at random, the application of standard methods to analyse the incomplete data may not be appropriate.

Appendix 8.1: creation of samples

Step 1: create patients | Three thousand samples are created with 200 patients each.

Step 2: assign age | A value is assigned to *age* which is normally distributed: $age \sim N(55.5; 100)$ truncated at 30 and 80.

Step 3: create quality of life (qol) variables | A value is assigned to *qol* for each time interval, using a multivariate normal distribution. Qol_1 is correlated to *age*, and qol_i ($i = 2,\ldots,10$) is correlated to qol_{i-1}. The parameters of the distribution are chosen such that the mean and standard deviation at each time interval will be approximately 60 and 15 respectively, and that the correlation of qol between intervals is approximately 0.75. All qol values are truncated at 0 and 100:

$qol_1 = X \times ((1 - 0.005 \times age) / 0.725)$ with $X \sim N(60; 225)$ and

$qol_i \sim N(mu_{qoli}, var_{qoli})$ with $mu_{qoli} = 60 + 0.75 \times (qol_{i-1} - 60)$ and

$var_{qoli} = 15^2 \times (1 - 0.75^2)$ for $i = 2,\ldots,10$.

Step 5: creation of cost data | Costs were assigned various distributions. In the first set of samples, costs in each time interval were distributed multivariate normal. In the second set of samples, costs were distributed multivariate lognormal. In the third set of samples, costs consisted of multiple components: a cost based on a multivariate lognormal distribution and an additional cost representing the treatment of medical events.

Multivariate normal distribution of costs | A cost value is assigned to each patient for each time interval using a multivariate normal distribution. $Costs_1$ is correlated to qol_1 and $costs_i$ ($i = 2,\ldots,10$) is correlated to qol_i and $costs_{i-1}$. The parameters of the distribution are chosen such that the mean and standard deviation at each time interval will be approximately 700 and 250 respectively and that the correlation between $costs_i$ and qol_i is approximately - 0.35 and between $costs_i$ and $costs_{i-1}$ approximately 0.70. All cost values are truncated at a minimum value of 1:

$cost_1 \sim N(700 - 5 \times (qol_1 - 60); 300^2)$

$cost_i \sim N(mu_{costi}, var_{costi})$ with $mu_{costi} = 700 + 0.66 \times (cost_{i-1} - 700) - 3.5 \times (qol_i - 60)$ and $var_{qoli} = 250^2 \times (1 - 0.55^2)$ for i = 2,...,10

Multivariate lognormal distribution | A cost value is assigned to each patient for each time interval using a multivariate lognormal distribution. $Costs_1$ are correlated to qol_1 and $costs_i$ (i = 2,...,10) is correlated to qol_i and $costs_{i-1}$. The parameters of the distribution are chosen such that the median at each time interval is approximately 700 and that the correlation between $costs_i$ and qol_i is approximately - 0.30 and between $costs_i$ and $costs_{i-1}$ approximately 0.50. All *cost* values are truncated at a minimum value of 1:

$cost_1 \sim$ lognormal$((720 - 10.8 \times (qol_1 - 60)), 0.65)$

$cost_i \sim$ lognormal$((683 + 0.38 \times (cost_{i-1} - 700) - 7.4 \times (qol_i - 60)), ((0.6^2) \times (1 - (0.4^2)))^{0.5})$ for i = 2,...,10.

The first parameter specifies the median of the costs, and the second the standard deviation of the log(costs).

Multivariate lognormal distribution enlarged with costs of events | A cost value is assigned to each patient for each time interval using a multivariate lognormal distribution, increased with the costs of a medical event that may or may not lead to hospitalisation. Starting point is the assignment of costs distributed multivariate lognormal as described above. The assignment of the event costs consist of three steps:

1) *assignment of events*

The probability to experience an event during a time interval is related to the quality of life at that time and whether an event was already experienced previously. Parameters were chosen such that the probability to experience an event increased from approximately 0.40 in period 1 to 0.75 in period 10. The variable $prev\_ev_i$ keeps track of previous events, and is 1 if an event occurred at any previous time interval. Furthermore:

$p\_event_i \sim$ uniform$(0, (0.8 - (0.40 \times (qol_i / 60)) + (0.15 \times prev_ev_i)))$,

if $(p\_event_i > 0.45)$ then $event_i = 1$.

2) *assignment of hospital admissions*

Given that an event occurred, the probability that a patient will be hospitalised is 0.70.

3) *assignment of event costs*

Costs of an event without hospitalisation are normally distributed $N(1250, 400^2)$.

Costs of an event with hospitalisation are lognormally distributed with median 12000 and standard deviation of the logcost of 0.4.

■ APPENDIX 8.2: CREATION OF DAR AND ID IN THE COMPLETE SAMPLES

To create DAR, a score variable is calculated for each patient with observed data for that time interval. This score (S) is a function of the standardised values of cost at t-1, the difference in costs between t-2 and t-1, quality of life at t-1, the difference in quality of life (qol) between t-2 and t-1, age and a random value drawn from the standard normal distribution to create 'noise':

$s_t = cost_{t-1} + (cost_{t-2} - cost_{t-1}) \times 0.25 - qol_{t-1} - (qol_{t-2} - qol_{t-1}) \times 0.25 + age \times 0.5 + N(0,1) \times 0.25.$

All variable names refer to the standardised variables. Standardisation of variables is obtained by taking the difference between an observation and the mean of the variable, divided by the standard deviation. The score variable s_t is used to rank order patients with observed data. At each time interval, a random sample with size x is drawn from the y patients with the highest values of s_t. Cost and quality of life values of these x patients are made missing at t and all subsequent time intervals. The values of x and y depend on sample size and the proportion of dropouts. With sample size n = 200 and 18%, 30% and 60% dropout, x takes the value of 4, 7 and 13, whereas y takes the value of 10, 20 and 40 respectively.

The procedure to create ID was almost identical to the procedure to create DAR. To create ID the formula above is rewritten into:

$s_t = cost_t + (cost_{t-1} - cost_t) \times 0.25 - qol_t - (qol_{t-1} - qol_t) \times 0.25 + age \times 0.5 + N(0,1) \times 0.25.$

Acknowledgements

The study was supported by a grant from the Netherlands Organisation for Scientific Research (NWO) as part of the Health Care Research Efficiency Program 2001.

CHAPTER 9

GENERAL DISCUSSION AND CONCLUSIONS

9.1 INTRODUCTION

This thesis is based on various publications that relate to the calculation and analysis of costs in economic evaluations. This chapter addresses the four research areas identified in chapter 1. These are the standardisation of costs in relation to the comparability of economic evaluations, the generalisability of the results of economic evaluations to other countries, the impact of expensive medical events on the analysis of costs, and the analysis of incomplete cost data due to premature withdrawal. The first section of this chapter discusses the findings with regard to the economic evaluation of tiotropium in patients with COPD.

9.2 ECONOMIC EVALUATION OF TIOTROPIUM

Chapters 4 and 5 show two complementary approaches to cost-effectiveness analyses. The first approach (chapter 4) is a cost-effectiveness analysis embedded in a clinical trial. It was found that the mean (SE) number of exacerbations over one year was reduced from 1.01 (0.10) in the ipratropium group to 0.74 (0.08) in patients treated with tiotropium, while the numbers of patients with a relevant improvement on the SGRQ increased from 34.6% in the ipratropium group to 51.2% in the tiotropium group. In addition, tiotropium was associated with increased costs of € 180 (228). An advantage of this approach to economic evaluations is that all outcome measures in terms of exacerbations, quality of life and use of medical resources are based on patient-level data of the same population. Other advantages include the randomised and controlled experimental design, and several guidelines have expressed a clear preference for economic evaluations that are performed alongside clinical trials (Jacobs et al., 1995; Langley, 1996). However, the results of trial-based economic evaluations only apply to the country in which they are performed and results may not be generalised to other settings. In the economic evaluations in chapter 4, about 85% of the patients were treated in the Netherlands and these results cannot be applied to other countries without modifications. Another drawback of this approach is that the primary aim of nearly all RCTs is to determine the safety and efficacy of a new treatment and economic outcomes often remain secondary. This situation may hamper the adequate estimation of resource use and costs in a trial situation. The trial-based cost-effectiveness analysis of tiotropium showed that scheduled trial visits and the instruction to physicians to keep the dose of concomitant medication constant throughout the trials, may have biased the proper estimation of maintenance costs.

The second approach to cost-effectiveness analysis was demonstrated in chapter 5 in which we constructed a decision-analytic model to assess the costs and effects of different bronchodilators, using data from the clinical trial program of tiotropium. In this analysis, the mean (SE) number of exacerbations over 1 year was reduced from 1.14 (0.13) in the ipratropium group, to 1.02 (0.10) in the salmeterol group and to 0.85 (0.03) in the tiotropium group. In addition, for the Netherlands, model-based estimates of the mean costs of patients in the tiotropium group were € 170 (197) lower than the costs of patients in the ipratropium group and € 42 (210) lower than the costs of patients in the salmeterol group. In Canada, costs were consistently lower than in the Netherlands and nearly the same in all treatment groups. The analysis in chapter 5 clearly shows the additive value of a model-based economic evaluation. Firstly, the model was used to determine the costs and effects of tiotropium in

other countries. In this model, resource use estimates were required for disease severity states and for exacerbations, and these estimates were collected from a variety of sources, either inside or outside the trials. The ability to obtain estimates of the cost-effectiveness of tiotropium in different countries was, in fact, the primary aim of the model and is further discussed in section 9.4. Secondly, the model was built to compare tiotropium with two other bronchodilators, salmeterol and ipratropium, within the same framework by combining data from different trials. Because we restricted ourselves to the phase III clinical trial program of tiotropium, the design and patient populations (i.e. the inclusion and exclusion criteria) in these trials were highly comparable. Nevertheless, it appeared that the trials in which tiotropium was compared to salmeterol, showed higher exacerbation rates than the other trials, a finding that could not be explained from characteristics of the patient population or the duration of the trials. In the base case analysis of the Markov model, exacerbation rates were based on the combined data from all trials, Consequently, model-based estimates of the numbers of exacerbations were higher than found in the prospective economic evaluation and, because of the high costs of (severe) exacerbations, overall treatment costs in the model were also higher than estimated in the prospective economic evaluation. To account for the uncertainty about the true rate of exacerbations, chapter 5 contained various additional analyses in which the sensitivity of outcomes to the rate of exacerbations was further explored. Finally, by assigning real-life estimates of resource use to maintenance treatment and exacerbations, estimates of costs in the model-based economic evaluation were adjusted for protocol-driven costs. These adjustments mainly relate to the costs of maintenance treatment. In the prospective economic evaluation, scheduled trial visits were excluded from the cost calculations and physicians were instructed to keep the dose of concomitant medication constant throughout the trials. In the model, costs of maintenance therapy were assigned to disease states, and costs were assumed to increase with disease severity. Another adjustment that was made relates to the daily costs of ipratropium. In the prospective economic evaluation, the daily cost of ipratropium was based on administration through the metered dose inhaler. This was the inhaler device that was used in all clinical trials. In the Markov model, the daily cost of ipratropium was based on the average price of the metered dose inhaler and the (more expensive) dry powder inhaler, weighted by the actual use of these devices in the Netherlands.

The results of these economic evaluations raise the question as to how to interpret the outcomes of these analyses. As bronchodilator treatment was not associated with substantial improvements in quality adjusted life, it is not possible to compare our findings with the results of economic evaluations performed in other diseases. The interpretation of the outcome measure 'cost per exacerbation avoided' is not straightforward and information about what would be an acceptable ratio is not available. If, in the future, more information about the cost-effectiveness of COPD treatments becomes available, a comparison with these studies may provide further insight. Considering the outcomes of the model-based analysis, in which tiotropium was associated with cost savings when compared to ipratropium and at least cost neutral when compared to salmeterol, and in which tiotropium was associated with maximum expected net benefit for nearly all values of the ceiling ratio, the conclusion that

tiotropium has acceptable cost-effectiveness seems to be justified. Uncertainty about the outcomes mainly relates to the exacerbation rates of the different treatments and, in order to reduce uncertainty, future research should concentrate on this outcome. In future analyses we will also extend the time frame of the model beyond the one-year study period in order to determine the long-term cost-effectiveness of tiotropium.

