

Cost and effects of pharmacotherapy in asthma and COPD

Citation for published version (APA):

Rutten-van Mólken, M. P. M. H. (1994). *Cost and effects of pharmacotherapy in asthma and COPD*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.19941124mr>

Document status and date:

Published: 01/01/1994

DOI:

[10.26481/dis.19941124mr](https://doi.org/10.26481/dis.19941124mr)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Costs and effects of pharmacotherapy in asthma and COPD

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Rutten-van Mólken, Maureen Petronelle Marie Hubertine

Costs and effects of pharmacotherapy in asthma and COPD /
Maureen Petronelle Marie Hubertine Rutten-van Mólken. –
Maastricht : Universitaire Pers Maastricht. – Ill.
Thesis Maastricht. – With ref.

ISBN 90-5278-158-3

Subject headings: pharmacotherapy ; economic aspects. /
asthma.

Cover design: RL Design, Maastricht

Production: Datawyse | Universitaire Pers, Maastricht

Lay-out: Jan van Emmerik, Maastricht

Sponsor: GLAXO B.V., Zeist

Costs and effects of pharmacotherapy in asthma and COPD

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Rijksuniversiteit Limburg te Maastricht, op gezag van de Rector Magnificus, Prof.Dr. H. Philipsen, volgens het besluit van het College van Dekanen, in het openbaar te verdedigen op donderdag 24 november 1994 om 14.00 uur

door

Maureen Petronelle Marie Hubertine Rutten-van Mólken
geboren te Sittard op 6 juli 1966

Promotores:

Prof.Dr. F.F.H. Rutten (oud hoogleraar Rijksuniversiteit Limburg/Erasmus
Universiteit Rotterdam)

Prof.Dr. E.K.A. Van Doorslaer (Universitaire Instellingen Antwerpen)

Beoordelingscommissie:

Prof.Dr. J.M.J.P. van der Linden (voorzitter)

Prof.Dr. R.M. Leidl

Prof.Dr. D.S. Postma (Rijksuniversiteit Groningen)

Prof.Dr. C. van Weel (Katholieke Universiteit Nijmegen)

Prof.Dr. E.F.M. Wouters

*hij zag het over het hoofd
het lieve en het schone
omdat hij op zijn hoge trap
met al zijn koele wetenschap
geen oog had voor 't gewone*

Toon Hermans

Voor mijn ouders

Contents

1. General introduction	1
1.1 Abstract	2
1.2 Prevalence and incidence	2
1.3 Economic burden of illness	3
1.4 Pharmacotherapy	4
1.5 Economic evaluation of asthma and COPD care	6
1.6 Potential uses for economic evaluations of pharmaceuticals	7
1.7 Issues in the standardization of methodologies	10
1.7.1 Quality of life	11
1.7.2 Utility measurement	12
1.7.3 Analysing cost data	14
1.8 Aims and outline of this thesis	15
1.9 References	17
2. Economic appraisal of asthma and COPD care:	
A literature review 1980-1991	27
2.1 Abstract	28
2.2 Introduction	28
2.3 Selection of the literature	29
2.4 Nature of the economic appraisals	30
2.5 Alternatives compared	35
2.5.1 Health education for adults	35
2.5.2 Health education for children	36
2.5.3 Pulmonary rehabilitation	36
2.5.4 Domiciliary oxygen therapy	37
2.5.5 Home care	38
2.5.6 Aerosol delivery of bronchodilator drugs	39
2.5.7 Mechanical ventilation	40
2.5.8 Alpha-1 Antitrypsin Replacement Therapy	40
2.6 Outcomes	41
2.6.1 Clinical measures	41
2.6.2 Disease-specific quality of life measures	41
2.6.3 Generic quality of life measures	42
2.6.4 Utilities and QALYs	42
2.7 Costs	43
2.7.1 Costs of organizing and operating a particular health care programme	43
2.7.2 Costs of utilization of other health services	44
2.7.3 Out-of-pocket expenses of patients and their families	44

2.7.4	The loss of productivity due to absence from work, imputed volunteer time or family time	45
2.7.5	Costs versus expenditures and charges	46
2.7.6	Discounting and sensitivity analysis	46
2.8	Quality of the effectiveness evidence	47
2.9	Conclusions and discussion	49
2.10	Acknowledgements	51
2.11	References	51
3.	Cost-effectiveness of inhaled corticosteroid plus bronchodilator therapy versus bronchodilator monotherapy in children with asthma	55
3.1	Abstract	56
3.2	Introduction	56
3.3	Methods	58
3.3.1	Clinical endpoints	59
3.3.2	Economic endpoints	60
3.3.3	Statistical analysis	61
3.4	Results	62
3.4.1	Baseline characteristics	62
3.4.2	Withdrawals	63
3.4.3	Effectiveness	64
3.4.4	Costs	65
3.5	Incremental cost-effectiveness	69
3.6	Discussion	71
3.7	Acknowledgements	75
3.8	References	75
4.	Costs and effects of inhaled corticosteroids and bronchodilators in asthma and COPD	79
4.1	Abstract	80
4.2	Introduction	80
4.3	Methods	82
4.3.1	Clinical endpoints	83
4.3.2	Economic endpoints	83
4.3.3	Statistical analyses	84
4.4	Results	85
4.4.1	Baseline characteristics	85
4.4.2	Withdrawals	86
4.4.3	Efficacy	88
4.4.4	Costs	89
4.4.5	Subgroup analyses	92
4.4.6	Cost-effectiveness ratio	93

4.5	Conclusion and discussion	94
4.6	Acknowledgements	97
4.7	References	98
5.	Two-year bronchodilator treatment in patients with mild airflow obstruction: Contradictory effects on lung function and quality of life	101
5.1	Abstract	102
5.2	Introduction	102
5.3	Patients and methods	103
5.3.1	Patients	103
5.3.2	Study design	104
5.3.3	Outcome measures	105
5.3.4	Statistical analysis	106
5.4	Results	107
5.4.1	Baseline patient characteristics	107
5.4.2	Baseline quality of life impairment compared to the general population	108
5.4.3	Effect of treatment on lung function and quality of life: Salbutamol versus ipratropium bromide	110
5.4.4	Effect of treatment on lung function and quality of life: Continuous versus symptomatic	111
5.4.5	Within-patient comparison of lung function and quality of life	113
5.5	Discussion	113
5.6	Acknowledgements	116
5.7	References	117
6.	Comparing the performance of four different instruments in evaluating the effects of salmeterol on asthma quality of life	119
6.1	Abstract	120
6.2	Introduction	120
6.3	Methods	122
6.3.1	Patients	122
6.3.2	Study design	122
6.3.3	Clinical outcome measures	123
6.3.4	Quality of life instruments	123
6.3.5	Analysis	126
6.4	Results	127
6.4.1	Patients	127
6.4.2	Clinical outcome measures	128
6.4.3	Responsiveness	129
6.4.4	Cross-sectional validity	131
6.4.5	Longitudinal validity	132

6.4.6	Reproducibility	134
6.5	Discussion	134
6.6	Acknowledgements	139
6.7	References	139
7. Methodological issues of patient utility measurement:		
	Experience from two clinical trials	143
7.1	Abstract	144
7.2	Introduction	144
7.3	Methods	146
7.3.1	Patients and studies	146
7.3.2	Methods of utility measurement	147
7.3.3	Statistical analysis	150
7.4	Results	151
7.4.1	Comparing rating scale and standard gamble utilities	151
7.4.2	Reliability of the rating scale and the standard gamble	153
7.4.3	Internal consistency of the standard gamble	156
7.4.4	A health state worse than death	159
7.5	Conclusions and discussion	161
7.6	Acknowledgements.	164
7.7	References	164
8. Statistical analysis of cost outcomes in a randomized controlled clinical trial		
	clinical trial	169
8.1	Abstract	170
8.2	Introduction	170
8.3	The asthma/COPD trial	172
8.4	Problems encountered in the asthma/COPD trial	173
8.4.1	Problem 1: Selective treatment-related withdrawal	173
8.4.2	Problem 2: Skewed cost distributions	176
8.5	Selective withdrawal and cost estimation	177
8.5.1	Complete data only	177
8.5.2	Cumulative cost approach	178
8.5.3	Patient-year approach	178
8.5.4	Comparing the approaches	179
8.6	Skewness and statistical testing	180
8.7	Discussion	186
8.8	Acknowledgements	187
8.9	References	188

9. General discussion and areas for future research	191
9.1 Specific pharmaco-economic evaluations	192
9.2 Economic evaluation of screening	194
9.3 Population-based studies	196
9.4 Further development of quality of life measures	197
9.5 Linking utility measures and generic quality of life measures	199
9.6 Validity of the utility measures	199
9.7 Generalizability of the approach suggested for analyzing skewed cost data	200
9.8 References	201
 Samenvatting	 203
Dankwoord	209
Curriculum Vitae	213

1

General introduction

1.1 Abstract

This thesis addresses various aspects of cost-effectiveness in the treatment of asthma and chronic obstructive pulmonary disease (COPD). Part of the thesis deals with methodological aspects of measuring and analysing quality of life, utilities and costs, but it also presents results from the empirical implementation of the proposed methods. Although it is hoped that the methodological findings will be of interest in their own right, the choice of asthma and COPD as the focus for these studies will first be justified by providing a brief description of the importance of asthma and COPD, the scale of the economic burden which they present, and of the pharmacotherapy costs and options which are available. Given the clinical and economic importance of pharmacotherapy in the treatment of asthma and COPD, it is surprising that there have been very few full cost-effectiveness studies of pharmacotherapies. Of those which were found, several had serious methodological shortcomings. The justification for more cost-effectiveness studies of pharmaceuticals has been strengthened in The Netherlands by recent proposals to include cost-effectiveness as an additional evaluation criterion in the future. It is hoped that this thesis will go some way towards addressing both the lack of methodological clarity and the need for empirical studies in this important field.

1.2 Prevalence and incidence

Asthma is a chronic inflammatory disease of the airways characterized by reversible airflow obstruction associated with exacerbations of coughing, wheezing, chest tightness and difficult breathing.¹ It is usually encountered in young, allergic and non-smoking subjects. COPD, which refers primarily to chronic bronchitis and emphysema, is characterized by chronic airflow obstruction causing progressive loss of lung function and is associated with symptoms of chronic cough with purulent sputum.² COPD is usually encountered in older, often non-allergic subjects. Smoking is the most important risk factor for the development of this disease. In routine clinical practice patients often present overlapping symptoms and are not easily characterized as having either asthma or COPD. Furthermore, the distinction between the diagnoses is difficult because long-standing asthma may evolve into COPD.

Asthma and COPD are major health problems in the Netherlands, which consume an increasing share of health care resources. The exact extent of the problem is, however, difficult to measure. Prevalence estimates vary considerably according to the source used. Epidemiological studies in the Dutch population have produced prevalence estimates of about 80 per 1000 persons under 65 years of age.³ The number of patients known to general practitioners is considerably lower. The Continuous Morbidity Registration Project in General Practitioners from Nijmegen

(CMR-Nijmegen) shows that the standardized prevalence rate of asthma and COPD in general practice is approximately 36 per 1000 persons.⁴ The prevalence is higher in men (almost 45 per 1000) than in women (28 per 1000). Men's higher prevalence rates may be due to their higher smoking or ex-smoking rates and higher occupational exposure to toxic materials. Asthma is most frequently present in people under 25 years of age whereas COPD is more common in people over 65 years of age. The prevalence of asthma decreases with age whereas the prevalence of COPD increases with age.⁵ Based on demographic projections which show the population becoming ever older, the prevalence rates of asthma and COPD are expected to increase by 4% and 23%, respectively, between 1990 and 2005.³

The real increase may be higher, because these projections do not take into account recent efforts to reduce the under-diagnosis of these diseases. It is estimated that the percentage of asthma and COPD patients recognized by physicians will increase from 50% in 1990 to 75% in 2005.³ Several studies in The Netherlands and other countries have reported under-diagnosis or inappropriate diagnosis of asthma and COPD.⁶⁻¹¹ This may explain part of the difference between prevalence estimates based on epidemiological studies and those based on registration projects in general practice.

Information on the incidence, as opposed to the prevalence, of asthma and COPD is relatively scarce, but the ratio between incidence and prevalence is known to be about 1:7.⁵ In general the incidence reveals much the same pattern as the prevalence.

1.3 Economic burden of illness

The burden of asthma and COPD for the individual and for society is measurable in terms of premature death, decreased productivity and the use of health care services. In 1990 asthma and COPD (as defined by the ICD-9 codes 490-496) caused 6.4% of total mortality in men and 2.9% of total mortality in women. About 4% of disability and 0.4% of the absence from work due to illness is caused by asthma and COPD.¹² The latter figure must be viewed with caution, since Dutch sick leave statistics cover only half of the work force, and the reason for the absence is often not known.

Most asthma and COPD patients are treated by the general practitioner.¹³⁻¹⁵ Compared to other chronic diseases in general practice, the "use-need" ratio of asthma and COPD is relatively low.¹⁶ For every 100 known asthma and COPD patients, only 30 will visit their general practitioner in any 3 month period.¹⁶ Only 5 to 10% of the patients are referred to a specialist. Asthma and COPD as primary diagnoses account for 1.6% (25,000) of the total number of hospitalizations and 2.3% (420,000) of the total number of hospitalization days. Three quarters of all hospitalizations for these diseases were emergency hospitalizations.³ Recently a

significant increase has been observed in hospital admissions due to asthma in the 0-4 years age group and in the male 5-9 years age group.¹⁷ It has been estimated that, in 1990, asthma and COPD were associated with about 216 general practitioner contacts, 93 specialist contacts, 90 community nurse contacts, 6 hospitalizations with about 98 hospitalization days and 182 nursing home days per 100 patients.³ Based on demographic projections, by the year 2005, health care utilization in the Netherlands due to these diseases will increase by 11% for general practitioners contacts, 14% for specialist contacts, 19% for the number of hospitalizations and 28% for the number of nursing home days.³

There have been several recently published estimates of the cost of asthma and/or COPD.¹⁸⁻²³ These estimates show that the distribution of health care resources associated with asthma and COPD is skewed toward control of the advanced stages of the diseases. In the Netherlands, diseases of the respiratory system accounted for somewhat more than 3% of total health care costs in 1988.¹² Health care costs associated with asthma and COPD in particular (ICD-9 codes 490-496) were estimated to be 633 million guilders in 1988, which was approximately 1.1% of the total health care costs. Total health care utilization due to these diseases is highest for people between 65 and 79 years of age and lowest for those between 20 and 44 years of age.³

1.4 Pharmacotherapy

Pharmacotherapy is the most commonly prescribed therapy for asthma and COPD. In more than 50% of the GP visits (more than 70% in asthma) some form of pharmacotherapy is prescribed. Although there is as yet no pharmacotherapy to cure the patients, the available drugs seem fairly effective in controlling the diseases by minimizing symptoms, optimizing lung function and preventing exacerbations. Two main categories of drugs to treat asthma and COPD can be identified: 1) bronchodilators and 2) anti-inflammatory drugs.²⁴ In 1993, both categories belong to the top-15 of commonly prescribed drugs for chronic disorders.²⁵ For chronic patients under 20 years of age these drugs even form the top-2.²⁵

The most commonly used bronchodilators are inhaled β_2 -sympathomimetics (salbutamol, terbutaline, fenoterol, salmeterol, formoterol) and inhaled anticholinergics. These drugs are fairly effective in providing relatively rapid symptom relief in both asthma and COPD. However the safety of long-term maintenance treatment with inhaled β_2 -sympathomimetics is currently under heavy debate, because studies have suggested that this treatment is related to an increased risk of dying from asthma^{26,28} and adverse effects of airways obstruction and hyperresponsiveness.^{29,30} Inhaled anticholinergics may be somewhat more effective in COPD than in asthma.^{31,32} The long-term effects of inhaled anticholinergics are relatively unknown, but have recently been questioned.³⁰ The recently introduced

long-acting β_2 -sympathomimetics salmeterol and formoterol have an acute bronchodilating effect which lasts about twice as long (about 12 hours) as the effects of the short-acting β_2 -sympathomimetics. They seem effective in improving lung function and reducing symptoms and exacerbations³³⁻³⁷ especially during the night.³⁸ However, there have been conflicting results about a possible tolerance to the potential protective effects of salmeterol after long-term use.^{39,40} Salmeterol was selected for evaluation in the quality of life study which is described in chapter 6.

The most commonly used anti-inflammatory drugs are inhaled corticosteroids. Recent studies have shown that inhaled corticosteroids are very beneficial in adult asthmatic patients,⁴¹⁻⁴⁷ but there is less unanimity about the effects of corticosteroids in patients with COPD.⁴⁸⁻⁵⁴ In a recent review of studies on inhaled corticosteroids in asthmatic children, Selroos concludes that pulmonary function was good, symptoms were well controlled and growth rate was not affected.⁵⁵ Moreover, a number of studies in asthmatic children have shown that inhaled corticosteroids decreased bronchial hyperresponsiveness.⁵⁶⁻⁵⁸ National and international guidelines for comprehensive asthma management have recently been published for adults and children.^{1,59-65} These guidelines give an important place to inhaled corticosteroids. Chapter 3 and 4 describe cost-effectiveness studies which include these drugs.

An important problem associated with pharmacotherapy is the relatively low compliance of patients. In 50% of patients, or even more, compliance is insufficient.^{66,67} This is particularly troublesome when patients are prescribed inhaled corticosteroids, because these drugs do not have a direct symptom-relieving effect.

The total costs of asthma and COPD drugs were estimated to be NLG 210 million in 1990 (excluding the mark-up for pharmacy expenses). Bronchodilators accounted for 60% of these costs and anti-inflammatory drugs for 40%.⁶⁸ In reality, the costs of pharmacotherapy for asthma and COPD must be higher than this, as some medications such as antibiotics and anti-cough medication, which are very frequently prescribed in case of exacerbations, were not included in these calculations. In asthma and COPD the number of prescriptions for antibiotics is even higher than the number of prescriptions for β_2 -sympathomimetics.⁶⁸ Nevertheless, although the number of prescriptions for anti-asthmatic drugs is high (more than 400 million prescriptions per year), the economic value of this market is relatively low, because most products were introduced between 1960 and 1980 and consequently have a relatively low price.⁶⁹ Depending on the assumptions made, the economic value of the Dutch market for these drugs can be expected to grow at between 1% (scenario in which the risks of pharmacotherapy are stressed, the fear for inhaled corticosteroids increases and the homeopathic market becomes more important) and 12% (scenario in which new drugs and new inhalation techniques are marketed and underdiagnosis and undertreatment is reduced).⁶⁸

1.5 Economic evaluation of asthma and COPD care

Estimates of the total costs of illness, such as those presented above, are of limited value in making decisions about the appropriate allocation of health care resources.^{70,71} At best, they can be seen as baseline estimates of the current economic burden of the disease, against which the impact of changes in treatments can be compared. Of greater value are economic evaluations in which both the costs and benefits of several different diagnostic and treatment options are compared in order to identify the most efficient interventions. However, cost-of-illness studies may be useful to identify areas that need further attention and research. Recent cost-of-illness studies have identified pharmacotherapy as one of the most costly areas of asthma and COPD management, which justifies the concentration on pharmacotherapy in the economic evaluations presented in this thesis.

Economic evaluations can also be used to evaluate whether treatment guidelines recommend the most cost-effective choices. Recent guidelines have recommended increased use of more expensive drugs such as inhaled corticosteroids. This may cause a substantial increase in the cost of medications, but may ultimately reduce overall costs by decreasing morbidity and the need for health care due to asthma and COPD. To test this hypothesis, cost-effectiveness studies of pharmacotherapy, such as those presented in chapter 3 and 4 of this thesis, are needed. Also, the development of new long-acting bronchodilators has stimulated a debate over the most cost-effective mix of drug therapies for symptomatic and prophylactic treatment of asthma and COPD.⁷²

Unfortunately, there have been few economic evaluation studies in the field of respiratory medicine, and even fewer such studies of respiratory drugs. Chapter 2 describes the results of a literature search for economic evaluations of asthma and COPD interventions, showing that between January 1980 and December 1991 only 20 such studies were published. The majority of the studies evaluated the cost-effectiveness of health education programmes, most of which seemed to be rather cost-effective. Three further studies published since 1992⁷³⁻⁷⁵ also dealt with educational interventions and their results were conflicting. The lack of any cost-effectiveness studies of pharmaceutical interventions was most striking. After this review was completed we found two cost-effectiveness studies of respiratory drugs.^{76,77} Both were retrospective, uncontrolled record-based studies, one showing that the addition of cromolyn sodium to common treatment⁷⁶ and the other showing that the use of albuterol inhalers compared to metaproterenol inhalers⁷⁷ led to significant net health care savings. However, these studies were not full cost-effectiveness studies because they used utilization of the health care system as the only outcome, and it is commonly agreed that utilization should be treated as a cost, so that it appears in the numerator of the cost-effectiveness ratio rather than in the denominator.

In 1993 Sculpher and Buxton published a cost-effectiveness analysis of maintenance treatment with formoterol versus salbutamol in patients needing daily

rescue medication during a run-in placebo period.⁷² Although the aim of this analysis was primarily to illustrate a new composite measure of effect, the episode-free day, it clearly showed that the average costs per episode-free day of formoterol were 50 to 60% higher than those for salbutamol. However, they only included the costs of pharmacotherapy and rescue medication.

Thus decision-makers and physicians are poorly provided with information they would need to encourage cost-effective clinical respiratory practice. To ensure that resources are being directed into the most cost-effective areas of asthma and COPD care, we need more well-conducted economic evaluation studies of interventions in this area. Besides calculations of the costs of the treatments under study, clinical trials of asthma should therefore include at least utilization of health care services and absence from work as an outcome measure. Furthermore, since improvement in the quality of life is one of the most important effects to be achieved, quality of life should also become a widely used outcome measure.