9.3 STANDARDISATION AND COMPARABILITY OF COSTS IN ECONOMIC EVALUATIONS

In chapter 2 we discussed some of the key issues in relation to the standardisation of costs and how each of these could be used to contribute to the comparability of economic evaluations. Key issues that emerged were the use of similar basic principles, methods for measurement and valuation, standard costs, standard values and the presentation of results. The cost analysis in the economic evaluation of tiotropium was largely conducted in accordance with the recommendations from the manual, and several aspects of the key issues were applied in our study to safeguard the comparability with other economic evaluations. Especially with regard to the valuation, we carefully balanced the relative contribution of the cost item to the total and incremental costs and the precision of the estimation of unit costs. Unit cost estimates of resources with limited contributions to total costs were based on hospital reimbursement fees (laboratory tests) or standard costs (GP and physiotherapist visits), whereas unit costs of inpatient days and outpatient visits to the pulmonologist were based on unit cost calculations in several hospitals. A total of 7 internal and pulmonary wards and outpatient clinics were selected for the calculation of unit costs. All of these sites were included in the costing study described in chapter 3. In addition, the time physicians spent per inpatient day or outpatient visit was based on a survey among 30 pulmonologists participating in the clinical trials of tiotropium. Many of the standard values mentioned in chapter 2 were used in the unit cost calculations.

The economic evaluation of tiotropium also shows that it may be difficult to cover all the basic principles that are mentioned in the Dutch manual and pharmacoeconomic guidelines. Major deviations from the basic principles in the primary analysis of the economic evaluation of tiotropium relate to the time span of the study and the failure to include costs outside the healthcare sector. In accordance with the duration of the clinical trials comparing tiotropium to ipratropium, the time span of the economic evaluations in chapter 4 and 5 was restricted to one year. For a chronic and largely irreversible disease like COPD, this period may be too short to incorporate all relevant costs and effects. Currently, we are constructing a five-year model to assess the long-term costs and effects of tiotropium compared to other bronchodilators. However, it may be difficult to populate the model with five-year data, as data about the costs and effects of tiotropium after one year are not available. This constitutes a common problem in pharmacoeconomic evaluations of chronic diseases, where long-term assessment of a new drug may be required to obtain regulatory approval, while data on the long-term costs and effects are not available. Prolongation of the clinical trials may be unethical when the availability of potential effective medicines is further postponed. In addition, prolongation of the trial periods may meet specific problems, such as increasing

dropout rates that hamper the interpretation of the long-term results. Hence, it may be unavoidable to base decisions about registration and reimbursement of new medications on studies with time spans that are not sufficient to incorporate all relevant costs and effects. Long-term modeling studies may assist in decision-making by evaluating the most likely scenarios.

In addition to the direct healthcare resource use, also data about absence from work and the number of days patients were unable to perform usual daily activities were collected in the ipratropium-controlled trials. Compared to patients in the ipratropium group, the number of days that patients were unable to perform their usual daily activities, including paid work, was 18% less in the tiotropium group. Because only small and different proportions of patients in each treatment group had a paid job, and because comparable information about inactivity days before start of treatment was not available, these indirect costs were not included in the cost-effectiveness analysis.

It would be interesting to know whether the Dutch manual for costing has led to more standardization of costs and a better comparability of studies performed in the Netherlands. It would also be interesting to know which of the key issues as identified in chapter 2 contributes to the standardisation of economic evaluations in practice. These questions cannot be answered without a review of Dutch economic evaluations, assessing all different aspects of the costing methodology and such a review has not been performed. Available data show that, among the key issues for standardisation, standard costs are most often used and that they may have the largest potential to further standardise costs in economic evaluations. In 2003 a formal evaluation of the manual among its users left no doubt that the ability to obtain standard costs that could directly be applied to value costs in economic evaluations was one of the main reasons for its use. The evaluation also showed that recommendations of users for future versions of the manual most often referred to an improvement or further extension of the set of standard costs. In a search for Dutch papers published between January 2002 and April 2004, I identified 80 economic evaluations. The proportion of papers that referred to the manual gradually increased from 53% (16 out of 30) in 2002 to 57% (21 out of 37) in 2003 and 69% (9 out of 13) in 2004. In the majority of studies, the use of standard costs was the main reason for reference to the manual. Torrance et al. have also argued that, '*to the extent possible, standard cost values should be used in costing out the utilisation of resources*' (Torrance et al., 1996). They further argue that high priority should be given to the development of such a list of standard costs and that this list should be considered mandatory (Torrance et al., 1996). Whether the use of standard costs is appropriate depends on the quality and availability of these standard costs, but also relates to the aim of the study. The use of standard costs becomes more appropriate in studies at a higher level of aggregation that aim to generalise the results to other settings. Hence, in pharmacoeconomic evaluations that aim to contribute to decision making about reimbursement of new medications at a national level, the use of standard costs may become the method of choice for the valuation, and is to be preferred above a valuation that is based on detailed, setting-specific cost calculations. In 2004 a new version of the manual will be published in which the list with standard costs will be extended with unit costs of a large number of inpatient medical procedures. This extended

list with standard costs may further improve the standardisation and comparability of costs in Dutch economic evaluations.

9.4 GENERALISABILITY OF ECONOMIC OUTCOMES ACROSS COUNTRIES

An increasing number of countries now require economic data to support reimbursement decisions for newly registered medications. Authorities require these data to represent country-specific estimates of the costs and effects and it is usually not sufficient to submit economic results from another country or results that are based on some multinational average. Given that these cost-effectiveness data are often required in countries shortly after regulatory approval and that sufficient clinical and economic data are rarely available in all these countries, investigators are increasingly faced with the challenge of generating cost-effectiveness estimates for different settings where no prospective cost-effectiveness study can be performed. This problem is acknowledged in the pharmacoeconomic guidelines of an increasing number of countries. Several of these guidelines acknowledge the fact that transferring data across countries may sometimes be inevitable. Further, they mention the need to adapt the pharmacoeconomic data to the local setting, and request as much transparency as possible (Canadian Coordinating Office for Health Technology Assessment, 1997; Riteco et al., 1999). However, none of these guidelines give practical guidance as to how these adaptations should be performed.

In the health economics literature, several approaches to deriving country-specific cost-effectiveness estimates from economic evaluations performed in other countries have been applied (Pang, 2002). Many of these are simple approaches in which, for instance, multinational resource use data are valued against country-specific estimates of unit cost. These approaches do not take account of potential differences in resource utilisation between health-care systems in different countries and the value of the outcomes can be questioned. More advanced approaches to transfer the results of cost-effectiveness analyses across countries are those based on regression analysis. The most elaborated example of this approach is provided by Willke et al. (Willke et al., 1998). They *'examine how clinical and economic outcomes interact when estimating treatment effects on costs and proposes empirical methods for capturing these interactions and incorporating them when making country-specific estimates'* (Willke et al., 1998). However, this methodology can only be applied to countries that participate in multinational trials with sufficient sample size per country. As insufficient sample size per country is often the reason to conduct a multi-country analysis, the value of this approach to multinational economic evaluations in practice is limited. Given the limitations of these approaches, Greiner et al. stated that *'if permitted by the study design and the subject of the investigation, the decision analysis approach should be the method of choice for transferring data to foreign health-care systems'* (Greiner et al., 2000).

In chapter 5, a decision-analytic Markov model was constructed with the specific aim of estimating the cost-effectiveness of bronchodilator therapy in COPD patients in different countries. Considering this aim, several characteristics of the model are worth mentioning. A first characteristic of the model is its probabilistic nature, as opposed to deterministic models, that enables the analysis of uncertainty around the estimates of costs and effects, and the

resulting ratios. The importance of adequately describing the uncertainty of the results of economic evaluations will be further discussed in section 9.5. The second characteristic is that all model inputs related to the effectiveness of treatment are based on patient-level trial data. Because RCTs provide the highest level of evidence for the efficacy of new medications compared to existing treatments, this means a strong support for models that are directly based on these data. A direct consequence of being trial-based is that it provides the opportunity to validate the model with the clinical trial data. Validation of models is strongly recommended (Weinstein et al., 2003), but often difficult to realize because necessary data are not available (McGuire and Morris, 2000). In chapter 5 we showed that the outcomes of the model in terms of exacerbations closely resembled the results of the clinical trials. A major concern of decision-makers with program or regulatory responsibility relates to the transparency of the model and the possibilities model constructors may have to manipulate the outcomes. The ability to compare some of the major outcomes of the model with the original trial data may overcome these concerns and may thereby increase the acceptance of models by local reimbursement authorities. A final characteristic of this model is the assumption that, given disease state and the presence or absence of an event, estimates of resource use and unit costs do not differ between treatment arms. Hence, no assumptions have to be made about differences between treatment arms in the cost per event or the maintenance cost per disease severity state. All differences between treatment groups are accounted for by differences in transition and event probabilities. This characteristic can be considered as a necessary requirement for models that aim to estimate cost-effectiveness in different countries or settings and allow resource use to be estimated from a variety of sources and settings inside or outside the trials.

■ 9.5 ANALYSIS OF EVENT-DRIVEN COSTS

Chapter 6 provides an analysis of the costs of exacerbations as observed in the trial-based economic evaluation of tiotropium versus ipratropium in chapter 4. In this post-hoc analysis, we determined whether costs were associated with exacerbations or with maintenance treatment of the disease. It was found that exacerbations contributed to approximately 34% of the total respiratory-related healthcare costs. In addition, it was found that 16% of the exacerbations that were associated with a hospitalisation accounted for 90% of the total costs of exacerbations. Chapter 4 also showed that, except for the costs of study medication, the difference in costs between tiotropium and ipratropium was almost entirely due to a difference in costs associated with exacerbations.

Despite the high costs of events and the resulting skewed distribution of costs, the sample size in the trial-based cost-effectiveness analysis of tiotropium was large enough to apply common methods to analyse the cost data. Bootstrapping and normal approximation resulted in almost identical confidence intervals of the difference in costs. Costs excluding study medication in the tiotropium group were found to be € 273 lower than in the ipratropium group, a reduction of almost 20%. Despite this considerable difference, the sample size was not sufficiently large to detect statistically significant differences in costs. A sample size of approximately 1570 patients per treatment group would have been required to detect a difference in

costs of € 250, with a type I error of 5% and a power of 80%. However, the statistical analysis of the outcomes in economic evaluations focuses on describing uncertainty rather than hypothesis testing. In chapters 4 and 5 we put current principles regarding the analysis of uncertainty in economic evaluations into practice. In chapter 4 we present the incremental cost-effectiveness of tiotropium versus ipratropium on the CE plane (figure 4.1) and the CE acceptability curve (figure 4.2). The CE planes show the ellipses containing 5%, 50% and 95% of the probability density of the difference in costs and effects, facilitating a rapid and visual interpretation of the uncertainty around the cost-effectiveness ratio. The CE acceptability curve shows the probability that a treatment is cost effective for different values of the ceiling ratio. In chapter 5, we compared three treatments within the same framework (figure 5.3). Based on the assumption that these treatments are, in principle, mutually exclusive, we present separate curves for each treatment instead of incremental curves for each mutual comparison. The acceptability frontier follows the curve of the treatment with the highest expected net benefit, hence determining the optimal treatment from a cost-effectiveness perspective for each value of the ceiling ratio. The presentation of the results of the sensitivity analysis of the model-based economic evaluation in table 5.4 fits well into this approach to uncertainty analysis. For each sensitivity analysis, this table shows for which values of the ceiling ratio each treatment is deemed to be cost effective. Such a presentation of the results of sensitivity analyses may be more relevant to decision makers than the separate presentation of differences in costs and effects as is commonly the case. The sensitivity analyses in chapter 5 clearly showed that exacerbations were the main determinant of differences in cost-effectiveness between treatments. If, for instance, similar exacerbation rates were assumed for all treatments, tiotropium was associated with the maximum expected net benefit, only for very large values of the ceiling ratio.

Skewness may not be the only problem of data characterised by high costs of clinical events. Costs may also be characterised by a bimodal distribution and a large variation in costs within the same patient over time. Patients experiencing a clinical event may accrue very high costs in one time interval, while costs in preceding time intervals did not substantially differ from patients without clinical events. In this thesis, this problem became apparent in relation to the analysis of incomplete data as discussed in chapter 8. It was shown that the bootstrap EM algorithm, MI regression and MI MCMC were robust to deviations from the normal distribution and still provided unbiased estimates of the mean costs in case of lognormal (skewed) distributions and DAR. When high costs of events were added, even these methods consistently underestimated the mean costs. Apparently, these methods were unable to take account of the high costs of events during the time intervals patients were no longer observed. This even was true despite data being DAR and the situation that the likelihood to experience an event was associated with higher costs and worse quality of life in preceding time intervals. The inability to deal with incomplete follow-up data characterised by high costs of events emphasises the importance to collect data of all patients for the entire study period. Because high costs of events are often due to hospital admission, it might be considered whether it is possible to collect hospital resource use data from the hospital records. In that case, resource use of the main cost drivers and events may even be collected after patient withdrawal.

Because of the impact of exacerbations on costs, they are a major determinant of cost-effectiveness in COPD. Important factors to consider are the definition of exacerbations and exacerbation severity, the measurement of resource use associated with exacerbations, and the relationship between these factors. It is difficult to identify the start and end of COPD exacerbations due to the fluctuation of symptoms (Rodriguez-Roisin, 2000); comorbidities also complicate the definition. Attempts to define exacerbations have focused on symptoms like increased dyspnoea, sputum production and sputum purulence (Anthonisen et al., 1987). In later definitions, exacerbations were not only defined in terms of the patient's condition or symptoms, but were also assumed to necessitate a change in regular medication (Rodriguez-Roisin, 2000). Clinical trials in COPD have used many varieties of this latter definition. Definitions of exacerbation severity have also frequently been related to the way exacerbations were treated. In these definitions, severe exacerbations are usually associated with hospitalisation or a visit to an emergency department, moderate exacerbations are associated with an outpatient visit to a GP or respirologist, whereas mild exacerbations are usually self-managed or associated with a change in regular medication. These 'working definitions' of exacerbation and exacerbation severity have the advantage that they may be easy to apply by physicians participating in RCTs. However, the treatment of exacerbations is unlikely to be the same across centres and countries, and an exacerbation that is classified as severe in one centre, may be classified as moderate in another centre. Similarly, physicians in different countries may have different thresholds before they prescribe antibiotics or oral steroids to a patient. Hence, the use of different and/or treatment-based definitions of exacerbations in clinical trials may lead to exacerbation rates that are incomparable across studies, and differences may reflect differences in treatment patterns rather than actual differences in the number of exacerbations. The incomparability of studies also considerably complicates the construction of decision-analytic models, because exacerbation rates and resource use associated with exacerbations are often derived from multiple sources. In order to avoid this problem in the model-based economic evaluation of tiotropium, we adopted definitions of exacerbations and exacerbation severity that were solely based on respiratory symptoms and the patient's condition. These definitions were available in all trials that were part of the clinical trial program of tiotropium. Hence, treatment differences between centres or countries have not affected the estimates of the rate of exacerbations in these trials and our model.

9.6 ANALYSIS OF INCOMPLETE COST DATA DUE TO PREMATURE WITHDRAWAL

In the empirical cost-effectiveness analysis in chapter 4, we applied the MI propensity score method to analyse the data of patients with incomplete follow-up. As far as we know this is the first empirical economic evaluation in which this method has been applied. Because it was unknown whether the use of this method was appropriate for analysing incomplete economic data, we extensively explored the impact this method had on the outcomes of the analysis and compared the results with those of various non-principled methods. The analysis in chapter 7 showed that patients with incomplete follow-up were largely responsible for the difference in costs between tiotropium and ipratropium. It was also shown that the point estimate

and the 95% CI of the difference in costs between these treatments were affected considerably by the different methods. Of the methods applied, the results obtained with multiple imputation were at least as conservative as obtained with hot decking, linear extrapolation and predicted mean. Because only multiple imputation takes account of the uncertainty due to missing data in a formal way, we felt safe to adopt the outcomes based on the MI propensity score method as our base case analysis.

Our findings in the economic evaluation of tiotropium were an important motivation to further explore the area of incomplete data due to premature withdrawal. In the simulation studies in chapter 8, we not only applied the MI propensity score method, but also other principled methods like the product-limit estimator, the EM algorithm, MI regression and MI MCMC. These simulation studies showed that some of these methods, like MI regression and MI MCMC, provided more consistent estimators of the mean costs under DAR than the propensity score method. Had we known these results before the analysis of the economic evaluation of tiotropium we might have considered the use of one of these methods instead. However, as outlined in the discussion in chapter 8, the distribution of costs in the tiotropium data set most closely resembled the simulations in which costs were 'distributed multivariate lognormal enlarged with costs of events'. When applied to these simulated data sets, MI regression and MI MCMC did not provide unbiased estimates of the true mean and SE. In later analyses, we applied MI regression and MI MCMC to another data set, containing three-year resource use and costs of patients with COPD. In this data set, only costs of the treatment of exacerbations and maintenance medication costs were included. When applied to these data, MI regression and MI MCMC did not provide meaningful outcomes, confirming that the use of these methods may not be appropriate in case of such complex distributions.

The review of Barber and Thompson showed that 50% of the trial-based economic evaluations passed peer review and editorial judgement without providing any information about the completeness of follow-up. Another 25% of the papers left out all patients with incomplete follow-up data, without any further motivation (Barber and Thompson, 1998). This is an alarming situation. Chapter 7 shows that the impact of incomplete data on the outcomes of the analyses may be substantial, even when the dropout rate and the proportion of missing observations are relatively modest. The lack of information about the completeness of the follow-up and the failure to adequately incorporate these data in the analysis, question the quality and comparability of these economic evaluations. The results of our study in chapter 8 may provide some guidance as to how to deal with the data of patients with incomplete follow-up. This advice can be summarised into three steps: 1) investigation of the distribution of costs; 2) investigation of the amount and pattern of dropout; and 3) selection of methods to deal with incomplete data.

Step 1: Investigate the distribution of costs

1) Summarise the data using descriptive statistics like mean costs, variance, median, mode, skewness, outliers and percentiles.

2) Plot the data, for instance histograms of mean costs at each time interval, a line diagram of total costs over time, line diagrams of costs of individual patients over time.

3) Determine the major cost drivers, for instance the contribution of individual resource items to total costs, the contribution of the 10% of most expensive patients to total costs, the impact of costs associated with clinical events.

Step 2: Investigate the amount and pattern of dropout

1) Determine the amount of dropout by treatment group in terms of the proportion of patients who withdraw, the average observation period of patients with incomplete follow-up and the proportion of missing observations.

2) Investigate the pattern of dropout.

- The likelihood of the dropout pattern to be DCAR may be investigated by comparing the dropout rate between treatment groups. A significant difference in dropout rates between treatment groups suggests that the dropout mechanism is not completely at random.
- The likelihood of the dropout pattern to be DCAR may further be investigated by studying the reason for dropout. Lack of therapeutic benefit or side-effects are indicative of a correlation between dropout and costs.
- A comparison of the costs per day of the completers and dropouts during the observation period may also help to assess the likelihood of DCAR.
- A more formal approach to assess whether the dropout pattern is completely at random constitutes the use of a logistic regression model (Curran et al., 1998; Diggle, 1989; Ridout, 1991). The dependent variable in such a model indicates missingness, whereas cost observations in previous periods, background characteristics, clinical variables and quality-of-life variables may be used as independent variables. The model can be used to test the hypothesis that all coefficients in the model are zero. Rejection of the hypothesis implies that the dropout mechanism is either DAR or ID.
- It is not possible to determine whether dropout is DAR or ID. In methods that make use of a model to analyse the incomplete data, the plausibility of the data being DAR (and not ID) is enhanced by including as many observed characteristics of each patient as possible in the analysis. Increasing the pool of observed variables decreases the degree to which missingness depends on unobservables given the observed variables (Gelman et al., 1995).

Step 3: Selection of methods to deal with incomplete data

1) If the dropout rate is low and comparable between treatment groups and the assumption of DCAR is justified, naïve approaches like complete cases or linear extrapolation can be used. Of these, the method with the most conservative outcome should be adopted.

2) In all other situations, the use of principled methods should be considered. If the assumption of DCAR is justified, the product-limit estimator, the bootstrap EM algorithm or any of the MI approaches can be used. The need to perform further analyses in a complete (imputed) data set may direct the choice to one of the MI approaches.

3) If there is no evidence of DCAR, the bootstrap EM algorithm, MI regression and MI MCMC are the preferred methods. However, in this case, the distribution of costs should always be considered. Whereas these methods are able to deal with considerable skewness of the data, mixed or bimodal distributions and a large variation of costs within patients over time may

prevent these methods from working properly. In data sets with such complex distributions of costs, and with substantial and different dropout rates between treatment groups, the applicability of standard methods to analyse the incomplete data should be questioned.

9.7 APPLICABILITY OF RESULTS TO OTHER DISEASES

The discussion of the standardisation, comparability and generalisability of costs was not specifically directed towards economic evaluations of patients with COPD. The findings with regard to these subjects may apply to the same extent to economic evaluations in other diseases. The analysis of the impact of events and incomplete data are mainly based on the economic evaluation of tiotropium in patients with COPD and it may be questioned whether the results apply to other diseases as well.

COPD is a chronic disease. The typical distribution of costs, the large variation of costs within a patient over time and the high impact of event-driven costs occurring in only a few patients may be characteristic for many chronic diseases. The analysis of incomplete data due to premature withdrawal is also likely to be a problem in many longitudinal, prospective economic evaluations of chronic diseases. In fact, the simulation study in chapter 8, in which the probability to withdraw was associated with worse health status and higher costs, was designed to reflect a 'typical' chronic disease, rather than COPD only.

The applicability of findings to acute or life-threatening diseases may not be as straightforward. Treatment of patients diagnosed with, for instance, cancer or acute cardiovascular diseases may often generate very high costs during the first months after diagnosis. Costs at later stages may constitute only a fraction of the initial treatment costs. In these studies, all patients may accrue very high costs and the presence of patients with low or zero costs may be rare. Hence, the impact of high costs of events after initial diagnosis may not affect the distribution of costs as strongly as in chronic diseases that progress more slowly. Similarly, because of the high costs at the start of treatment, dropout may be less of a problem, as costs in later stages may be relatively modest when compared to the initial treatment costs. On the other hand, studies in acute or life-threatening diseases may be confronted with other typical problems. For instance, outcome measures in clinical studies of acute diseases may involve the occurrence of a major clinical event. Data of patients after the occurrence of these events may not always have been collected. In addition, if the study aims to estimate lifetime costs, the outcome of interest may not have been observed at the end of the study. The occurrence of this type of censoring may also considerably complicate the cost analysis (Lin et al., 1997).

In conclusion, the findings of this thesis may apply to many diseases other than COPD. Standardisation and generalisability are essential determinants of almost all studies incorporating resource use and costs. The impact of high-cost events and the problem of incomplete data due to premature withdrawal, as discussed in this thesis, deserve more attention in virtually all economic evaluations.

REFERENCES

Al MJ, van Hout BA, Michel BC, Rutten FFH. *Sample size calculation in economic evaluations.* Health Econ 1998;7:327-35.

American Thoracic Society. *Lung function testing: selection of reference values and interpretative strategies.* Am Rev Respir Dis 1991;144:1202-18.

American Thoracic Society. *Dyspnea: mechanisms, assessment, and management. A consensus statement.* Am J Respir Crit Care Med 1999;159(1):321-40.

Andersson F, Borg S, Jansson SA, Jonsson AC, Ericsson A, Prutz C, Ronmark E, Lundback B. *The costs of exacerbations in chronic obstructive pulmonary disease (COPD).* Respir Med 2002;96(9):700-8.

Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. *Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease.* Ann Intern Med 1987;106(2):196-204.

Badia X, Nocea G, Rovira J. *Differences in the methodology and data of economic evaluations of a health programme. The case of hepatitis B vaccination programmes in Spain.* Pharmacoeconomics 1997;11(2):175-83.

Bang H, Tsiatis AA. *Estimating medical costs with censored data.* Biometrika 2000;87:329-43.

Barber JA, Thompson SG. *Analysis and interpretation of cost data in randomised controlled trials: review of published studies.* BMJ 1998;317:1195-200.

Barnes PJ. *Chronic obstructive pulmonary disease: new treatments for COPD.* Thorax 2003;58(9):803-8.

Beacon HJ, Thompson SG. *Multi-level models for repeated measurement data: application to quality of life data in clinical trials.* Stat Med 1996;15:2717-32.

Boer B. *Onderzoek op maat. Een verkenning van factoren voor het gebruik van Technology Assessment. (Customised research: an investigation into factors for the use of Technology Assessment.)* PhD thesis. In Dutch: Erasmus University Rotterdam; 2002.

Borg S, Ericsson A, Wedzicha JA, Gulsvik A, Lundbäck B, Donaldson GC, Sullivan SD. *A computer simulation model of the natural history and economic impact of chronic obstructive pulmonary disease.* Value Health 2004;7(2):153-67.

Briggs AH. *Handling uncertainty in cost-effectiveness models.* Pharmacoeconomics 2000;17(5):479-500.

Briggs AH, Ades AE, Price MJ. *Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework.* Med Decis Making 2003;23(4):341-50.

Briggs AH, Clark T, Wolstenholme J, Clarke P. *Missing... presumed at random: cost-analysis of incomplete data.* Health Econ 2003;12(5):377-92.

Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. *Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease.* Med Decis Making 2002;22(4):290-308.

Briggs AH, Gray A. *The distribution of health care costs and their statistical analysis for economic evaluation.* J Health Serv Res Policy 1998;3(4):233-45.