1.6 Potential uses for economic evaluations of pharmaceuticals

An influential committee (the Dunning Commission) was asked to advise the State Secretary of Health on ways to put limits on new medical technologies, and to deal with problems caused by scarcity, rationing of health care and the selection of patients for care. Its report, "Choices in Health Care",⁷⁸ which was published in 1992, provides a significant justification for more economic evaluation studies. The committee suggests that the process of deciding what kinds of health care should be included in the basic health insurance package can be seen as a funnel containing four sieves, through which every intervention must pass. The first sieve allows through only necessary care, the second selects for effectiveness, the third for efficiency and the fourth for whether it is possible to leave that care to each individual's own responsibility. The third sieve, in this proposal for setting priorities in health care, explicitly takes into account the cost-effectiveness of care. Specifically with regard to drugs, the committee recommended the introduction of a national list of essential drugs to be complemented by negative lists for certain groups of drugs. It would be desirable to base these lists on cost-effectiveness evidence.

The information generated by the economic evaluation of drugs could potentially be used in decisions about drug approval, price negotiations, reimbursement decisions, decisions about the inclusion of a drug in positive or negative drug usage lists, or in treatment guidelines, and in marketing activities to stimulate the wide adoption of a drug.⁷⁹

In The Netherlands, a company submitting a drug for market approval by the Dutch Medicines Evaluation Board (the Dutch version of the Food and Drug Administration in the USA) only needs to provide evidence of efficacy and safety.

There is no official governmental requirement to undertake economic evaluations, which may explain why the number of economic evaluations of drugs is still limited. Once a drug has been approved, a pharmaceutical company can set its price without any preceding price negotiations.

Approval of a drug does not necessarily imply that it will be reimbursed. Reimbursement decisions are taken separately by the State Secretary of Health after consulting the National Health Insurance Executive Board. In The Netherlands, reimbursement is based on a reference price system (*geneesmiddelen vergoedingen systeem*, or GVS), which was adopted by the Dutch government in 1991. A detailed description of the GVS in English is provided by Redwood.⁸⁰ The essential feature of this system is that 'interchangeable' drugs are placed within the same therapeutic groups. Then the average price for all products in a group is calculated, and the reference price is the price of the drug nearest below the group average. This reference price is the reimbursement limit. The manufacturers are free to charge more than the reference price, but in that event the patient has to pay the difference between the reference price and the actual price. Economic evaluations could play a role in decisions about the assignment of drugs to a therapeutic group. At the moment, such decisions are based only on criteria such as mode of action, range of indications, method of administration etc. Thus, the reference price system does not encourage pharmaceutical companies to undertake economic evaluations. The system seems primarily driven by cost containment concerns and not by value-for-money concerns. Preliminary evaluations of the GVS indicate that it is indeed successful in containing costs.⁸¹

The reference price system does not limit the reimbursement of new, innovative drugs that cannot (yet) be placed in a group of interchangeable drugs. However, in July 1993 the State Secretary of Health limited the reimbursement of all new drugs which are not interchangeable or for which interchangeability has not yet been established, except for new drugs to treat diseases for which no pharmacotherapy was available at all.⁸¹ As a consequence, some new drugs which may have a high therapeutic value may not be granted access to the market.⁸² Again, value-for-money concerns did not play any role in this decision.

As part of the policy associated with the introduction of the reference price system, the government stimulates pharmacotherapeutic consultation (*farmacotherapieoverleg*, or FTO) between physicians and pharmacists on a structural basis. The FTO is intended to rationalize the drug prescription and delivery process and improve its quality and efficiency, partly by means of the development of local treatment guidelines. There is a potential role for economic evaluation to guide this rationalization process.

Economic evaluations could also play a role in determining which pharmaceuticals to include in the benefits package. In 1993, the Second Chamber (i.e. lower house) of the Dutch Parliament carried a motion (the Laning-Boersema motion) to pass the pharmaceutical package through the 'Dunning funnel'. Although the State Secretary of Health has specifically asked that this process should be

focused on the fourth sieve, i.e. on whether a particular drug should be paid for by the patient himself, this motion will also mean that more attention will be given to efficiency, the third sieve of the funnel.⁸¹

As a result of the Laning-Boersema motion, the Dutch government installed a committee (the Van Winzum Commission) to work out the practical application of the four sieves or criteria of the 'Dunning funnel' in the pharmaceutical sector. Instead of passing drugs through the funnel the committee advised to assess drugs only in combination with the indication(s) for which they are used.²⁵ Assessing whether drug treatment is necessary for a particular indication is the first sieve in the decision procedure. The committee has developed a detailed classification system to define levels of necessity. In general, treatment is judged necessary when the disorder leads to premature death or prevents an individual from functioning at a level that is normal for the community. For registered drugs, the second sieve, assessing the efficacy in the restricted sense (called qualitative efficacy by the committee), has already been passed through. Efficiency, the third sieve should be judged in comparison with other drugs for the same indication or by means of a broad community discussion when there is only one drug for an indication. To assess efficiency, costs should be related to community effectiveness (called quantitative efficacy by the committee). This three-step decision process should be applied first to categories of drugs and second to individual drugs within those categories. In contrast to the advice of the State Secretary of Health, the committee did not incorporate the fourth sieve into the decision process because, in general, drugs were judged to be useful, efficient, and relatively cheap compared to other health care services. Thus, when a drug for a specific indication is found to be necessary, effective, and efficient, the drug should be included in the benefits package, regardless of whether the patient can buy that drug on his own expense.²⁵

To summarize, economic evaluation can potentially play a role at various levels of the decision-making process with regard to pharmaceuticals. Thus far, however, national policy has focused mainly on cost-containment rather than on the relation between costs and effects. If efficiency or value-for-money considerations were to be taken into account, the goal would be to improve the overall allocation of health care resources so that total health gains would increase. This implies that drugs that are significantly superior to others or to non-pharmaceutical alternatives such as surgical procedures must, to some extent, be allowed higher reimbursement levels.

In view of the developments described above, and recognizing the possible contributions of economic evaluations in selecting appropriate drug treatments, pharmaceutical companies are becoming increasingly involved in generating cost-effectiveness data, hoping that this may improve their chances under the reference price system.

1.7 Issues in the standardization of methodologies

If economic evaluation is to serve the potential uses as described in the previous section, more attention will have to be given to methodological issues and the approaches used will have to be consistent. Canada and Australia^{83,84} are the only countries in which pharmaceutical companies are required to provide cost-effectiveness information to support applications for reimbursement, and the only countries which provide official methodology guidelines for carrying out these economic evaluations. Although the guidelines of Canada and Australia are both based on the methodological principles of a sound economic evaluation as described by Drummond⁸⁵ they differ, for example with respect to the choice of the alternative, the importance of including indirect costs and benefits (i.e. production losses and gains), and the choice of outcome measures to which the costs can be related.⁸⁶ However, at the national level, such guidelines improve the comparability of study results. This is especially relevant when health care programmes are ranked according to their costs per QALY, i.e. in so called QALY league tables, in order to support decision-making in resource allocation.

The needs to ensure comparability of results and to guarantee high methodological standards are the main motivations for current efforts to standardise economic evaluation methodology.⁸⁷ In The Netherlands, the standardization of methodologies for calculating costs has been stimulated by the recent publication of practical guidelines for cost calculation by the Institute for Medical Technology Assessment.⁸⁸ This book is a toolkit describing the choices, principles and methods in the process of calculating costs. It contains recommendations, but these are not meant to be followed blindly. On the effect side, the formation of a Dutch Working Group on Health Status Assessment has stimulated the standardization of quality of life measurements.^{89,90} This group, which consists of representatives from virtually all Dutch health status research groups, focuses on improving the comparability of quality of life measurements in evaluation research. After reviewing all generic quality of life measurements available in Dutch, the members of this group have recently agreed to use the COOP/WONCA¹⁾ charts¹¹⁴ as a 'hanger-on' instrument to complement other measures in any evaluation study. This agreement ensures a minimal form of standardization.

A certain level of standardization has already been attained without any special efforts or formalization: for example, the terminology of economic evaluation, the use of marginal costs, the importance of considering alternatives, the specification of the viewpoint, the principle of discounting, and the performance of sensitivity analyses.⁸⁷ However, a number of important methodological issues remain controversial, such as the calculation of indirect costs, discounting of health

1) COOP/WONCA: Dartmouth Primary Care Cooperative Information Project / World Organisation of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians.

benefits, the choice of comparison program, methods of measuring quality of life and utilities and ways of analysing differences in the costs of different programmes. The research described in this thesis can contribute especially to discussions in the latter two areas. It addresses theoretical and methodological aspects of quality of life measurements, utility measurements and cost analyses and describes the results of empirical applications of the techniques.

1.7.1 Quality of life

Buxton suggested that one of the reasons for the limited number of cost-effectiveness studies in asthma and COPD is the absence of a single dominant measure of clinical effectiveness.⁹¹ The effectiveness of interventions in asthma and COPD is usually expressed in a variety of commonly used clinical measures such as forced expiratory volume in one second (FEV_1), peak expiratory flow rate (PEFR), the provocative dose of histamine causing a 20% fall in FEV_1 (PC_{20}), symptom scores, use of rescue medication, sleep disturbances etc. Data on side effects and adverse events such as asthma attacks and exacerbations are reported separately. It is possible to use these clinical outcome measures in a cost-effectiveness analysis, but it is known from the literature and the results of our studies presented in chapter 5 and 6 that these measures are relatively poor proxies for overall quality of life.⁹²⁻⁹⁶ Sculpher suggested the construction of a composite measure of effectiveness, called the episode-free day, which combines various clinical outcomes.⁷² An episode can be either an asthma attack, an adverse event, the need for rescue medication, or sleep disturbance caused by asthma. This measure somewhat resembles the symptom-free days that we have used as one of the outcome measures in the two cost-effectiveness studies of combined inhaled bronchodilator and corticosteroid therapy in chapter 3 and 4. However, even a measure such as the episode-free days is only a pragmatic solution because it is still entirely based on clinical outcome measures. Thus it is not a good substitute for quality of life measures.

For patients with asthma or COPD, loss of quality of life is likely to be perceived as the major consequence of their illness. Fortunately instruments for measuring quality of life, specific to asthma and COPD, have recently been developed,⁹⁷⁻¹⁰³ even for children.^{104,105} These measures provide a very useful description of the quality of life aspects that are characteristic of the disease. And, as we also found when comparing the sensitivity to change of four different quality of life questionnaires in chapter 6, these disease-specific questionnaires are usually more sensitive to changes in health status than generic instruments such as health profiles. However, because these measures focus only on those aspects of quality of life that are relevant to the disease, they do not provide the overall picture of quality of life that is needed to compare the effects of interventions across different patient populations. If such an overall picture is required, or when the changes in

Different aspects of quality of life do not point in the same direction, generic quality of life instruments such as health profiles or utility measures are needed.