Briggs AH, O'Brien BJ. *The death of cost-minimization analysis?* Health Econ 2001;10(2):179-84.

Briggs AH, O'Brien BJ, Blackhouse G. *Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies.* Annu Rev Public Health 2002;23:377-401.

Briggs AH, Wonderling DE, Mooney CZ. *Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation.* Health Econ 1997;6:327-40.

Brouwer WBF, Koopmanschap MA, Rutten FFH. *Productivity costs measurement through quality of life? A response to the recommendations of the Washington Panel.* Health Econ 1997;6:253-9.

Brouwer WBF, Koopmanschap MA, Rutten FFH. *Patient and informal caregiver time in cost-effectiveness analysis. A response to the recommendations of the Washington Panel.* Int J Technol Assess Health Care 1998;14(3):505-13.

Brusasco V, Hodder R, Miravitlles M, Korducki L, Towse L, Kesten S. *Health outcomes following treatment for six months with once daily tiotropium compared with twice daily salmeterol in patients with COPD.* Thorax 2003;58(5):399-404.

Burge PS, Calverley PMA, Jones PW, Spencer S, Anderson JA, Maslen TK. *Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the Isolde trial.* BMJ 2000;320:1297-303.

Canadian Coordinating Office for Health Technology Assessment. *A guidance document for the costing process.* Ottawa; August 1996.

Canadian Coordinating Office for Health Technology Assessment. *Guidelines for Economic Evaluation of Pharmaceuticals: Canada.* Ottawa; November 1997. www.ccohta.ca/entry_e.html.

Carides GW, Heyse JF, Iglewicz B. *A regression-based method for estimating mean treatment cost in the presence of right-censoring.* Biostatistics 2000;1(3):299-313.

Casaburi R, Mahler DA, Jones PW, Wanner A, San PG, ZuWallack RL, Menjoge SS, Serby CW, Witek TJ, Jr. *A long-term evaluation of once-daily inhaled tiotropium in chronic obstructive pulmonary disease.* Eur Respir J 2002;19(2):217-24.

Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, Pinto Plata V, Cabral HJ. *The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease.* N Engl J Med 2004;350(10):1005-12.

Celli BR, Snider GL, Heffner J, Tiep B, Ziment I, Make BJ, Braman S, Olsen G, Phillips Y. *Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease; official statement of the American Thoracic Society.* Am J Respir Crit Care Med 1995;152:S77-S120.

Centraal Bureau voor de Statistiek (Statistics Netherlands). *Statline (Central database of Statistics Netherlands).* 2004; http://statline.cbs.nl.

Commissie modernisering curatieve zorg. *Gedeelde zorg: betere zorg; Achtergrondstudies van de Commissie modernisering curatieve zorg* (In Dutch). Roosendaal; January 1994.

Commonwealth Department of Health and Ageing. *Manual of Resource Items and their Associated Costs.* Canberra; 2002. www.health.gov.au/pbs/general/pubs/manual/index.htm.

Commonwealth Department of Health and aging. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee.* Canberra; 2002. www.health.gov.au/pbs/general/pubs/guidelines/index.htm.

Cook NR. *An imputation method for non-ignorable missing data in studies of blood pressure.* Stat Med 1997;16:2713-28.

Curran D, Bacchi M, Schmitz SF, Molenberghs G, Sylvester RJ. *Identifying the types of missingness in quality of life data from clinical trials.* Stat Med 1998;17(5-7):739-56.

Cuvelier A, Muir JF, Benhamou D, Weitzenblum E, Zuck P, Delacenserie R, Taytard A, Iacono P. *Dry powder ipratropium bromide is as safe and effective as metered-dose inhaler formulation: a cumulative dose-response study in chronic obstructive pulmonary disease patients.* Respir Care 2002;47(2):159-66.

Davidoff AJ, Powe NR. *The Role of Perspective in Defining Economic Measures for the Evaluation of Medical Technology.* Int J Technol Assess Health Care 1996;12-1:9-21.

Decramer M, Dekhuijzen PN, Troosters T, van Herwaarden C, Rutten-van Molken MPMH, van Schayck CP, Olivieri D, Lankhorst I, Ardia A. *The Bronchitis Randomized On NAC Cost-Utility Study (BRONCHUS): hypothesis and design.* BRONCHUS-trial Committee. Eur Respir J 2001;17(3):329-36.

Dempster AP, Laird NM, Rubin DB. *Maximum likelihood estimation from complete data via the EM algorithm (with discussion).* Journal of the Royal Statistical Society Series B, 1977;39:1-38.

Dewan NA, Fafique S, Kanwar B, Satpathy H, Ryschon K, Tillotson GS, Niederman MS. *Acute exacerbation of COPD, factors associated with poor treatment outcome.* Chest 2000;117:662-71.

Diggle P. *Testing for random dropouts in repeated measurement data.* Biometrics 1989;45:1255-8.

Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. *Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease.* Thorax 2002;57(10):847-52.

Donohue JF, van Noord JA, Bateman ED, Langley SJ, Lee A, Witek TJ, Jr., Kesten S, Towse L. *A 6-month, placebo-controlled study comparing lung function and health status changes in COPD patients treated with tiotropium or salmeterol.* Chest 2002;122(1):47-55.

Drummond MF. *Comparing cost-effectiveness across countries: the model of acid-related disease.* Pharmacoeconomics 1994;5(Suppl 3):60-7.

Drummond MF, Bloom BS, Carrin G, Hillman AL, Hutchings AC, Knill-Jones RP, de Pouvourville G, Torfs K. *Issues in the cross-national assessment of health technology.* Int J Technol Assess Health Care 1992;8(4):671-82.

Drummond MF, Brandt A, Luce BR, Rovira J. *Standardizing methodologies for economic evaluation in health care.* Int J Technol Assess Health Care 1993;9(1):26-36.

Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* 2 Ed. New York: Oxford University Press; 1997.

Feenstra TL, van Genugten ML, Hoogenveen RT, Wouters EF, Rutten-van Mölken MPMH. *The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands.* Am J Respir Crit Care Med 2001;164(4):590-6.

Fenwick E, Claxton K, Sculpher M. *Representing uncertainty: the role of cost-effectiveness acceptability curves.* Health Econ 2001;10(8):779-87.

Ferguson B. *NHS database of reference costs is severely flawed.* BMJ 2001;323:106.

Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis (Ch. 17).* 1st Ed. London: Chapman&Hall; 1995.

Gimeno F, van Veenen R, Berg WC, Steenhuis EJ. *A placebo-controlled comparison between the bronchodilatory effects of ipratropium bromide inhaled as a dry powder and by metered dose inhaler in chronic obstructive pulmonary disease.* Ann Allergy 1988;61(5):341-3.

Goeree R, Gafni A, Hannah M, Myhr T, Blackhouse G. *Hospital Selection for Unit Cost Estimates in Multicentre Economic Evaluations; does the choice of hospitals make a difference.* Pharmacoeconomics 1999;15(6):561-72.

Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in Health and Medicine.* 1 Ed. New York: Oxford University Press; 1996.

Grasso ME, Weller WE, Shaffer TJ, Diette GB, Anderson GF. *Capitation, managed care, and chronic obstructive pulmonary disease.* Am J Respir Crit Care Med 1998;158:133-8.

Graves N, Walker D, Raine R, Hutchings A, Roberts JA. *Cost data for individual patients included in clinical studies: no amount of statistical analysis can compensate for inadequate costing methods.* Health Econ 2002;11(8):735-9.

Greiner W, Schöffski O, Graf-von der Schulenburg JM. *The transferability of international economic health-economic results to national study questions.* Hepac 2000;1:94-102.

Groeneveld JF. *Specialistenhonoraria* (In Dutch). Medisch Contact 1996;27(28):919-21.

Hakkaart-van Roijen L. *Societal perspective on the cost of illness.* PhD thesis: Erasmus University Rotterdam; 1998.

Haycox A, Walley T. *Pharmacoeconomics: evaluating the evaluators.* Br J Clin Pharmacol 1997;43:451-456.

Hilleman DE, Dewan N, Malesker M, Friedman M. *Pharmacoeconomic evaluation of chronic obstructive pulmonary disease.* Chest 2000;118:1278-85.

Hjelmgren J, Berggren F, Andersson F. *Health economic guidelines – similarities, differences and some implications.* Value Health 2001;4(3):225-50.

Hogan JW, Laird NM. *Mixture models for the joint distribution of repeated measures and event times.* Stat Med 1997;16:239-57.

Hoogendoorn M, Feenstra T, Hoogenveen R, van Genugten ML, Rutten-van Mölken MPMH. *A health policy model for COPD: Effects of smoking cessation.* Rotterdam: iMTA, Erasmus MC; November 2003.

Horngren CT. *Cost accounting.* 5 Ed. Englewood Cliffs: Prentice-Hall; 1982.

Hvizdos KM, Goa KL. *Tiotropium bromide.* Drugs 2002;62(8):1195-203.

Institute of Health Economics. *A national list of provincial costs for health care: Canada, 1997/8.* Edmonton.

Jacobs P, Bachynsky J. *Costing Methods in the Canadian Literature on the Economic Evaluation of Health Care.* Int J Technol Assess Health Care 1996;12(4):721-34.

Jacobs P, Bachynsky J, Baladi J-F. *A Comparative Review of Pharmacoeconomic Guidelines.* Pharmacoeconomics 1995;8(3):182-9.

Jacobs P, Baladi J-F. *Biases in Cost Measurement for Economic Evaluation Studies in Health Care.* Health Econ 1996;5:525-29.

Jacobs P, Roos NP. *Standard cost lists for healthcare in Canada.* Pharmacoeconomics 1999;15(6):551-60.

Jacobson L, Herzman P, Löfdahl C-G, Skoogh B-E, Lindgren B. *The economic impact of asthma and chronic obstructive pulmonary disease (COPD) in Sweden in 1980 and 1991.* Respir Med 2000;94(3):247-55.

Jansson SA, Andersson F, Borg S, Ericsson A, Jonsson E, Lundback B. *Costs of COPD in Sweden according to disease severity.* Chest 2002;122(6):1994-2002.

Johannesson M. *Avoiding double-counting in pharmacoeconomic studies.* Pharmacoeconomics 1997;11(5):385-88.

Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. *Assessing the costs of healthcare technologies in clinical trials.* Health Technol Assess 1999;3(6):1-76.

Jones PW. *Interpreting thresholds for a clinically significant change in health status in asthma and COPD.* Eur Respir J 2002;19(3):398-404.

Jones PW, Lasserson D. *Relationship between change in St. George's Respiratory Questionnaire score and patients' perception of treatment efficacy after one year of therapy with nedocromil sodium.* Am Rev Respir Dis 1994;149:A211.

Jones PW, Quirk FH, Baveystock CM. *The St. George's Respiratory Questionnaire.* Respir Med 1991;85 Suppl B:25-31.

Jones PW, Wilson K, Sondhi S. *Cost-effectiveness of salmeterol in patients with chronic obstructive pulmonary disease: an economic evaluation.* Respir Med 2003;97(1):20-6.

Kind P, Macran S, Spencer MD. *The quality of life deficit in COPD: evidence from a national population survey.* Poster presented at ERS; Stockholm; 2002.

Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. *The friction cost method for measuring indirect costs of disease.* J Health Econ 1995;14:171-89.

Koopmanschap MA, van Ineveld BM. *Towards a New Approach for Estimating Indirect Costs of Disease.* Soc Sci Med 1992;34:1005-10.

Langley PC. *The November 1995 Revised Australian Guidelines for the Economic Evaluation of Pharmaceuticals.* Pharmacoeconomics 1996;9(4):341-52.

Lavori P, Dawson R, Shera D. *A multiple imputation strategy for clinical trials with truncation of patient data.* Stat Med 1995;14:1913-25.

Lin D, Feuer E, Etzioni R, Wax Y. *Estimating medical costs from incomplete follow-up data.* Biometrics 1997;53:419-34.

Lin DY. *Linear regression analysis of censored medical costs.* Biostatistics 2000;1(1):35-47.

Little R, *Rubin D. Statistical analysis with missing data.* 1st Ed. New York: Wiley; 1987.

Little RJA. *Modelling the dropout mechanism in repeated measure studies.* J Am Stat Assoc 1995; 90: 1112-21.

Little RJA, Wang Y. *Pattern-mixture models for multivariate incomplete data with covariates.* Biometrics 1996;52:98-111.

Lundback B, Lindberg A, Lindstrom M, Ronmark E, Jonsson AC, Jonsson E, Larsson LG, Andersson S, Sandstrom T, Larsson K. *Not 15 but 50% of smokers develop COPD? – Report from the Obstructive Lung Disease in Northern Sweden Studies.* Respir Med 2003;97(2):115-22.

Mahler DA, Weinberg DH, Wells CK, Feinstein AR. *The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes.* Chest 1984;85(6):751-8.

Manning WG, Mullahy J. *Estimating log models: to transform or not to transform?* J Health Econ 2001;20(4):461-94.

Mason J. *The generalisability of pharmacoeconomic studies.* Pharmacoeconomics 1997;11(6):503-514.

Mason J, Drummond M. *Reporting guidelines for economic studies.* Health Econ 1995;4:85-94.

Matsuyama Y, Ohashi Y. *Mixed models for bivariate response repeated measures data using Gibbs sampling.* Stat Med 1997;16:1587-601.

McGuire A, Morris S. *What is it to be a model. Trials and tribulations in economic evaluation.* Hepac 2000;1:33-6.

Miller DK, Homan SM. *Determining transition probabilities: confusion and suggestions.* Med Decis Making 1994;14(1):52-8.

Ministerie van Financiën. *Handboek financiële administratie en informatie rijksoverheid* (In Dutch). 1 Ed. Den Haag; 1995.

Miravitlles M, Murio C, Guerrero T, Gisbert R. *Pharmacoeconomic evaluation of acute exacerbations of chronic bronchitis and COPD.* Chest 2002;121(5):1449-55.

Miravitlles M, Murio C, Guerrero T, Gisbert R. *Costs of chronic bronchitis and COPD: a 1-year follow-up study.* Chest 2003;123(3):784-91.

Murray CJL, Lopez AD. *Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study.* Lancet 1997;349:1498-504.

National Obesity education initiative. *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults; the evidence report: National Institutes of Health, National Heart, Lung, and Blood Institute; September 1998.* http://www.nhlbi.nih.gov/guidelines/obesity/ob_gdlns.pdf.

Nishimura K, Izumi T, Tsukino M, Oga T. *Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD.* Chest 2002;121(5):1434-40.

O'Callaghan F. *A multiple imputation strategy for missing data in longitudinal studies.* Kwantitatieve methoden 1999;62:111-22.

O'Hagan A, Stevens JW. *Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality?* Health Econ 2003;12(1):33-49.

O'Hagan A, Stevens JW. *On estimators of medical costs with censored data.* J Health Econ 2004;23(3):615-25.

Oostenbrink JB, Koopmanschap MA, Rutten FFH. *Manual for Costing: Methods and standard costs for economic evaluations in health care.* In Dutch. Amstelveen: Health Insurance Council; March 2000.

Osman IM, Godden DJ, Friend JA, Legge JS, Douglas JG. *Quality of life and hospital re-admission in patients with chronic obstructive pulmonary disease.* Thorax 1997;52(1):67-71.

Pang F. *Design, analysis and presentation of multinational economic studies: the need for guidance.* Pharmacoeconomics 2002;20(2):75-90.

Paterson C, Langan CE, McKaig GA, Anderson PM, Maclaine GD, Rose LB, Walker SJ, Campbell MJ. *Assessing patient outcomes in acute exacerbations of chronic bronchitis: the measure your medical outcome profile (MYMOP), medical outcomes study 6-item general health survey (MOS-6A) and EuroQol (EQ-5D).* Qual Life Res 2000;9(5):521-7.

Pauwels R, Calverley P, Buist AS, Rennard S, Fukuchi Y, Stahl E, Lofdahl CG. *COPD exacerbations: the importance of a standard definition.* Respir Med 2004;98(2):99-107.

Pauwels RA, Buist S, Calverley PMA, Jenkins CR, Hurd SS. *Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease.* Am J Respir Crit Care Med 2001;163(5):1256-76.

Pauwels RA, Buist S, Calverley PMA, Jenkins CR, Hurd SS. *Global initiative for Chronic Obstructive Lung Disease.* Bethesda: NHLBI/WHO; 2003.

Pechevis M, Fagnani F, Brin S, De Zelicourt M, Morales M. *Infections respiratoires récidivantes du sujet atteint de bronchite chronique obstructive: prise en charge médicale et coûts* (In French). Rev Mal Resp 1996;13:507-12.

Polder JJ. *Cost of illness in the Netherlands* (PhD thesis): Erasmus University Rotterdam; 2001.

Price MJ, Briggs AH. *Development of an economic model to assess the cost effectiveness of asthma management strategies.* Pharmacoeconomics 2002;20(3):183-94.

Prismant. *Rekeningschema voor zorginstellingen* (In Dutch). Utrecht; November 2003. Issued on cd-rom.

Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. *Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society.* Eur Respir J Suppl 1993;16:5-40.

Raikou M, Briggs A, Gray A, McGuire A. *Centre-specific or average unit costs in multi-centre studies? Some theory and simulation.* Health Econ 2000;9:191-8.

Ramsey SD, Berry K, Etzioni R, Kaplan RM, Sullivan SD, Wood DE. *Cost effectiveness of lung-volume-reduction surgery for patients with severe emphysema.* N Engl J Med 2003;348(21):2092-102.

Rees PJ. *Tiotropium in the management of chronic obstructive pulmonary disease.* Eur Respir J 2002;19:205-6.

Ridout MS. *Testing for random dropouts in repeated measurement data – reader reaction.* Biometrics 1991;47:1617-21.

Riteco JA, de Heij LJM, van Luijn JCF, Wolff I. *Dutch guidelines for pharmacoeconomic research.* Report. Amstelveen: Dutch Health Insurance Council; March 25, 1999.

Rizzo JD, Powe NR. *Methodological hurdles in conducting pharmacoeconomic analyses.* Pharmacoeconomics 1999;15(4):339-55.

Rodriguez-Roisin R. *Toward a consensus definition for COPD exacerbations.* Chest 2000;117:398S-401S.

Rosenbaum PR, Rubin DB. *Reducing bias in observational studies using subclassification in the propensity score.* J Am Stat Assoc 1984;79:516-24.

Rovira J. *Standardization of the Economic Evaluation of Health Technologies, European Developments.* Medical Care 1996;34(suppl):DS182-88.

Rovira J, Antonanzas F. *Economic Analysis of Health Technologies and Programmes.* Pharmacoeconomics 1995;8(3):245-52.

Rubin DB. *Multiple imputation for nonresponse in surveys.* 1 Ed. New York: J. Wiley & Sons; 1987.

Rubin DB. *Multiple imputation after 18+ years.* J Am Stat Assoc 1996;91:473-89.

Rubin DB, Schenker N. *Multiple imputation in healthcare databases: an overview and some applications.* Stat Med 1991;10:595-8.

Ruchlin HS, Dasbach EJ. *An economic overview of chronic obstructive pulmonary disease.* Pharmacoeconomics 2001;19(6):623-42.

Rutten-van Mölken MPMH, Postma MJ, Joore MA, van Genugten M, Leidl R, Jager C. *Current and future medical costs of asthma and chronic obstructive pulmonary disease in The Netherlands.* Respir Med 1999;93(11):779-87.

Rutten-van Mölken MPMH, van Doorslaer EKA, van Vliet RCJA. *Statistical analysis of cost outcomes in a randomized controlled clinical trial.* Health Econ 1994;3:333-45.

Rychlik R, Pfeil T, Daniel D, Pfeil B, Mast O, Thate-Waschke I, Lorenz J. *Socioeconomic relevance of acute exacerbations of chronic bronchitis in the Federal Republic of Germany.* A prospective cost of illness study. Dtsch Med Wochenschr 2001;126(13):353-9.

Scanlon PD, Connett JE, Waller LA, Altose MD, Bailey WC, Buist AS. *Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease.* The Lung Health Study. Am J Respir Crit Care Med 2000;161:381-90.

Schafer J, *Analysis of incomplete multivariate data.* 1st Ed. London: Chapman & Hall; 1997.

Schafer JL, Yucel RM. *Computational strategies for multivariate linear mixed-effects models with missing values.* J Comp Graph Stat 2002; 11: 434-57.

Sculpher MJ, Price M. *Measuring costs and consequences in economic evaluation in asthma.* Respir Med 2003;97(5):508-20.

Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. *Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease.* Am J Respir Crit Care Med 2000;161:1608-13.

Siafakas NM, Vermeire P, Pride NB, Paoletti P, Gibson J, Howard P, Yernault JC, Decramer M, Higenbottam T, Postma DS, Rees J. *Optimal assessment and management of chronic obstructive pulmonary disease.* Eur Respir J 1995;8:1398-420.

Siegel JE, Torrance GW, Russell LB, Luce BR, Weinstein MC, Gold MR. *Guidelines for Pharmacoeconomic Studies.* Pharmacoeconomics 1997;11(2):159-168.

Siegel JE, Weinstein MC, Russell LB, Gold MR. *Recommendations for reporting cost-effectiveness analysis.* JAMA 1996;276:1339-41.

Sin DD, Golmohammadi K, Jacobs P. *Cost-effectiveness of inhaled corticosteroids for chronic obstructive pulmonary disease according to disease severity.* Am J Med 2004;116(5):325-31.

Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J, Pride NB. *Recent trends in physician diagnosed COPD in women and men in the UK.* Thorax 2000;55(9):789-94.

Spencer S, Jones PW. *Time course of recovery of health status following an infective exacerbation of chronic bronchitis.* Thorax 2003;58(7):589-93.

Stang P, Lydick E, Silberman C, Kempel A, Keating ET. *The prevalence of COPD: using smoking rates to estimate disease frequency in the general population.* Chest 2000;117(5 Suppl 2):354S-9S.

Strassels SA, Smith DH, Sullivan SD, Mahajan PS. *The costs of treating COPD in the United States.* Chest 2001;119(2):344-52.

Sullivan SD, Ramsey SD, Lee TA. *The economic burden of COPD.* Chest 2000;117:5S-9S.

Thompson SG, Barber JA. *How should cost data in pragmatic randomised trials be analysed?* BMJ 2000;320(7243):1197-200.

Torrance G, Walker V, Grossman R, Mukherjee J, Vaughan D, La Forge J, Lampron N. *Economic evaluation of Ciprofloxacin compared with usual antibacterial care for the treatment of acute exacerbations of chronic bronchitis in patients followed for 1 year.* PharmacoEconomics 1999;16 (5pt1):499-520.

Torrance GW, Blaker D, Detsky A, Kennedy W, Schubert F, Menon D, Tugwell P, Konchak R, Hubbard E, Firestone T. *Canadian guidelines for Economic Evaluation of Pharmaceuticals.* Pharmacoeconomics 1996;9(6):535-59.

Touloumi G, Babiker AG, Pocock SJ, Darbyshire JH. *Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study.* Stat Med 2001;20:3715-28.

U.S. Department of Health and Human Services. *The Health Benefits of Smoking Cessation. A Report on the Surgeon General.* Washington D.C.: U.S. Government Printing Office; 1990.

van Agthoven M, Vellenga E, Fibbe WE, Kingma T, Uyl-de Groot CA. *Cost analysis and quality of life assessment comparing patients undergoing autologous peripheral blood stem cell transplantation or autologous bone marrow transplantation for refractory or relapsed non-Hodgkin's lymphoma or Hodgkin's disease. a prospective randomised trial.* Eur J Cancer 2001;37(14):1781-9.

van Buuren S, Oudshoorn CGM. *Flexible multivariate imputation by MICE.* Leiden: TNO Preventie en Gezondheid. Report No.: TNO/VGZ/PG 99.054 (see www.multiple_imputation.com for pdf).

van Enckevort PJ, TenVergert EM, Bonsel GJ, Geertsma A, van der Bij W, de Boer J, Koopmanschap MA, Al MJ, Rutten FFH. *Technology assessment of the Dutch lung transplantation program.* Int J Technol Assess Health Care 1998;14(2):344-56.

van Hout BA, Al MJ, Gordon GS, Rutten FFH. *Costs, effects and c/e-ratio's alongside a clinical trial.* Health Econ 1994;3:309-19.

Vincken W, van Noord JA, Greefhorst AP, Bantje TA, Kesten S, Korducki L, Cornelissen PJ. *Improved health outcomes in patients with copd during 1 year's treatment with tiotropium.* Eur Respir J 2002;19(2):209-16.

Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. *Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices – Modeling Studies.* Value Health 2003;6(1):9-17.

Willan AR, Chen EB, Cook RJ, Lin DY. *Incremental net benefit in randomized clinical trials with quality-adjusted survival.* Stat Med 2003;22(3):353-62.

Willke RJ, Glick H, A, Polsky D, Schulman K. *Estimating Country-Specific Cost-Effectiveness from Multinational Clinical Trials.* Health Econ 1998;7:481-93.

Witek TJ, Jr., Mahler DA. *Meaningful effect size and patterns of response of the transition dyspnea index.* J Clin Epidemiol 2003;56(3):248-55.