Health profiles tend to cover the complete concept of health by measuring various dimensions of quality of life. These aspects may be combined into a few sub-dimension scores or into one overall summary score (in that case the instrument is called a health index). The generic instruments that generate a single index score may be suitable for determining cost-effectiveness ratios. Furthermore, generic instruments, because they are more broadly focused, may detect unanticipated effects on the quality of life from new therapies, effects that are not detected by a disease-specific instrument. Health profiles that are available in Dutch are the Sickness Impact Profile (SIP),^{106,107} the Nottingham Health Profile (NHP),^{108,109} the Medical Outcomes Study 20-item questionnaire (MOS-20),^{110,111} the Short Form 36 (SF-36),^{110,112} and the COOP/WONCA charts.^{113,114} In this thesis we describe our experience in using the NHP and the SIP. The NHP was used in the study comparing continuous with symptomatic bronchodilator therapy (chapter 5) and the SIP was used in our study of salmeterol versus salbutamol (chapter 6). In both studies these generic instruments showed the quality of life of the patient population to be impaired when compared to the general population, but they also proved to be relatively unresponsive to changes in health status that were clearly recorded by the clinical outcome measures.

1.7.2 Utility measurement

As indicated above, it is not easy to find a single outcome measure that represents the multi-dimensional changes a 'successful' therapy can bring about. In order to relate such a multidimensional health outcome to the costs of achieving it, a system is required that weights the changes in the various dimensions of health so that a single overall index value can be assigned to the package of changes.¹¹⁵ This can be achieved by means of utility measurement. In contrast to the disease-specific and generic quality of life instruments mentioned above, which primarily aim at *describing* the quality of life, a utility measurement goes one step beyond that by explicitly *valuing* the quality of life. A utility is a single number on a scale ranging from perfect health, which has a utility of 1, to death, which has a utility of zero. This number represents the value or preference weight assigned to a particular health state relative to perfect health and death.¹¹⁶

Utilities can be obtained from the general public, health care providers or patients.⁸⁵ In our studies we have chosen to obtain the utilities directly from patients, since patient utilities reflect the relative value of different health states to the people who should benefit from the services provided by the health care system.¹¹⁷ However it can also be argued, that since the health care sector is primarily publicly financed and pertains to public policy decisions, the utilities should be obtained from the general public.⁸⁵

The most commonly used methods to elicit utilities are the rating scale, the standard gamble and the time trade-off method, but other methods such as equivalence¹¹⁸ and magnitude estimation¹¹⁹ exist. A comprehensive description of the major techniques and their accuracy is given by Torrance.¹²⁰ Of these techniques, only the standard gamble is well-founded on an economic theory, the Von Neumann and Morgenstern expected utility theory.¹²¹ In 1989, Froberg and Kane published an extensive review of strategies and methods to measure utilities, as well as of context variables and population characteristics that influence the utilities.¹²²⁻¹²⁴ High on the research agenda which they also published were more studies on the reliability and validity of the preference values produced by different methods of utility measurement.¹²⁵ Our studies, described in chapter 6 and 7, contribute to increased knowledge on these aspects. In 1991 Gerard reviewed 51 cost-utility studies published between 1980 and 1990.^{126,127} She concluded that the technical execution of the limited number of utility measurement studies was often of poor quality. She also found that the majority of the studies did not measure utilities directly, but rather used the utilities that were obtained by the original developers of various utility measurement instruments. Currently, the most commonly-used utility measurement instruments are Kaplan's Quality of Well-being scale,¹²⁸ Torrance's Health Utility Index,¹²⁹ Rosser's Disability Distress scale,¹³⁰ and the EuroQol.¹³¹ In our studies we have not used the utilities available from the original developers, but rather conducted utility measurements by means of the Maastricht Utility Measurement Questionnaire, a translated and slightly adapted version of the McMaster Health Utility Index,^{132,133} which in turn was based on Torrance's Health Utility Index. An introduction to this instrument, which contains both the rating scale and the standard gamble method, and a description of the feasibility of these techniques can be found in Bakker *et al.*¹³⁴

Utility measurements are useful in studies of asthma and COPD care, because this care is primarily directed towards the prevention or reduction of morbidity rather than mortality. However, some elements of this care, for example the treatment of asthma attacks or oxygen therapy in severe COPD or emphysema, also have an effect on survival. If quality of life changes and mortality effects have to be taken into account together, the construction of a measure incorporating both the quality and the length of life is needed. The Quality Adjusted Life Year (QALY) is the most familiar of such measures.¹³⁵ In calculating QALYs, the remaining life years are weighted using a quality index for the patients' health state during these years. These weights can be elicited by performing utility measurements. Utilities are not the same as QALYs, but are used as weights to adjust life years for the quality of life in order to calculate QALYs.¹³⁶ Theoretically, the QALY allows comparisons to be made between interventions that mainly effect quality of life, interventions that effect survival and interventions that effect both quality of life and survival.

However, the QALY approach has been widely criticized.¹³⁷⁻¹⁴² Recently an alternative for the QALY, the Healthy Year Equivalent (HYE), has been

proposed.^{117,143-145} The HYE deals with one of the major problems underlying the QALY, i.e. the assumed independence between life years and quality of life. This assumption allows the direct multiplication of life years by utilities in the QALY approach. However it is likely that the utility that a patient assigns to his or her health state will be determined not only by that particular health state but also by the number of years this health state is expected to last. However, doubts have been raised as to whether the HYE is fundamentally different to the time trade-off method.¹⁴⁶ This issue of HYE versus QALY is not further addressed in this thesis.

Another major point of criticism is associated with the comparability of utilities and QALYs, which can only be guaranteed when they are based on the same underlying methods of utility measurement and the same underlying dimensions of health. Despite the criticism, the concepts of utilities and QALYs are useful. By means of utility measurement, value judgements which otherwise implicitly guide decisions about the distribution of health care technologies are made explicit. However, more effort should be directed to strengthening the empirical basis of utility measurements. Chapter 6 and chapter 7 of this thesis are intended to contribute to that goal.

1.7.3 Analysing cost data

Until recently, evidence of differences in the economic data for different treatment groups was often not required to be as 'hard' as the clinical evidence.¹⁴⁷ Economic analysis is often restricted to a straight-forward calculation of cost differentials between alternatives. Uncertainty in economic analysis is dealt with by means of sensitivity analysis, in which critical assumptions in the cost analyses are varied in order to test the sensitivity of the results to such changes.⁸⁷ This has undoubtedly been because economic evaluations were often done retrospectively and the economic data were not collected on a patient-specific basis, so that the variance in cost data cannot be calculated. This makes formal statistical testing of differences in cost data impossible.

However at the moment there is a strong tendency for the data on costs and effects required for an economic evaluation to be gathered in a randomized controlled trial. Although a randomized controlled trial may not always reflect average clinical practice, it is the best design to minimize various types of bias in the estimation of the economic data. As a result it is increasingly being accepted that, when testing differences in cost data, the statistical procedures used should be similar to the ones used in testing outcome data. Unfortunately this is not so straightforward as it seems.¹⁴⁸ One complicating factor is that cost data are usually heavily skewed to the right. A relatively small proportion of the patients in a trial are responsible for a large proportion of the costs. This skewness is often caused by a small number of hospitalizations or other high-cost events. As a result, parametric

procedures which require an underlying normal distribution are inapplicable. Moreover, the sample sizes of clinical trials are usually calculated to be sufficient to identify important differences in one or more clinical outcomes, rather than the economic outcomes. Due to the skewness, this sample size may be too small to test whether economic differences are due to chance.

O'Brien *et al.* have recently proposed a method for the formal statistical testing of economic hypotheses, in which they concentrate on the calculation of confidence intervals for cost-effectiveness ratios.¹⁴⁹ It remains relevant however, to also know the size and significance of absolute differences in costs, separately from the effects. An appropriate procedure for estimating and statistically testing differences in longitudinally-gathered sets of cost data which are heavily skewed is given in chapter 8. This approach is illustrated using the cost data gathered alongside the SGO adult trial of inhaled corticosteroids.

1.8 Aims and outline of this thesis

In this thesis various aspects of cost-effectiveness in asthma and COPD are dealt with. Part of the thesis deals with methodological aspects of measuring and analysing quality of life, utilities and costs, but it also presents the results of the empirical implementation of the proposed methods. The collection of studies in this thesis aim to answer the following questions:

1. What is known about the cost-effectiveness of asthma and COPD care and what gaps in this knowledge can be identified?
2. What is the cost-effectiveness of maintenance pharmacotherapy directed at both increasing airway calibre and suppressing airway inflammation (with inhaled β_2 -agonist and inhaled corticosteroid), as compared to maintenance therapy directed at increasing airway calibre alone (inhaled β_2 -agonist only) in asthmatic children?
3. What is the cost-effectiveness of maintenance pharmacotherapy directed at both increasing airway calibre and suppressing airway inflammation (with inhaled β_2 -agonist and inhaled corticosteroid), as compared to maintenance therapy directed at increasing airway calibre alone (inhaled β_2 -agonist only), in adults with obstructive airways disease?
4. What is the cost-effectiveness of maintenance pharmacotherapy directed at maximal bronchodilation (with inhaled β_2 -agonist and inhaled anticholinergic), as compared to maintenance therapy with only one bronchodilator (inhaled β_2 -agonist), in adults with obstructive airways disease?
5. Is maintenance bronchodilator therapy more effective in improving the quality of life of asthma and COPD patients than symptomatic bronchodilator therapy?
6. Is the new long-acting β_2 -agonist salmeterol more effective in improving the quality of life of asthma patients than the short-acting β_2 -agonist salbutamol?

7. What are the psychometric properties (validity, reliability and sensitivity to change) of 2 asthma-specific quality of life questionnaires (developed by Hyland and Juniper), the Sickness Impact Profile and patient utilities measured by means of the rating scale and the standard gamble method, in asthmatic patients given outpatient pharmacotherapy?
8. How do the magnitude and reliability of patient utilities elicited by means of the rating scale compare to utilities elicited by means of the standard gamble?
9. What is the internal consistency of the standard gamble method?
10. How can costs be calculated when economic data contain missing elements due to highly selectively treatment-related drop-out?
11. What is the most appropriate approach for analysing, in the context of economic appraisal alongside a clinical trial, highly-skewed cost distributions?