Wouters EF. *Economic analysis of the Confronting copd survey: an overview of results.* Respir Med 2003;97(suppl c):S3-14.

Zeger SL, Liang KY. *Longitudinal data analysis for discrete and continuous outcomes.* Biometrics 1986;42:121-30.

ABBREVIATIONS

BMI body mass index

BDI baseline dyspnoea index

CE ACCEPTABILITY CURVE cost-effectiveness acceptability curve

CE PLANE cost-effectiveness plane

CI confidence interval

COPD chronic obstructive pulmonary disease

CRF case report form

CP coverage probability

DAR dropout at random

DCAR dropout completely at random

DRG diagnostic-related group

DPI dry powder inhaler

EM expectation maximisation

ER emergency room

EQ-5D Euroqol questionnaire, 5 dimensions

FEV_1 forced expiratory volume in one second

FVC forced vital capacity

GP general practitioner

HTA health technology assessment

ICER incremental cost-effectiveness ratio

ICU intensive care unit

ID informative dropout

LVCF last value carried forward

MI multiple imputation

MCMC Monte Carlo Markov chain

MDI metered dose inhaler

ML millilitre

RCT randomised controlled trial

QALY quality-adjusted life-year

QOL quality of life

SA sensitivity analysis

SD standard deviation

SE standard error

SEE standard error estimator

SGRQ St. George's Respiratory Questionnaire

SSE sampling standard error

TDI transitional dyspnoea index

µG microgram

UI uncertainty interval

SUMMARY

Introduction | The rapid increase in the costs of healthcare during the last decades in many Western countries has increased the awareness that limits must be set to the growth of the costs of healthcare. Instead of the automatic influx of new technologies, the need arose to assess these technologies in terms of their costs and benefits in order to decide upon registration, reimbursement and pricing. These developments have led to a significant increase in the number and variety of economic evaluations in healthcare and, in order to improve comparability, many authors have argued for more standardisation of the methodology of economic evaluations.

The aim of this thesis is to contribute to the comparability and generalisability of economic evaluations by standardising and improving methods for the calculation and analysis of costs. The economic evaluation of tiotropium in patients with COPD will be used as a critical case to address the following methodological research areas:

- the standardisation and comparability of costs in economic evaluations;
- the generalisability of the results of economic evaluations to other countries;
- the impact of expensive medical events on the analysis of costs;
- the analysis of incomplete cost data due to dropout.

Economic evaluation of tiotropium in patients with COPD | Chronic obstructive pulmonary disease (COPD) is a disease of the respiratory system characterized by slowly progressive airflow limitation that is not fully reversible. The anticholinergic tiotropium is a long-acting bronchodilator for the treatment of patients with COPD. In this thesis, the costs and effects of the treatment with tiotropium will be considered using two complementary approaches to economic evaluations. The first approach is a prospective cost-effectiveness analysis performed alongside two one-year randomised controlled trials in the Netherlands and Belgium, comparing tiotropium to the short-acting anticholinergic ipratropium. The second approach to cost-effectiveness analysis concerns a decision-analytic model, using data from six trials from the clinical trial program of tiotropium. The aim of this probabilistic Markov model is to compare the cost-effectiveness of tiotropium, salmeterol (a long-acting beta-agonist) and ipratropium for the treatment of patients with COPD in different countries. In this thesis, a comparison is made between the Netherlands and Canada. An advantage of the prospective economic evaluation is that all outcomes are obtained in a randomised situation and that they are based on data from individual patients from the same population. The model-based economic evaluation, on the other hand, provides the opportunity to compare several treatments within the same framework, to use data from a variety of sources, to adjust cost estimates in order to compensate for trial-related protocol-driven costs, and to generalise the estimates of costs and effects to other settings and countries.

The prospective and model-based economic evaluations resulted in comparable estimates of the reduction in exacerbations due to treatment with tiotropium. The estimated reduction was approximately 35% (difference: 0.27, 95% confidence interval (CI): -0.02; 0.37) when tiotropium was compared to ipratropium, and approximately 20% (difference: 0.17, 95% CI: -0.02; 0.37) when tiotropium was compared to salmeterol. Differences between the

prospective and model-based approach became apparent in the estimates of costs. In the prospective economic evaluation, tiotropium was associated with an increase in costs of € 180 (95% CI: -268; 627) when compared to ipratropium. In the model-based analysis for the Netherlands, the mean costs of patients in the tiotropium group were € 170 (95% CI: -812; 335) lower than the costs of patients in the ipratropium group and € 42 (95% CI: -484, 353) lower than the costs of patients in the salmeterol group. The difference with the prospective analysis is mainly due to adjustments for protocol-driven costs, the estimated price of the study-medication ipratropium, and the estimated number of severe exacerbations. In Canada, costs were consistently lower than in the Netherlands and almost the same in all treatment groups. Differences between the two countries were mainly due to a longer length of hospital stay in case of an exacerbation in the Netherlands.

Standardisation and comparability of costs in economic evaluations | To measure and value costs in the economic evaluations of tiotropium we made use of the Dutch 'Manual for costing: methods and standard costs for economic evaluations in healthcare'. The manual was published as a response to differences between economic evaluations with regard to the way costs were measured and valued. The manual aims to improve the quality and comparability of the costing in Dutch (pharmaco-) economic evaluations.

This thesis discusses the content of the manual and some of the key issues related to the standardisation of costs. In the manual, we introduce a six-step procedure for costing and these steps address: 1) the scope of the study; 2) the choice of cost categories; 3) the identification of units; 4) the measurement of resource use; 5) the monetary valuation of units; and 6) the calculation of unit costs. During each step, choices have to be made and these together define the approach taken. Five key issues related to the standardisation of costs are distinguished. These are the use of: 1) basic principles; 2) methods for the measurement and valuation; 3) standard costs; 4) standard values; and 5) the presentation of outcomes. By the use of these key issues, the manual tries to find a balance between standardisation of costs and the necessity to tailor the approach to a specific study setting.

Costs of inpatient hospital days are the main driver of total treatment costs in many diseases, therefore we performed a study to determine unit costs of inpatient hospital days from 22 general wards and 11 intensive care units (ICUs) from 15 hospitals. The mean costs per inpatient day varied from € 230 in general hospitals, to € 323 in university hospitals, and to € 1125 in ICUs. Nursing costs contributed to 38% of the total costs on a general ward to 48% of the total costs on an ICU. Despite the use of uniform costing methods, this study showed that there were considerable differences between the estimates of costs from the different centres. Often, these differences cannot be explained from the type of hospital or the case-mix of patients in a hospital. The results from this study were used to derive a standard cost for inpatient days for the Dutch manual. The use of standard costs for the valuation may eliminate some of the differences that are not due to actual differences in the resources consumed and thus contribute to the standardisation and comparability of economic evaluations.

Generalisability of economic outcomes across countries | The prospective economic evaluation of tiotropium was based on two clinical trials conducted in the Netherlands and in Belgium. Information about the cost-effectiveness of tiotropium in other countries was not available and the question arose how such information could be obtained. Whereas it may often be assumed that the biological effect of a treatment is more or less the same for patients from different countries, this clearly is not a valid assumption for resource use and costs. Many factors that are pertinent to resource use and costs are known to vary across countries and, clearly, resource use and costs may not be generalised to other countries without appropriate adjustment.

The primary aim of the model-based economic evaluation of tiotropium was to assess the cost-effectiveness of tiotropium in different countries. Considering this aim, several characteristics of the model are worth discussing. A first characteristic of the model is its probabilistic nature, as opposed to deterministic models, that enables the analysis of uncertainty around the estimates of costs and effects, and the resulting ratios. The second characteristic is that all model inputs related to the effectiveness of treatment are based on patient-level trial data. Hence, outcomes of the model could be validated with the clinical trial data, thereby increasing the acceptance of models by local reimbursement authorities. A final characteristic of this model is the assumption that, given disease state and the presence or absence of an event, estimates of resource use and unit costs do not differ between treatment arms. Hence, no assumptions have to be made about differences between treatment arms in the cost per event or the maintenance cost per disease severity state. All differences between treatment groups are accounted for by differences in transition and event probabilities. This characteristic can be considered a necessary requirement for models that aim to estimate cost-effectiveness in different countries or settings, and allow resource use to be estimated from a variety of sources and settings inside or outside the trials.

Analysis of event-driven costs | In a post-hoc analysis of the prospective cost-effectiveness analysis of tiotropium, we determined the costs of COPD-related exacerbations. Exacerbations contributed to approximately 34% of the total respiratory-related healthcare costs. Estimates of the mean (SD) costs of exacerbations ranged from € 86 (223) for a mild exacerbation to € 579 (1227) for a moderate exacerbation to € 4007 (5922) for a severe exacerbation. Of these exacerbations, 16% was associated with a hospitalisation and accounted for 90% of the total costs of exacerbations. This typical situation of a few patients incurring rare but highly expensive costs and many patients having few or no costs causes the distribution of costs to be severely skewed, with important implications for the calculation and analysis of cost data. As many empirical economic evaluations are performed alongside randomised controlled trials powered on some clinical outcome measure, the sample size is often not sufficient to detect differences between treatments that are 'statistically significant' at conventional levels. The high variance in costs and cost-effectiveness outcomes is one of the driving forces behind the concept that an adequate description of uncertainty is more relevant for decision-making than classical hypothesis testing based on some arbitrary threshold of statistical significance.

Despite the high costs of events and the resulting skewed distribution of costs, the sample size in the prospective cost-effectiveness analysis of tiotropium was large enough to apply common methods to analyse the cost data. Bootstrapping and normal approximation resulted in almost identical confidence intervals of the difference in costs. To describe the uncertainty around the cost-effectiveness ratios in the prospective and model-based economic evaluations, we put current principles into practice. In the prospective evaluation, we presented the incremental cost-effectiveness of tiotropium versus ipratropium on the cost-effectiveness (CE) plane and the CE acceptability curve (figure 4.2). The CE planes show the ellipses containing 5%, 50% and 95% of the probability density of the difference in costs and effects, facilitating a rapid and visual interpretation of the uncertainty around the CE ratio. The CE acceptability curve shows the probability that a treatment is cost-effective for different values of the ceiling ratio. For instance, if the willingness to pay to avoid one exacerbation or to have one additional patient with a relevant improvement on the St. George's Respiratory Questionnaire was set at € 2000, the probability that tiotropium is acceptable was 80% and 72% respectively. In the model-based economic evaluation, we compared three treatments within the same framework. Based on the assumption that these treatments are, in principle, mutually exclusive, we presented separate curves for each treatment instead of incremental curves for each mutual comparison. The acceptability frontier follows the curve of the treatment with maximum expected net benefit, hence determining the optimal treatment from a cost-effectiveness perspective for each value of the ceiling ratio. The acceptability frontier of exacerbations showed that tiotropium was associated with maximum expected net benefit for all values of the ceiling ratio above € 0 in the Netherlands and above € 10 in Canada.

Analysis of incomplete cost data due to premature withdrawal | About 83% of the patients in the prospective economic evaluation of tiotropium completed the entire 1-year study. About 15% of the patients in the tiotropium group and about 21% of the patients in the ipratropium group prematurely withdrew before the scheduled end date. To deal with the data of patients with incomplete follow-up, we applied the multiple imputation (MI) propensity score method and compared the results with those of various non-principled methods. It was shown that the point estimate and the 95% CI of the difference in costs between the treatments were considerably affected by the different methods. Many economic evaluations of chronic diseases are based on prospective data collected during longitudinal studies. In these studies, the occurrence of incomplete data due to premature withdrawal of patients is a common problem, but the attention that has been paid to this problem in relation to the analysis of costs in economic evaluations is surprisingly small.

We designed a simulation study to investigate how standard methods for dealing with incomplete data performed when applied to cost data with various distributions and various types of dropout. The simulations constituted the creation of samples with patient characteristics and costs, the creation of dropout in these samples, and the analysis of the incomplete samples. In the different sets of samples costs were distributed either 1) multivariate normal; 2) multivariate lognormal; or 3) multivariate lognormal enlarged with costs of events. Dropout in each sample was created either 1) completely at random (i.e. the occurrence of dropout

is independent from the observed and unobserved data; DCAR); 2) at random (i.e. the occurrence of dropout is related to data that are observed in the periods before dropout; DAR); or 3) as informative dropout (i.e. the occurrence of dropout is related to the unobserved data; ID). Various naïve and principled methods were selected to deal with the missing data. Selected principled methods included the product-limit estimator of Lin et al., the expectation maximisation (EM) algorithm and various types of multiple imputation.