In this thesis we first review the literature available on the cost-effectiveness of various health care interventions in asthma and COPD (Chapter 2). The answer to question 2 is given in the third chapter, which describes the results of a study assessing the additional costs and effects (in terms of medical consumption and clinical outcomes) of combined inhaled bronchodilator and corticosteroid therapy compared to β_2 -agonist monotherapy in asthmatic children. This cost-effectiveness study was linked to a randomized controlled multicentre clinical trial in children (SGO-children trial of the Dutch Chronic Non-Specific Lung Disease Study Group). The answers to questions 3 and 4 are given in chapter 4, describing the results of a cost-effectiveness study linked to a clinical trial in which combined inhaled bronchodilator and corticosteroid therapy, combined inhaled β_2 -agonist and anticholinergic therapy and β_2 -agonist monotherapy are compared (SGO-adults trial of the Dutch Chronic Non-Specific Lung Disease Study Group). Chapter 5 addresses the quality of life effects of continuous versus symptomatic bronchodilator therapy. Chapter 6 deals with the effects on quality of life of inhaled salmeterol as compared to salbutamol. The answer to question 7 is also given in chapter 6. A comparison between the rating scale and the standard gamble method can be found in parts of chapter 6 as well as in chapter 7, which addresses various methodological issues associated with patient utility measurement. Chapter 7 also deals with the internal consistency of the standard gamble. This chapter is the only part of this thesis that involved rheumatic patients rather than respiratory patients. However, this chapter focuses on methodological issues of utility measurement techniques, issues which are not limited to a particular patient-population. Moreover, our experience with the utility measurements described in chapter 7 determined the methods used to measure the utilities in chapter 6. Chapter 8 outlines an approach which can be used to deal with a heavily skewed cost-distribution that contains missing data due to selective treatment-related drop-out, thereby answering questions 10 and 11. Finally, a general discussion and recommendations for further research are given in chapter 9.

1.9 References

1. National Heart Lung and Blood Institute. International consensus report on diagnosis and treatment of asthma. *European Respiratory Journal* 1992; 5: 601-641
2. American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *American Review Respiratory Disease* 1987; 136: 225-244.
3. Stuurgroep Toekomstscenario's Gezondheidszorg. Chronische Ziekten in het jaar 2005. Deel 2. Scenario's over CARA 1990-2005. Bohn Stafleu van Loghum, Houten/Antwerpen, 1990
4. Bottema BJAM, Van den Hoogen HJM, Nijhoff S, Schadé E, Van Weel C. CARA en acute bronchitis in de huisartspraktijk, 1971-1989. *Huisarts en Wetenschap* 1992; 35: 305-310
5. Ruwaard D, Gijsen R, Verkley H. Chronische Aspecifieke Respiratoire Aandoeningen (CARA). In: Ruwaard D, Kramers PGN (eds.) *Volksgezondheid Toekomst Verkenning. De gezondheidstoestand van de Nederlandse bevolking in de periode 1950-2010*. Sdu Uitgeverij, Den Haag, 1993
6. Huygen FJA, Van Eijk J, Van den Hoogen H, Van Heeswijk A, Van Iersel J, Roozen R, Sechterberger O, Vlaar N, Van der Wert C. Een praktijk doorgelicht op CARA (1). *Huisarts en Wetenschap* 1977; 20: 383-386
7. Speight ANP, Lee DA, Hey EN. Underdiagnosis and undertreatment of asthma in childhood. *British Medical Journal* 1983; 286: 1253-1256
8. Kaptein AA, Dekker FW, Gill K, Van der Waart MAC. Undertreatment of asthma in Dutch general practice. *Family Practice* 1987; 4: 219-225
9. Levy M. Delay in diagnosing asthma- is the nature of general practice to blaim. *Journal of the Royal College of General Practitioners* 1986; 36: 52-53
10. Dompeling E, Van Grunsven PM, Molema J, Verbeek ALM, Van Schayck CP, Van Weel C. Early detection of patients with fast progressive asthma or chronic bronchitis in general practice. *Scan J Prim Health Care* 1992; 10: 143-150
11. Dekker FW. Asthma and COPD in general practice. Studies on the quality of care. Dissertation. University of Leiden, 1993
12. Koopmanschap MA, Van Roijen L, Bonneux L. Kosten van ziekten in Nederland. Instituut Maatschappelijke Gezondheidszorg, Instituut voor Medische Technology Assessment, Erasmus Universiteit Rotterdam, 1991
13. Weel C van, Van den Bosch WJHM, Van den Hoogen HJM. De continue morbiditeitsregistratie Nijmegen. Een gegevensbestand voor longitudinaal patiëntgebonden onderzoek in de huisartspraktijk. *Huisarts en Wetenschap* 1986; 29: 373-377
14. Gregg I. The importance of asthma to the general practitioner. *The Practitioner* 1987; 231: 471-477
15. Swinkels H. Contact met de huisarts naar praktijkvorm, 1981-1985. Maandbericht Gezondheidsstatistiek CBS. Volume 6, no. 5, p. 5-18, Staatsuitgeverij, s' Gravenhage, 1987
16. Velden van der J. De rol van de huisartspraktijk in Nederland; Nationale studie van ziekten en verrichtingen in de huisartspraktijk. *Medisch Contact* 1990; 45: 605-608
17. Wever-Hess J, Wever AMJ, Yntema JL. Mortality and morbidity from respiratory diseases in childhood in the Netherlands, 1980-1987. *European Respiratory Journal* 1991; 4: 429-433
18. Strauss MH, Conrad D, LoGerfo JP, Hudson LD, Bergner M. Costs and outcome for patients with chronic obstructive lung disease. *Medical Care* 1986; 24: 915-924

19. Van Mólken MPMH, Van Doorslaer EKA, Rutten FFH, CARA in Cijfers. Rijksuniversiteit Limburg, Institute for Medical Technology Assessment, Maastricht, 1989
20. Velde LJK van der, Gijsen R, Ruwaard D, Verkleij H, Casparie AF. Economische aspecten van CARA. Appendix 2. In: Stuurgroep Toekomstscenario's Gezondheidszorg. Chronische Ziekten in het jaar 2005. Deel 2. Scenario's over CARA 1990-2005. Bohn Stafleu van Loghum, Houten/Antwerpen, 1990
21. Weiss KB, Gergen PJ, Hodgson A. An economic evaluation of asthma in the United States. *New England Journal of Medicine* 1992; 326: 862-866.
22. Mellis CM, Jennifer KP, Woolcock AJ. The Cost of Asthma. Can it be Reduced? *PharmacoEconomics* 1993; 3: 205-219.
23. Weiss KB, Dullivan SD. The Economic Costs of Asthma. A review and conceptual model. *PharmacoEconomics* 1993; 4: 14-30
24. Barnes PJ. Drug therapy. A new approach to the treatment of asthma. *N Eng J Med* 1989; 321: 1517-1527
25. Commissie criteria geneesmiddelenkeuze (Commissie Van Winzum). Verdeling door Verdunning. Stichting Gezondheidszorg en Publicaties, Geldermalsen, 1994
26. Crane J, Glatt A, Jackson R, Ball M, Pearce N, Burgess C, Kwong T, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981-1983: case control study. *Lancet* 1989; 8644: 917-922.
27. Pearce N, Crane J, Burgess C, Jackson R, Beasley R. Review: beta agonists and asthma mortality: deja vue. *Clinical and Experimental Allergy* 1991; 21: 401-410
28. Spitzer WO, Suissa S, Ernst P, Horwitz RI, Habbick B, Cockcroft D, Boivin J-F, McNutt M, Buist AS, Rebuck AS. The use of β -agonist and the risk of death and near death from asthma. *The New England Journal of Medicine* 1992; 326: 501-506.
29. Sears MR, Taylor DR, Print G, Lake DC, Qingqing L, Flannery EM, Yates DM, Lucas MK, Herbison GP. Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* 1990; 336: 1391-1396.
30. Schayck Van CP, Dompeling E, Herwaarden Van CLA, Folgering H, Verbeek ALM, Hoogen Van der HJM, Weel Van C. Bronchodilator treatment in moderate asthma or chronic bronchitis: continuous or on demand? *British Medical Journal* 1991; 303: 1426-1431.
31. Schayck Van CP, Folgering HTM, Harbers HJM, Maas KL, Van Weel C. Effects of allergy and age on responses to salbutamol and ipratropium bromide in moderate asthma and chronic bronchitis. *Thorax* 1991; 46: 355-359
32. Braun SR, McKenzie WN, Copeland W, Kingman L, Ellersieck M, A comparison of the effect of ipratropium bromide and albuterol in the treatment of chronic obstructive airway disease. *Archives of Internal Medicine* 1989; 149: 544-547
33. Ullman A, Svedmyr N. Inhaled salmeterol - a new β_2 -adrenoceptor agonist produces sustained bronchodilation in asthmatic patients without causing tachyphylaxis. (Abstract) *American Review of Respiratory Disease* 1988; 137: 32
34. Wallin A, Melander B, Rosenhall L, Sandstrom T, Wahlander L. Formoterol, a new long acting beta2 agonist for inhalation twice daily, compared with salbutamol in the treatment of asthma. *Thorax* 1990; 45: 259-261
35. Ullman A, Hedner J, Svedmyr N. Inhaled Salmeterol and Salbutamol in Asthmatic Patients. An evaluation of asthma symptoms and the possible development of tachyphylaxis. *American Review of Respiratory Disease* 1990; 142: 571-575

36. Pearlman DS, Chervinsky P, LaForce G, Seltzer JM, Southern DL, Kemp JP, Dockhorn RJ, Grossman J, Liddle RF, Yancey SW, Cocchetto DM, Alexander WJ, Van As A. A comparison of salmeterol with albuterol in the treatment of mild-to-moderate asthma. *The New England Journal of Medicine* 1992; 327: 1420-1425
37. Kesten S, Chapman KR, Broder I, Cartier A, Hyland RH *et al.* A three-months comparison of twice daily inhaled formoterol versus four times daily albuterol in the management of stable asthma. *American Review of Respiratory Disease* 1991; 144: 622-625
38. Fitzpatrick MF, Mackay T, Driver H, Douglas NJ. Salmeterol in nocturnal asthma: a double blind, placebo controlled trial of a long acting inhaled β_2 agonist. *British Medical Journal* 1990; 301: 1365-1368
39. Cheung D, Timmers MC, Zwinderman AH, Bel EH, Dijkman JH, Sterk PJ. Long-term effects of a long-acting β_2 -adrenoceptor agonist, salmeterol, on airway hyperresponsiveness in patients with mild asthma. *The New England Journal of Medicine* 1992; 327: 1198-1203
40. Booth H, Fishwick K, Harkawat R, Devereux G, Hendrick DJ, Walters EH. Changes in methacholine induced bronchoconstriction with the long acting β_2 agonist salmeterol in mild to moderate asthmatic patients. *Thorax* 1993; 48: 1121-1124
41. Molema J, Herwaarden Van CLA, Folgering HThM. Effects of long-term treatment with inhaled cromoglycate and budesonide on bronchial responsiveness in patients with allergic asthma. *European Respiratory Journal* 1989; 2: 308-316.
42. Barnes PJ. Effect of corticosteroids on airway hyperresponsiveness. *American Review Respiratory Disease* 1990; 141: S70-S76.
43. Bel EH, Timmers MC, Hermans J, Dijkman JH, Sterk PJ. The long-term effects of nedocromil sodium and beclomethasone dipropionate on bronchial responsiveness to methacholine in nonatopic asthmatic subjects. *American Review Respiratory Disease* 1990; 141: 21-28.
44. Juniper EF, Kline PA, Vanzielegheem MA, Ramsdale EH, O'Byrne PM, Hargreave FE. Effect of long-term treatment with an inhaled corticosteroid (budesonide) on airway hyperresponsiveness and clinical asthma in nonsteroid-dependent asthmatics. *American Review Respiratory Disease* 1990; 142: 832-836.
45. Juniper EF, Kline PA, Vanzielegheem MA, Ramsdale EH, O'Byrne PM, Hargreave FE. Long-term effects of budesonide on airway responsiveness and clinical asthma severity in inhaled steroid-dependent asthmatics. *European Respiratory Journal* 1990; 3: 1122-1127.
46. Vathenen AS, Knox AJ, Wisniewski A, Tattersfield AE. Time course of change in bronchial reactivity with an inhaled corticosteroid in asthma. *American Review Respiratory Disease* 1991; 143: 1317-1321.
47. Haahtela T, Jarvinen M, Kava T, Kiviranta K, Koskinen S, Lehtonen K, Nikander K, Persson T, Reinikainen K, Selroos O, Sovijardi A, Stenius-Aarniala B, Svahn T, Tammivaara R, Laitinen L. Comparison of a beta2-agonist, terbutaline, with an inhaled corticosteroid, budesonide, in newly detected asthma. *The New England Journal of Medicine* 1991; 325: 388-392.
48. Eliasson R, Hoffman J, Trueb D, Frederick D, McCornick JR. Corticosteroids in COPD. A clinical trial and reassessment of the literature. *Chest* 1986; 89: 484-490.
49. Postma DS, Peters I, Steenhuis EJ, Sluiter HJ. Moderately severe chronic airflow obstruction. Can corticosteroids slow down progression? *European Respiratory Journal* 1988; 1: 22-26.
50. Weir DC, Grove RI, Robertson AS, Burge PS. Corticosteroid trials in non-asthmatic chronic airflow obstruction: a comparison of oral prednisolone and inhaled beclomethasone dipropionate. *Thorax* 1990; 45: 112-117.