Almost all methods were unbiased in case of dropout completely at random (DCAR), but only the principled methods provided adequate estimates of the SE. The best estimates of mean and SE in case of dropout at random (DAR) were provided by the bootstrap EM algorithm, MI regression and MI Monte Carlo Markov Chain. These methods were able to deal with skewed cost data in combination with DAR and only became biased when costs also included the costs of expensive events. None of the methods was able to deal adequately with informative dropout. Based on these results it was concluded that, the EM algorithm with bootstrap, MI regression and MI MCMC are robust to the multivariate normal assumption and are the preferred methods for the analysis of incomplete cost data when the assumption of DCAR is not justified. The analyses also showed that skewness may not be the only problem of data characterised by high costs of clinical events, and that mixed or bimodal distributions and large variation in costs within a patient over time may constitute the real obstacles for the adequate analysis of incomplete cost data. In data sets with such complex mixed distributions and with substantial and different dropout rates between treatment groups, the applicability of standard methods to analyse the incomplete data should be questioned.

Applicability of results to other diseases | The findings of this thesis may apply to many diseases other than COPD. Standardisation and generalisability are essential determinants of all studies incorporating resource use and costs. The impact of high-cost events and the problem of incomplete data due to premature withdrawal, as discussed in this thesis, deserve more attention in virtually all economic evaluations.

SAMENVATTING

Inleiding | Gedurende de laatste decennia zijn de kosten van de gezondheidszorg in Westerse landen snel gestegen en werd het duidelijk dat grenzen gesteld dienden te worden aan deze groei. De behoefte ontstond om nieuwe technologieën niet automatisch te laten instromen in de zorg, maar deze eerst te beoordelen om zodoende gefundeerde beslissingen te kunnen nemen inzake registratie, vergoeding en prijsvorming. Deze ontwikkelingen hebben geleid tot een sterke groei van het aantal economische evaluaties op allerlei terreinen in de gezondheidszorg en diverse auteurs hebben gepleit voor standaardisatie van methodologie, teneinde de vergelijkbaarheid van economische evaluaties te verbeteren.

Dit proefschrift beoogt een bijdrage te leveren aan de vergelijkbaarheid en generaliseerbaarheid van economische evaluaties door het standaardiseren en verbeteren van methoden voor het berekenen en analyseren van kosten. Aan de hand van de economische evaluatie van tiotropium bij patiënten met COPD zullen de volgende methodologische onderwerpen behandeld worden:

- de standaardisatie en vergelijkbaarheid van kosten in economische evaluaties;
- de generaliseerbaarheid van de resultaten van economische evaluaties naar andere landen;
- de invloed van dure medische gebeurtenissen (Engels: 'events') op de kostenanalyse; en
- de analyse van incomplete kostengegevens als gevolg van voortijdige uitval.

■ **Economische evaluatie van tiotropium bij patiënten met COPD** | Chronisch obstructieve longziekte (COPD) is een aandoening van het respiratoire systeem, gekarakteriseerd door een geleidelijke, progressieve vermindering van de longfunctie die grotendeels onomkeerbaar is. Het anticholinergicum tiotropium is een langwerkende bronchusverwijder voor de behandeling van patiënten met COPD. In dit proefschrift worden de kosten en effecten van de behandeling met tiotropium beschouwd aan de hand van twee complementaire vormen van economische evaluaties. De eerste vorm betreft een prospectieve kosteneffectiviteitsanalyse gekoppeld aan twee gerandomiseerde klinische studies in Nederland en België, waarin tiotropium wordt vergeleken met ipratropium, een kortwerkend anticholinergicum. De tweede vorm van economische evaluatie bestaat uit een besliskundig model, waarbij gegevens worden gebruikt van zes trials die onderdeel zijn van het klinische trialprogramma van tiotropium. Het doel van dit probabilistische Markov model is het vergelijken van de kosteneffectiviteit van tiotropium, salmeterol (een langwerkende beta-agonist) en ipratropium met betrekking tot de behandeling van patiënten met COPD in verschillende landen. Op basis van deze modelmatige analyse worden de resultaten van Nederland vergeleken met die van Canada. Het voordeel van een prospectieve economische evaluatie is dat alle uitkomsten verkregen zijn in een gecontroleerde en gerandomiseerde situatie en gebaseerd zijn op individuele patiëntgegevens uit dezelfde populatie. De modelmatige analyse daarentegen biedt de mogelijkheid om meerdere behandelingen binnen hetzelfde raamwerk te vergelijken, gegevens uit verschillende databronnen te combineren, kostenschattingen aan te passen aan de werkelijke praktijk, bijvoorbeeld door te corrigeren voor kosten die samenhangen met het trial-protocol, en de schattingen van kosten en effecten te generaliseren naar andere situaties en landen.

De prospectieve en de modelmatige analyse resulteren in een vergelijkbare reductie van het aantal exacerbaties als gevolg van de behandeling met tiotropium. Ten opzichte van ipratropium bedraagt de reductie in het aantal exacerbaties per patiënt per jaar circa 35% (verschil: 0.27; 95% betrouwbaarheidsinterval (BI): 0.02; 0.52) en ten opzichte van salmeterol circa 20% (verschil: 0.17, 95% BI: -0.02; 0.37). Verschillen tussen de prospectieve en modelmatige benadering komen met name tot uiting in de kostenschattingen. De prospectieve economische evaluatie laat zien dat de behandeling met tiotropium leidt tot een stijging van de gemiddelde kosten per patiënt ten opzichte van ipratropium van € 180 (95% BI: -268; 627). De modelmatige benadering daarentegen resulteert in lagere kosten voor tiotropium. Een verschil van € 170 (95% BI: -812; 335) ten opzichte van ipratropium en een verschil van € 42 ten opzichte van salmeterol (95% BI: -484; 353). Het verschil met de prospectieve analyse wordt onder meer veroorzaakt door een correctie voor kosten die samenhangen met het trial-protocol, de geschatte prijs van de studiemedicatie ipratropium en het geschatte aantal ernstige exacerbaties. De kosten in Canada waren aanzienlijk lager dan in Nederland en nagenoeg

gelijk in de drie behandelgroepen. De belangrijkste oorzaak voor het verschil tussen de twee landen was de langere opnameduur in geval van een exacerbatie in Nederland.

Standaardisatie en vergelijkbaarheid van kosten in economische evaluaties | In de economische evaluaties van tiotropium is voor het meten en waarderen van zorggebruik gebruik gemaakt van gegevens uit de 'Handleiding voor Kostenonderzoek: methoden en richtlijnprijzen voor economische evaluaties in de gezondheidszorg'. Deze handleiding is uitgebracht als reactie op de verschillen tussen economische evaluaties met betrekking tot de wijze waarop kosten worden gemeten en gewaardeerd. Met het uitbrengen van de handleiding is beoogd de kwaliteit en vergelijkbaarheid van kostenbepalingen in Nederlandse economische evaluaties te verbeteren.

In dit proefschrift, wordt de inhoud van de handleiding bediscussieerd en worden een aantal kernbegrippen met betrekking tot de standaardisatie van kosten behandeld. In de handleiding wordt een stappenplan voor kostenonderzoek geïntroduceerd, bestaande uit de volgende stappen: 1) het bepalen van de reikwijdte van het onderzoek; 2) de keuze van kostencategorieën; 3) de identificatie van eenheden; 4) de volumemeting; 5) de waardering van eenheden en 6) het berekenen van kostprijzen. Elke stap gaat gepaard met het maken van keuzes die uiteindelijk vastleggen op welke wijze de kostenbepaling wordt uitgevoerd. Kernbegrippen met betrekking tot de standaardisatie van kosten zijn 1) uniforme uitgangspunten; 2) methoden voor het meten en berekenen van kosten; 3) standaardkosten ('richtlijnprijzen'); 4) standaardrekenwaarden en 5) de rapportage van kosten. Door het gebruik van de verschillende elementen wordt beoogd een goede balans te vinden tussen de standaardisatie van kosten en het verschaffen van voldoende mogelijkheden om recht te doen aan de specifieke situatie waarin het onderzoek wordt uitgevoerd.

Bij veel aandoeningen worden de totale kosten van behandeling in belangrijke mate bepaald door de kosten van verpleegdagen in ziekenhuizen. Om inzicht te krijgen in de kosten van verpleegdagen is een onderzoek uitgevoerd waarbij kostprijzen zijn bepaald van ligdagen op 22 verpleegafdelingen en 11 intensive care afdelingen van in totaal 15 ziekenhuizen. De gemiddelde kosten per verpleegdag varieerden van € 230 voor een verpleegdag in een algemeen ziekenhuis, tot € 323 in een universitair ziekenhuis en tot € 1125 voor een ligdag op een intensive care afdeling. Het aandeel van de kosten voor verpleging varieerde van 38% op een verpleegafdeling tot 48% op een intensive care afdeling. Dit onderzoek laat zien dat ondanks het gebruik van uniforme methoden voor het verzamelen en berekenen van kosten er een grote variatie kan bestaan in kosten verkregen uit verschillende centra. Deze verschillen kunnen lang niet altijd verklaard worden door verschillen in behandeling of patiëntenpopulatie. De resultaten van dit onderzoek zijn gebruikt voor het berekenen van een standaard kostprijs voor verpleegdagen. Het gebruik van standaard kostprijzen uit de handleiding kan bijdragen aan het voorkomen van verschillen tussen economische evaluaties die geen verband houden met werkelijke verschillen in zorggebruik en aldus bijdragen aan de standaardisatie en vergelijkbaarheid van economische evaluaties.

Generaliseerbaarheid van economische evaluaties naar andere landen | De prospectieve economische evaluatie van tiotropium was gebaseerd op 2 klinische trials in Nederland en België. Informatie omtrent de kosteneffectiviteit van tiotropium in andere landen ontbrak en de vraag rees op welke wijze deze informatie verkregen kon worden. Terwijl vaak mag worden aangenomen dat het biologische effect van een behandeling min of meer gelijk is voor patiënten uit verschillende landen, mag deze aanname voor zorggebruik en kosten niet zomaar gemaakt worden. Er zijn diverse factoren die bijdragen aan verschillen in zorggebruik en kosten tussen landen en het is duidelijk dat generalisatie van zorggebruik en kosten naar andere landen niet zonder meer mogelijk is.

Het primaire doel van de modelmatige economische evaluatie van tiotropium, was het bepalen van de kosteneffectiviteit van tiotropium in verschillende landen. Gezien dit doel, zijn drie eigenschappen van het model met name van belang. Ten eerste is het model, in tegenstelling tot deterministische modellen, volledig probabilistisch, waardoor de onzekerheid in kosten, effecten en kosteneffectiviteit geanalyseerd kan worden. Ten tweede zijn alle input parameters van het model die gerelateerd zijn aan de effectiviteit van de behandelingen gebaseerd op empirische trial gegevens. Hierdoor is het mogelijk om de uitkomsten van het model te valideren aan de hand van de uitkomsten van de klinische trials en daarmee de transparantie en acceptatie van het model te vergroten. Ten derde zijn alle schattingen van zorggebruik en kostprijzen gekoppeld aan ziektestadia en exacerbaties en deze schattingen zijn voor alle behandelgroepen hetzelfde. Er hoeven dus geen aannames gemaakt te worden met betrekking tot verschillen in zorggebruik tussen behandelgroepen. Behoudens het verschil in de prijs van studiemedicatie hangen alle verschillen in kosten tussen behandelgroepen samen met verschillen in overgangs- en exacerbatiekansen. Deze eigenschap kan beschouwd worden als een minimum voorwaarde voor besliskundige modellen die beogen de kosten en kosteneffectiviteit te bepalen in verschillende landen. Door deze eigenschap wordt het tevens mogelijk om schattingen van zorggebruik niet alleen te baseren op gegevens die in de klinisch trial verzameld zijn, maar kunnen deze gegevens ook ontleend worden aan andere databronnen.

De invloed van dure medische gebeurtenissen op de analyse van kosten | In een post-hoc analyse van gegevens ontleend aan de prospectieve kosteneffectiviteitsanalyse van tiotropium is een berekening gemaakt van de kosten van COPD exacerbaties. Schattingen van de gemiddelde (standaarddeviatie) kosten van exacerbaties varieerden van € 86 (223) voor een milde exacerbatie tot € 579 (1227) voor een gemiddelde exacerbatie tot € 4007 (5922) voor een ernstige exacerbatie. Zestien procent van de exacerbaties resulteerde in een opname en deze exacerbaties waren verantwoordelijk voor 90% van de totale kosten van exacerbaties. Deze karakteristieke situatie waarin een klein aandeel van de patiënten zeer hoge kosten heeft en het grootste deel van de patiënten lage of geen kosten, zorgt voor een zeer scheve verdeling van de kosten met belangrijke consequenties voor het berekenen en analyseren van kosten. Omdat veel empirische economische evaluaties worden uitgevoerd in samenhang met een gerandomiseerde klinische studie waarbij de steekproefomvang gebaseerd is op klinische uitkomstmaten, zal het onderscheidingsvermogen van de studie doorgaans niet voldoende zijn om verschillen in kosten aan te tonen die, bij het gebruik van conventionele waarden,

'statistisch significant' zijn. De grote variantie in kosten en kosteneffectiviteit is een van de belangrijkste motivaties voor de opvatting dat een adequate beschrijving van onzekerheid relevanter is voor de besluitvorming dan de klassieke benadering gericht op het testen van hypotheses waarbij gebruik wordt gemaakt van een arbitrair afkappunt met betrekking tot statistische significantie.

Ondanks de hoge kosten van exacerbaties en de resulterende scheve verdeling van kosten in de prospectieve economische evaluatie van tiotropium, was de steekproefomvang groot genoeg om gebruikelijke methoden voor de analyse van kosten toe te passen. Bootstrap-simulatie en schattingen gebaseerd op de aannames van een normale verdeling resulteerden in nagenoeg identieke betrouwbaarheidsintervallen van het verschil in kosten. Voor het analyseren van de onzekerheid rondom de kosteneffectiviteitsratio's zijn bestaande uitgangspunten in praktijk gebracht. In de prospectieve economische evaluatie, is de incrementele kosteneffectiviteit van tiotropium versus ipratropium gepresenteerd aan de hand van een grafische weergave op het 'kosteneffectiviteitsvlak' (Engels: CE-plane) en aan de hand van de 'kosteneffectiviteit-aanvaardbaarheidscurves' (CE-acceptability curve). Op het kosteneffectiviteitsvlak worden ellipsen gepresenteerd die 5%, 50% en 95% van de kansmassa van het verschil in kosten en effecten omvatten en faciliteren daarmee een snelle en visuele interpretatie van de onzekerheid rondom de ratio. De kosteneffectiviteit-aanvaardbaarheidscurve geeft voor verschillende waarden van de kosteneffectiviteitslimiet, de kans weer dat een behandeling kosteneffectief is. Wanneer de limiet per voorkomen exacerbatie dan wel per patiënt met een relevante verbetering op de St. George's Respiratory Questionnaire wordt gesteld op bijvoorbeeld € 2000, dan is de kans dat tiotropium acceptabel is respectievelijk 80% en 72%. In de modelmatige economische evaluatie van tiotropium werden drie behandelingen vergeleken binnen hetzelfde raamwerk. Gebaseerd op de aanname dat deze behandelingen elkaar wederzijds uitsluiten, zijn, in plaats van incrementele curves voor de directe vergelijking van twee behandelingen, aanvaardbaarheidscurves gepresenteerd voor elke behandeling afzonderlijk. De aanvaardbaarheidsgrens volgt de curve van de behandeling met het hoogste verwachte netto profijt en bepaalt aldus voor elke waarde van de limiet wat de optimale behandeling is vanuit het perspectief van het kosteneffectiviteitscriterium. De aanvaardbaarheidscurve voor exacerbaties laat zien dat in Nederland en Canada tiotropium het hoogste verwachte netto profijt had voor alle waarden van de kosteneffectiviteitslimiet boven respectievelijk € 0 en € 10.

Analyse van incomplete kostengegevens vanwege voortijdige uitval | In de prospectieve economische evaluatie van tiotropium completeerde 83% van de patiënten de studie. Vijftien procent van de patiënten in de tiotropium groep en 21% van de patiënten in de ipratropium groep viel voortijdig uit. De gegevens van deze uitvallers zijn geanalyseerd door het toepassen van een vorm van multiple imputatie (MI), de propensity score methode. De resultaten van deze analyse zijn vergeleken met de uitkomsten van analyses verkregen door toepassing van een aantal niet-formele methoden. De analyses resulteerden in aanzienlijke verschillen in de puntschatting en het 95% BI van het verschil in kosten tussen de behandelgroepen. Veel economische evaluaties van chronische ziekten zijn gebaseerd op prospectieve, longitudinale studies, waarin voortijdige uitval van patiënten een veel voorkomend probleem zal zijn.

Desalniettemin is tot op heden nauwelijks aandacht besteed aan het probleem van voortijdige uitval in relatie tot de analyse van kosten in economische evaluaties.

Om te onderzoeken hoe standaardmethoden voor de analyse van incomplete gegevens presteren indien toegepast op kostengegevens, is in dit proefschrift een simulatiestudie opgezet. De simulaties bestaan uit het aanmaken van samples met patiëntkenmerken en kosten, het creëren van uitval in deze samples en de analyse van de incomplete samples. Kosten in de verschillende steekproeven worden gekenmerkt door één van de volgende verdelingen: 1) multivariaat normaal; 2) multivariaat lognormaal of 3) multivariaat lognormaal vermeerderd met de kosten van medische gebeurtenissen. Daarnaast is in elke sample één van de volgende vormen van uitval gecreëerd: 1) volslagen willekeurig, waarbij uitval onafhankelijk is van de geobserveerde en niet-geobserveerde data ('dropout completely at random'; DCAR), 2) willekeurige uitval, waarbij uitval gerelateerd is aan de geobserveerde data ('dropout at random'; DAR) of 3) informatieve uitval, waarbij de uitval gerelateerd is aan niet-geobserveerde data ('informative dropout'; ID). Vervolgens zijn verschillende naïeve en formele methoden geselecteerd voor het analyseren van de incomplete samples.

Nagenoeg alle methoden geven goede schattingen van de gemiddelde kosten in geval van DCAR, maar alleen de formele methoden resulteren in adequate schattingen van de SE. De beste schattingen van de gemiddelde kosten en de SE in geval van DAR worden verkregen door toepassing van het EM-algoritme met bootstrap, MI regressie en MI MCMC. Deze methoden resulteren in goede schattingen bij multivariaat lognormaal verdeelde kosten in combinatie met DAR en vertekening in de uitkomsten ontstaat alleen wanneer de kosten tevens de kosten van dure medische gebeurtenissen omvatten. Geen van de methoden presteerde goed in geval van ID. Op basis van deze resultaten is geconcludeerd dat het EM-algoritme met bootstrap, MI regressie en MI MCMC robuust zijn voor afwijkingen van de vereiste van een multivariaat normale verdeling. Deze methoden hebben dan ook de voorkeur boven andere methoden, wanneer de aanname van DCAR niet gerechtvaardigd is. Dit onderzoek laat ook zien dat een scheve verdeling niet het enige probleem is van kosten die mede bestaan uit hoge kosten van medische gebeurtenissen. Samengestelde of bimodale verdelingen en een grote variatie in kosten bij een patiënt in de tijd vormen de grootste obstakels voor de adequate analyse van incomplete kostendata. In geval van kosten met zulke complexe, samengestelde verdelingen en met aanzienlijke, ongelijke uitvalpercentages in de behandelgroepen, dient betwijfeld te worden of de dataset met behulp van standaardmethoden voor de analyse van incomplete kostengegevens geanalyseerd kan worden.

Geldigheid van de resultaten voor andere aandoeningen | De bevindingen van dit proefschrift hebben niet alleen betrekking op COPD, maar gelden ook voor een groot aantal andere aandoeningen. Standaardisatie en generaliseerbaarheid zijn essentiële determinanten van bijna alle studies op het gebied van zorggebruik en kosten. In nagenoeg alle economische evaluaties zou meer aandacht geschonken moeten worden aan de problemen van hoge kosten van medische gebeurtenissen en incomplete gegevens als gevolg van voortijdige uitval zoals beschreven in dit proefschrift.

DANKWOORD

'We naderen de toekomst, Joost. Als we daar zijn kunnen we even uitrusten', aldus Olivier B. Bommel tot zijn bediende in de cartoon die de wand van de koffiekamer op de BMG lange tijd sierde. De noodzaak om uit te rusten zal, denk ik, wel meevallen, maar het proefschrift is af en dat was lange tijd toekomst. Mijn dank gaat uit naar de vele mensen die de afgelopen jaren hun interesse en betrokkenheid bij mijn proefschriftproces getoond hebben. Zeker de laatste tijd kon ik niet meer op recepties of feestjes verschijnen zonder na het introducerende 'Jan, hoe is het?', onmiddellijk de 'proefschriftvraag' voorgeschoteld te krijgen. Toch was al die aandacht fijn en voor mij een belangrijke motivatie om door te gaan. Verder gaat mijn dank uit naar de co-auteurs, collega's en oud-collega's die op velerlei wijze bij dit proefschrift betrokken zijn geweest. Promotieclub, Elly, Herman, Marten, Xander en heel veel anderen, bedankt voor zowel het enthousiasme als geduld waarmee jullie alle versies van discussieparagrafen, samenvattingen, doelstellingen, proefschrifttitels en stellingen de afgelopen tijd van commentaar en suggesties hebben voorzien. Jullie input heeft daarmee in belangrijke mate bijgedragen aan de uiteindelijke focus en vormgeving van dit proefschrift.

Een aantal personen wil ik graag apart bedanken. En natuurlijk als eerste, Maureen. Na het glaucoom-onderzoek was het voor mij een bewuste stap om met jou als copromotor een promotietraject te starten. Daar heb ik nooit spijt van gekregen. Als geen ander was je telkens bereid om tijd te steken in en commentaar te leveren op voorstellen, abstracts en artikelen die ik schreef en zeker in het begin waren vaak vele versies nodig voordat de rode gloed van de correctiepen geheel achterwege bleef. Je hebt dan ook een grote invloed gehad op mijn schrijfstijl, publicaties en zeker ook dit proefschrift. Het tiotropium-onderzoek was een grote en wel heel langdurige klus, maar zorgde ook voor de mogelijkheid om ons te verdiepen in de vele aspecten gerelateerd aan de analyse van kosteneffectiviteit en onzekerheid. Jouw streven om de analyse steeds beter te doen zorgde voor een belangrijke motivatie om nieuwe technieken toe te passen en ons deze eigen te maken. Precies datgene wat het onderzoek zo leuk maakt. Ik prijs mij gelukkig dat jij mijn copromotor hebt willen zijn.

Beste Frans, ik had niet gedacht dat de uitnodiging in 1998 om een voorstel te schrijven voor de Ziekenfondsraad voor het opstellen van een standaardkostenlijst zou leiden tot een boekwerk dat later zoveel aandacht heeft gekregen. Jouw uitnodiging om mij het 'kostenhandleiding' project te laten doen heeft een belangrijke impuls gegeven aan mijn positie en carrière als onderzoeker. Thans staat het handleidingonderzoek ook nog aan de basis van het huidige boekwerk. Veel dank.

Brigitta, not often will researchers at iMTA have been confronted with a 'pharmaceutical counterpart' with so much knowledge about their research. We had to work hard to keep up with you in reading and understanding all the new literature about COPD and MTA research. Discussions about the tiotropium studies never concerned the outcomes of the study, but were always related to the methodological aspects of the analysis. Working with you and many of your colleagues from Boehringer Ingelheim has been a pleasure.

Maiwenn, bedankt voor je uitleg en advies bij al die statistische onderwerpen die ik de afgelopen jaren aan je heb voorgelegd. Wie had gedacht dat zo'n eenvoudige vraag als: 'we hebben een paar patiënten met incomplete follow-up, wat zullen we daar mee doen?', tot zoveel onderzoek heeft kunnen leiden. Ook nu nog zijn er veel ideeën om het dropout-onder-

zoek verder uit te werken. Pas dus nog even goed op al mijn mappen met missing data. The 'Al' in my publications matters.

Paul, je directe bijdrage aan dit proefschrift mag dan gering zijn; toch met recht in het dankwoord want jij maakt mijn onderzoeksbestaan veel leuker. Met jou als kamergenoot, op wat ongetwijfeld de gezelligste kamer van de BMG geweest is de afgelopen 8 (jaja!) jaar, geen kans om weg te zinken in eenzaam onderzoek. Bedankt voor het scherp houden van geest en snelheid. Bedankt voor alle vrolijkheid.

Marjolein, lieve Marjolein. Gaan samenwonen met jou is de beste beslissing die ik de laatste jaren genomen heb en duurt al bijna net zolang als het schrijven van een proefschrift. Proefschrift is klaar, wij nog niet. Nog 1 keer een 'promotiehoedje' als Sinterklaassurprise. Daarna samen verder naar nieuwe toekomst.

COLOFON

Financial support by Boehringer Ingelheim GmbH,
Ingelheim, Germany, and Pfizer Inc,
New York, USA, for the publication of this thesis
is gratefully acknowledged.

Lay-out, graphics and cover Maaike Klijn
Printed by Ridderprint, Ridderkerk

ISBN 90-5335-039-X