51. Auffarth B, Postma DS, Monchy De JGR, Mark Van der ThW, Boorsma M, Koëter GH. Effects of inhaled budesonide on spirometric values, reversibility, airway responsiveness and cough threshold in smokers with chronic obstructive lung disease. *Thorax* 1991; 46: 372-377.
52. Dompeling E, Van Schayck CP, Molema J, Folgering H, Van Grunsven PM, Van Weel C. Inhaled beclomethasone improves the course of asthma and COPD. *European Respiratory Journal* 1992; 5: 945-952
53. Dompeling E, Van Schayck CP, Van Grunsven PM, Van Herwaarden CLA, Akkermans R, Molema J, Folgering H, Van Weel C. Slowing the deterioration of asthma and COPD during bronchodilator therapy by adding inhaled corticosteroids. A four-year prospective study. *Annals of Internal Medicine* 1993; 118: 770-778
54. Kerstjens HAM, Brand PLP, Hughes MD, Robinson NJ, Postma DS, Sluiter HJ, Bleecker ER, Dekhuijzen R, de Jong PM, Mengelers HJJ, Overbeek SE, Schoonbrood DFME, and the Dutch CNSLD Study Group. A comparison of bronchodilator therapy with or without inhaled corticosteroid therapy in obstructive airways disease. *New England Journal of Medicine* 1992; 327: 1413-1419.
55. Selroos O. The effects of inhaled corticosteroids on the natural history of obstructive lung diseases. *European Respiratory Review* 1991; 1: 354-365
56. Kerrebijn KF, Van Essen-Zandvliet EEM, Neijens HJ. Effects of long-term treatment with inhaled corticosteroids and beta-agonists on the bronchial responsiveness in asthmatic children. *Journal of Allergy and Clinical Immunology* 1987; 79: 653-659
57. Bennati D, Piacentini GL, Peroni DG, Sette L, Testi R, Boner AL. Changes in bronchial reactivity in asthmatic children after treatment with beclomethasone alone or in association with salbutamol. *Journal of Asthma* 1989; 26: 359-364
58. Waalkens HJ, Gerritsen J, Koëter GH, Krouwels FH, Van Aalderen WMC, Knol K. Budesonide and terbutaline or terbutaline alone in children with mild asthma: effects on bronchial hyperresponsiveness and diurnal variation in peak flow. *Thorax* 1991; 46: 499-503
59. National Heart Lung and Blood Institute expert panel. Guidelines for the diagnosis and management of asthma. *Journal of Allergy and Clinical Immunology* 88 part 2 (suppl), 1991
60. British Thoracic Society. Guidelines for management of asthma in adults. I. Chronic persistent asthma. *British Medical Journal* 1990; 301: 651-653
61. British Thoracic Society. Guidelines for management of asthma in adults. II. Acute severe asthma. *British Medical Journal* 1990; 301: 797-800
62. Warner JO, Neijens HJ, Landau LI, Jones K, Asher MI *et al.* Asthma: a follow up statement of an international paediatric asthma consensus group. *Archives of Disease in Childhood* 1992; 67: 240-248
63. Bottema BJAM, Fabels EJ, Van Grunsven PM *et al.* NHP-Standaard CARA bij volwassenen: diagnostiek. *Huisarts en Wetenschap* 1992; 35: 430-436
64. Waart MAC van der, Dekker FW, Nijhoff S, *et al.* NHG-Standaard CARA bij volwassenen: beleid. *Huisarts en Wetenschap* 1992; 35: 437-443
65. Dirksen WJ, Geyer RMM, De Haan M, Kolnaar BGM, Merckx JAM, Romeijnders ACM, Dijkstra RH, Van der Laan JR. NHG-standaard astma bij kinderen. *Huisarts & Wetenschap* 1992; 35: 355-362
66. Steen van der JJ, Gill K. Astma en patient-compliance. *Airways* 1988; 7: 18-20
67. Dieleman FE, Dekker FW, Kaptein AA. Compliantie bij astma-medicatie. *Huisarts en Wetenschap* 1989; 32: 43-47

68. Stuurgroep Toekomstscenario's Gezondheidszorg. De toekomst van het geneesmiddel in de gezondheidszorg. Een scenario-analyse. Hoofdstuk 4, paragraaf 6, Chronische Aspecifieke Respiratoire Aandoeningen, pag. 131-137, Bohn Stafleu, Van Loghum, Houten/Zavetem, 1993
69. Anonymous. Scrip 1990; 1574: 16-18
70. Drummond M. Cost-of-illness studies. A major headache? *Pharmacoeconomics* 1992; 2: 1-4
71. Davey PJ, Leeder SR. The cost of migraine. More than just a headache? *Pharmacoeconomics* 1992; 2: 5-7
72. Sculpher MJ, Buxton MJ. The episode-free day as a composite measure of effectiveness. An illustrative economic evaluation of formoterol versus salbutamol in asthma therapy. *Pharmacoeconomics* 1993; 4: 345-352
73. Sondergaard B, Davidsen F, Kirkeby B, Rasmussen M, Hey H. The economics of an intensive education programme for asthmatic patients. A prospective controlled trial. *Pharmacoeconomics* 1992; 1: 207-212
74. Tougaard L, Krone T, Sorknaes A, Ellegaard H, and the PASTMA Group. Economic benefits of teaching patients with chronic pulmonary disease about their illness. *The Lancet* 1992; 339: 1517-1520
75. Trautner C, Richter B, Berger M. Cost-effectiveness of a structured treatment and teaching programme on asthma. *European Respiratory Journal* 1993; 6: 1485-1491
76. Ross RN, Morris M, Stanley R, Sakowitz SR, Berman BA. Cost-effectiveness of including cromolyn sodium in the treatment program for asthma: a retrospective record-based study. *Clinical Therapeutics* 1988; 10: 188-203
77. Tierce JC, Phil C, Meller W, Berlow B, Gerth WC. Assessing the cost of albuterol inhalers in the Michigan and California Medicaid programs: A total cost-of-care approach. *Clinical Therapeutics* 1989; 11: 53-61
78. Choices in Health Care. A report by the government committee (committee Dunning) on choices in health care. Ministry of Welfare, Health and Cultural Affairs, Zoetermeer, The Netherlands, 1992
79. Drummond M, Rutten F, Brenna A, Pinto CG, Horisberger B, Jönsson B, Le Pen C, Rovira J, Graf von der Schulenburg M, Sintonen H, Torfs K. Economic Evaluation of Pharmaceuticals. A European Perspective. *Pharmacoeconomics* 1993; 4: 173-186
80. Redwood H. The Dynamics of Drug Pricing and Reimbursement in the European Community. Where are control systems leading suppliers, prescribers and patients? Assessment and Prognosis. Oldwicks Press Limited, Felixtowe, Suffolk, England, 1992
81. State Secretary of Welfare, Health and Cultural Affairs. Simons HJ. Letter on the cost-containment of pharmaceuticals. Rijswijk, 22 april 1993
82. Ziekenfondsraad. Besluit farmaceutische hulp AWBZ: bijlage 4 (niet WTG geneesmiddelen). Rapportnr. 581, Amstelveen, 1993
83. Henry D. Economic Analysis as an Aid to Subsidization Decisions. The Development of Australian Guidelines for Pharmaceuticals. *Pharmacoeconomics* 1992; 1: 54-67
84. Detsky A. Guidelines for economic analysis of pharmaceutical products: a draft document for Ontario and Canada. *Pharmacoeconomics* 1993; 3: 354-361
85. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford, 1987
86. Drummond M. Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status. *Health Economics* 1992; 1: 85-92

87. Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation of health care. Practice, Problems and Potential. *International Journal for Technology Assessment in Health Care* 1993; 9: 26-36
88. Rutten FFH, Van Ineveld BM, Van Ommen R, Van Hout BA, Huijsman R. Kostenberekening bij Gezondheidszorgonderzoek. Richtlijnen voor de praktijk. Rapport in opdracht van de Stuurgroep Toekomstscenari'o's Gezondheidszorg, Uitgeverij Jan van Arkel, Utrecht, 1994
89. Essink-Bot M-L. on behalf of the Dutch Working Group on Health Status Assessment. Key issues for standardization in health status assessment. *Quality of Life News Letter* 1994; February-May: 1-2
90. Essink-Bot M-L, Bonsel GJ. Standaardisatie van meetinstrumentarium voor gezondheidstoestandmeting. *Huisarts en Wetenschap* 1994 (forthcoming)
91. Buxton MJ. Economic evaluation studies in respiratory medicine. *Respiratory Medicine* 1991; 85: Sb43-Sb46
92. McSweeney AJ, Heaton, RK, Grant I, Cugell D, Solliday N, Timms R. Chronic obstructive pulmonary disease; socioemotional adjustment and life quality. *Chest* 1980; 77: 309-311
93. Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773-778
94. Williams SJ, Bury MR. Impairment, disability and handicap in chronic respiratory illness. *Social Science and Medicine* 1989; 29: 609-616
95. Schrier AC, Dekker F, Kaptein AA, Dijkman JH. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990; 89: 894-899
96. Malo J-L, Boulet L-P, Dewitte J-D, Cartier A, L'Archeveque J, Coté J, Bedard G, Boucher S, Champagne F, Tessier G, Contandriopoulos A-P, Juniper EF, Guyatt GH. Quality of life of subjects with occupational asthma. *Journal of Allergy and Clinical Immunology* 1993; 91: 1121-1127
97. Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773-778
98. Hyland ME. The Living with Asthma Questionnaire. *Respiratory Medicine* 1991; 85: S13-S16
99. Marks GB, Dunn SM, Woolcock AJ. A scale for the measurement of quality of life in adults with asthma. *Journal of Clinical Epidemiology* 1992; 45: 461-472
100. Creer TL, Wigal JK, Kotses H, McConnaughy K, Winder JA. A life Activities Questionnaire for Adult Asthma. *Journal of Asthma* 1992; 29: 393-399
101. Jones PW, Quirk FH, Baveystock CM. The St. George's Respiratory Questionnaire. *Respiratory Medicine* 1991; 85: S25-S31
102. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47: 76-83
103. Maillé AR, Koning CJM, Zwinderman AH, Kaptein AA. De ontwikkeling van een kwaliteit van leven vragenlijst voor CARA-patiënten. *Gedrag en Gezondheid* 1994; 22: 35-43
104. Christie MJ, French D, Weatherstone L, West A. The patients' perceptions of chronic disease and its management: psychosomatics, holism and quality of life in contemporary management of childhood asthma. *Psychotherapy and Psychosomatics* 1991; 56: 197-203
105. Creer TL, Wigal JK, Kostas H, Hatala JC. A life activities questionnaire for childhood asthma. Athens, Ohio Universtidy, Department of Psychology, 1992

106. Bergner M, Bobbit RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care* 1981; 19: 787-805
107. Jacobs HM, Luttik A, Touw-Otten FWMM, De Melker RA. De 'Sickness Impact Profile'. Resultaten van een valideringsonderzoek van de Nederlandse versie. *Nederlands Tijdschrift voor Geneeskunde* 1990; 134: 1950-1954
108. Hunt SM, McEwen J, McKenna SP. *Measuring health status*. 1st ed. London, Croom Helm, 1986
109. Erdman RAM, Passchier J, Kooijman M, Stronks DL. The Dutch version of the Nottingham Health Profile: investigation of psychometric aspects. *Psychological Reports* 1992; 72: 1027-1035
110. Stewart AL, Ware JE (eds). *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*, Durham, NC: Duke University Press, 1992
111. Kempen GIJM. Het meten van de gezondheidstoestand van ouderen; een toepassing van de Nederlandse versie van de MOS-schaal. *Tijdschrift voor Gerontologie en Geriatrie* 1992; 23: 132-140
112. Van der Zee KI, Sanderman R, Heyink J. De psychometrische kwaliteiten van de MOS 36-item Short Form Health Survey (SF-36) in een Nederlandse populatie. *Tijdschrift voor Sociale Geneeskunde* 1993; 71: 183-191
113. Nelson EC, Wasson J, Kirk J *et al.* Assessment of function in routine clinical practice. Description of the COOP chart method and preliminary findings. *Journal of Chronic Diseases* 1987; 40: 55S-64S
114. Scholten JHG, Van Weel C. *Functional Status Assessment in Family Practice, The Dartmouth COOP Functional Health Assessment Charts/WONCA*. World Organization of Family Doctors. MediTekst, Lelystad, 1992
115. Feeny D, Labelle R, Torrance GW. Integrating Economic Evaluations and Quality of Life Assessments. In: Spilker B (ed.). *Quality of Life Assessment in Clinical Trials*. Ltd. New York, Raven Press, 1990.
116. Torrance GW. Utility approach to measuring health-related quality of life. *Journal of Chronic Disease* 1987; 40: 593-600
117. Mehrez A, Gafni A. Quality-adjusted life years, utility theory and health-years equivalents. *Medical Decision Making* 1989; 9: 142-149
118. Patrick DL, Bush JW, Chan MM. Methods for measuring levels of well-being for a health status index. *Health Services Research* 1973; 8: 228-245
119. Kaplan RM, Bush JW, Berry CC. Health status index: category ratings versus magnitude estimation for measuring levels of well-being. *Medical Care* 1979; 17: 501-525
120. Torrance GW. Measurement of health-state utilities for economic appraisal: A review. *Journal of Health Economics* 1986; 5: 1-30
121. Neumann Von J, Morgenstern O. *Theory of games and economic behavior*. Princeton University Press, 1944 (1st ed.), 1947 (2nd ed.).
122. Froberg DG, Kane RL. Methodology for measuring health-state preferences-I: Measurement Strategies. *Journal of Clinical Epidemiology* 1989; 42: 345-354.
123. Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: Scaling Methods. *Journal of Clinical Epidemiology* 1989; 42: 459-471.
124. Froberg DG, Kane RL. Methodology for measuring health-state preferences-III: Population and Context Effects. *Journal of Clinical Epidemiology* 1989; 42: 585-592.

125. Froberg DG, Kane RL. Methodology for measuring health-state preferences-IV: Progress and a Research Agenda. *Journal of Clinical Epidemiology* 1989; 42: 675-685.
126. Gerard K. A review of cost utility studies: assessing their policy-making relevance. Discussion paper 11/91. Health Economics Research Unit, University of Aberdeen, 1991
127. Gerard K. Cost-utility in practice: A policy maker's guide to the state of the art. *Health Policy* 1992; 21: 249
128. Kaplan RM, Bush JW. Health status: types of validity and the index of well-being. *Health Services Research* 1976; Winter: 478-507.
129. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Research* 1982; 30: 1043-1069.
130. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? *International Journal Epidemiology* 1978; 7: 347-358.
131. The EuroQoL Group. A new facility for the measurement of health related quality of life. *Health Policy* 1990; 16: 199-208
132. Bennett K., Torrance GW, Tugwell P. Methodologic challenges in the development of utility measures of health-related quality of life in rheumatoid arthritis. *Controlled Clinical Trials* 1991; 12: S118-S128
133. Bakker CH, Rutten-van Mólken M, Van Doorslaer E, Bennett K, Van der Linden Sj. Health Related Utility Measurement in Rheumatology: An Introduction. *Patient Education and Counseling* 1993; 20: 145-152
134. Bakker C, Rutten M, Van Doorslaer E, Bennett K, Van der Linden Sj. Feasibility of Utility Assessment by Rating Scale and Standard Gamble in Patients with Ankylosing Spondylitis and Fibromyalgia. *Journal of Rheumatology* 1994; 21: 269-274
135. Weinstein MC, Stasson WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England Journal of Medicine* 1977; 296: 716-721.
136. Torrance GW, Feeny D. Utilities and quality adjusted life years. *International Journal of Technology Assessment in Health Care* 1989; 5: 559-575
137. Smith A. Qualms about QALYs *Lancet* 1987; 1(8542): 1134-1136
138. Gafni A. The quality of QALY's (quality adjusted life years): do QALY's measure what they at least intend to measure? *Health Policy* 1989; 13: 81-83
139. Loomes G, McKenzie L. The use of QALY's in Health Care Decision Making. *Social Science and Medicine* 1989; 28: 299-308
140. Carr-Hill RA. Allocating resources to health care: is the QALY (Quality Adjusted Life Year) a technical solution to a political problem? *International Journal of Health Services* 1991; 21: 351-363
141. Gerard K, Mooney G. QALY league tables: Handle with Care. *Health Economics* 1993; 2: 59-64
142. Nord E. Toward quality assurance in QALY calculations. *International Journal of Technology Assessment in Health Care* 1993; 9: 37-45
143. Mehrez A, Gafni A. The healthy-years-equivalents: How to measure them using the standard gamble approach. *Medical Decision Making* 1991; 11: 140-146
144. Gafni A, Birch S. Searching for a common currency: Critical appraisal of the scientific basis underlying European harmonization of the measurement of Health Related Quality of Life (EuroQoL). *Health Policy* 1993; 23: 219-228
145. Bleichrodt H, Gafni A. HYE's: The second generation of outcomes measures in health. Institute for Medical Technology Assessment, Report number 93.29, Erasmus University Rotterdam, 1993

146. Buckingham K. Risks in utility assessment and risks of medical interventions. Letter to the editor. *Medical Decision Making* 1993; 13: 167-168
147. Buxton MJ. Problems in the economic appraisal of new technology: the evaluation of heart transplants in the UK. In: Drummond MF (ed). *Economic Appraisal of Health Technology in the European Community*. Oxford University Press, 1987
148. Drummond, M., O'Brien, B. Clinical importance, statistical significance and the assessment of economic and quality of life outcomes. *Health Economics*, 1993; 2: 205-212
149. O'Brien, B.J., Drummond, M.F., Labelle, R.J. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 1994; 32: 150-163

2

Economic appraisal of asthma and COPD care: A literature review 1980-1991

*Maureen P.M.H. Rutten-Van Mölken¹, Eddy K.A. Van Doorslaer²,
Frans F.H. Rutten².*

1 Department of Health Economics, University of Limburg.

2 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

Originally published as: Economic appraisal of asthma and COPD care: A literature review 1980-1991. Soc. Sci. Med. 1992; 35:161-175. Reprinted with permission of the publisher.

2.1 Abstract

Despite the considerable burden and costs of illness and despite the increasing need to set priorities on the basis of efficiency considerations, only 20 economic appraisals of asthma and COPD care have been published during the past 11 years. This paper provides a detailed summary of the cost-effectiveness 'evidence' given by these studies and a discussion of relevant methodological issues. The studies comparing programme costs of delivery methods for oxygen and for aerosol bronchodilator drugs, provide the most straightforward evidence in favour of the concentrator and the metered dose inhaler respectively. There also seems to be evidence in favour of hospital-based home care programmes as compared to community-based home care programmes. Health education, especially directed at asthmatic children seems to reduce health care costs and improve attitude, compliance behaviour and self-management skills. Information on the cost-effectiveness of pharmacotherapy and diagnostic technologies, both important interventions in asthma and COPD, was found to be totally lacking.

2.2 Introduction

The pressing need for more efficient allocation of resources in health care has stimulated interest in economic evaluation studies. Whereas the first studies tended to concentrate on the most visible applications of modern advanced diagnostic and therapeutic technology in medicine, the focus is gradually shifting to more routinely applied 'small-ticket' technologies. Although these are less expensive per unit of output, they often lead to much higher costs, because of their wide application in much larger patient populations, who often need long-term care. The treatment of asthma and chronic obstructive pulmonary disease (COPD) is characterized by such long-term, small-ticket, mostly drug therapy.

The terminology of these two diagnoses is often confusing. In clinical practice patients frequently have overlapping characteristics and are not easily categorized as having either asthma or COPD. The Dutch practice is to treat asthma and COPD as different expressions of the same underlying disease called chronic non-specific lung disease (CNSLD). However we will follow the English-language literature and report separately on asthma and COPD. COPD refers primarily to chronic bronchitis and emphysema. It causes progressive loss of lung function and is primarily related to long term cigarette smoking, age, occupational exposure and airways hyper-responsiveness. Asthma is characterized by attacks of dyspnea or wheezing, reversible airflow obstruction and hyper-responsiveness.¹

In the Netherlands, the two conditions impair the quality of life of 10 to 20% of men and 5 to 10% of women and children. The associated burden of illness is considerable. The annual costs to society in the Netherlands are estimated to exceed

1 billion Dutch guilders. This includes health care costs and costs due to absence from work.² In the Netherlands, most asthma and COPD patients are treated by general practitioners (GPs). Almost 9% of all GP consultations and about 13% of all absenteeism due to illness is caused by asthma and COPD.² The ageing of the population will almost certainly lead to even larger costs of illness, especially due to COPD, in the near future. In children of 4 to 12 years of age asthma is the main reason for absence from school.³ Also the number of hospitalizations due to asthma has been shown to be increasing significantly in children aged 0-4.⁴ Although they are generally considered to be of greater importance as a cause of disability and ill health than as a cause of death, asthma and COPD directly or indirectly cause 15% of all male mortality and 6% of all female mortality.

Since the prevalence and associated burden and costs of asthma and COPD are large and are expected to grow in the near future, it seems important to establish the cost-effectiveness of the various health care interventions which are in common use. Economic appraisals, by systematically comparing both the costs and the outcomes of alternative health care programmes, provide information on how to allocate scarce resources to obtain maximum positive health effects from those programmes. This article will review existing knowledge on the cost-effectiveness of asthma and COPD treatment and assess the quality of this information in a critical and constructive way. After describing the literature search, we review the topics of the studies and describe the nature of the economic appraisals. Then we deal with the alternative treatments and what is known about their cost-effectiveness. The next sections review the measurement of outcomes and cost, followed by a thorough assessment of the quality of the effectiveness evidence on which the economic evaluations are based. The last section contains the conclusions and a discussion.

2.3 Selection of the literature

A medline search for publications in English found 19 articles offering economic appraisals of a particular asthma or COPD-related health care intervention.⁵⁻²³ A publication from the U.S. Office of Technology Assessment was added to this.²⁴ A combination of the following entries was used: respiratory tract disease, obstructive lung disease, asthma, respiratory therapy, respiratory drug therapy, economics, costs and cost-analysis, cost-benefit analysis and economic evaluation. The entry "cost-benefit analysis" applied to all of the articles. The studies were all published between January 1980 and December 1991.

2.4 Nature of the economic appraisals

Like Williams,²⁵ who reviewed the psychosocial literature on COPD, we found that the majority of the economic evaluation studies published in international journals have been performed in the United States. Two studies were performed in the UK, 1 in Australia, 1 in Germany and 1 in Sweden.

A wide variety of COPD interventions have been subjected to an economic appraisal (See Table 1). Health educational programmes for adults and for children have been appraised relatively often. These programmes are primarily directed at asthmatics. Economic appraisals of pulmonary rehabilitation -a multidisciplinary programme which among other things includes physical exercise and health education- and of domiciliary oxygen therapy and home care, generally include patients with rather severe chronic airflow limitation. Aerosol bronchodilator delivery methods have been the subject of several studies involving asthma, COPD and other respiratory diseases. We have not found a single economic appraisal of pharmacotherapy, which is the main treatment in asthma and COPD.

Before we discuss the studies themselves, it is necessary to consider what constitutes an economic evaluation. Drummond *et al.*²⁶ make a useful distinction between partial evaluation and full economic evaluation. Whereas partial evaluation studies are restricted to either costs or outcomes, or consider no alternatives to the intervention being studied, in full economic evaluations some combination of costs is compared with some combination of outcomes for two or more alternatives. Depending on the way in which the outcomes are assessed, a full economic evaluation is labelled a cost-minimization analysis, a cost-effectiveness analysis, a cost-utility analysis or a cost-benefit analysis.^{26,27} All the articles reviewed have been labelled according to Drummond's classification (See last column of table 1).

In a *cost-minimization analysis* the central question is restricted to: "How to obtain the same outcome at the lowest cost?" Equal effectiveness is explicitly assumed, backed up by evidence from good quality studies. For example, the studies by Lowson *et al.*¹⁵ and Scheffler *et al.*²⁴ concerning alternative ways of delivering oxygen and aerosol bronchodilator drugs, explicitly assumed that the alternatives were equally effective in terms of lung function and bronchodilation respectively. So the analyses concentrate on the assessment of the costs only.

In contrast, the studies by McKeon *et al.*¹⁶ and Motwani *et al.*¹⁷, who also restricted their analyses to the costs only, were partial evaluation studies classified as *cost-analyses*. Motwani *et al.*, in comparing home care with institutional care, and McKeon *et al.*, in comparing the current practice of oxygen delivery with a rationalized programme, did not explicitly assume the equal effectiveness of these alternatives. Had they provided evidence for assuming equal effectiveness, their studies would have been labelled cost-minimization analyses.

If the outcome is measured in units such as life years gained, walking distance, prevented reduction in forced expiration volumes or improvement in score on a quality of life scale, then the economic evaluation is called a *cost-effectiveness*

analysis. For example, the study by Schmidt *et al.*¹⁸ on mechanical ventilation for various groups of patients estimated that the costs per life year gained ranged from \$460 for asthma patients to \$2155 for patients suffering from chronic bronchitis and emphysema.

In a *cost-utility analysis*, benefits are expressed in quality adjusted life years (QALYs). A quality of life correction factor called utility is used to adjust life years for their quality. This utility reflects the preference value for a particular health state. Since improving the quality of life might be seen as the ultimate goal of most asthma and COPD therapies, cost-utility analysis would be very appropriate to evaluate them. Although this is often recognised, only Toevs *et al.*²⁰ have attempted to measure this improvement in QALYs or 'well years' as they call them. They evaluated a behavioural programme directed at increasing compliance with a pulmonary rehabilitation programme and concluded that the programme produced a "well year" at a unit cost of \$24,256. Also Hay *et al.*¹² in their study of AAT replacement therapy applied several values for the quality of a restricted activity day, but these values were not actually measured. Since QALYs can be seen as a common unit for measuring the outcome of diverse interventions, cost-utility analysis has the potential additional advantage of allowing a comparison of interventions in asthma and COPD with those in other areas of the health care sector.

Finally, if the outcome is measured in monetary units, the evaluation is labelled a *cost-benefit analysis*. This is often limited to a comparison of items that can easily be expressed in money terms, such as the number of outpatient visits or the number of hospitalizations. Reviewing the literature, it became clear that there is a tendency to describe as cost-benefit analyses studies which concentrate on outcomes merely in terms of savings elsewhere in the health care sector. For example, in the study by Clark *et al.*⁷ a sample of 310 low income urban children was randomly divided into a control group and an experimental group which received health education to improve asthma management at home. The costs of delivering the health education programme were compared to the outcomes in terms of health cost savings mainly consisting of a reduced number of hospitalizations and emergency room visits. Although expecting a positive influence, the effect of health education on the ultimate goal of the intervention, i.e., improvement in quality of life, was not measured. As Guyatt *et al.*²⁷ and Drummond *et al.*²⁸ have shown, a cost-benefit analysis involves taking into account more than just the outcomes easily expressed in monetary units. It involves the monetary measurement of all relevant final outcomes, including health improvements. In fact none of the studies reviewed could be described as a full social cost-benefit analysis.

In addition to Drummond's classification, economic appraisals can be divided into empirical economic evaluation studies and studies based on modelling. In empirical evaluation studies, the economic analysis is linked to a clinical study. Modelling studies are in fact secondary analyses in which cost and outcome data obtained from several previous studies are combined into an overall cost-effectiveness model. The studies by Lowson *et al.*¹⁵ and Hay *et al.*¹² are good examples of the latter.

Table 1. Main characteristics of the studies ordered by intervention.

First author year	N Number groups	Random	Intervention	Diagnosis	Inpatient/ Outpatient	Main outcome measures	Results per outcome measure ^a	Costs ^b	Sens. analysis	Dis-counting ^c	Nature study ^d
<i>Children</i>											
Clark 1986	310 2	yes	health education	asthma	out	hospitalizations emergency room visits	- -	direct 1,2	no	irrelevant	CEA
Fireman 1981	26 2	no	health education	asthma	out	wheezing asthma attacks compliance school absenteeism emergency room visits hospitalizations	- + - * + + * + *	direct 1,2	no	no	CEA
Lewis 1984	76 2	yes	health education	asthma	out	knowledge/beliefs skills family dynamics compliance behavior emergency room visits hospitalizations hospitalization days	- * + + * + * + - +	direct 1,2	no	irrelevant	CEA
<i>Adults</i>											
Deter 1986	22 2	no	health education	asthma	out	absence due to illness hospitalization days primary phys. visits	+ - -	direct indirect 1,2,4	no	no	CEA
Wilson-Pessano 1987	300 4	yes	health education	asthma	out	self management skills morbidity functional impairment utilization of services patient knowledge/attitude					CEA
Windsor 1990	267 2	yes	health education	asthma	out	inhaler use inhaler adherence drug adherence total adherence	+ + + +	direct 1	no	no	CEA
Toevs 1984	76 2	yes	pulmonary rehabilitation	asthma COPD	out	QWB, well-years	+	direct 1,2	yes	5% 10%	CUA

First author year	N	Number groups	Random	Intervention	Diagnosis	Inpatient/ Outpatient	Main outcome measures	Results per outcome measure ^a	Costs ^b	Sens. analysis	Dis-counting ^c	Nature study ^d
Wright 1983	74	before-after	-	pulmonary rehabilitation	COPD	out	exercise ability subjective gains hospitalizations hospitalization days	+ +* +* +*	direct 1,2	no	irrelevant	CEA
Lowson 1981	-	-	-	oxygen delivery methods	-	out	-	-	direct 1,3	yes	7%	CMA
McKeon 1987	111	before-after	no	oxygen delivery methods	COPD Cardio-vascular	out	-	-	direct 1	no	no	CA
Ström 1990	41	2	no	oxygen delivery methods	COPD TBC	out	SIP ability to handle equipment daily life activities sleep mobility social contact	- +* -* -* +* +*	direct 1	no	no	CEA
Schmidt 1983	137	-	-	mechanical ventilation	Pulmonary Cardiovascular Neuromuscular Postoperative complications	in	life years gained	+*	direct 1	no	no	CEA
Campbell 1991	17	before-after	-	home care	asthma COPD bronchiectasis	out	hospitalizations hospitalization days emergency room visits	+ + +	direct 1,2	no	no	CEA
Dranove 1985	436	2 hospitals compared	-	home care	COPD CVD	out	length init. hosp. stay hospitalization days emergency room visits outpatient visits	+ + - +	direct 1,2	no	irrelevant	CEA
Motwani 1988	10	-	-	home care	severely disabled respiratory patients	out	-	-	direct 1,2	no	irrelevant	CA

First author year	N	Number groups	Random	Intervention	Diagnosis	Inpatient/ Outpatient	Main outcome measures	Results per outcome measure ^a	Costs ^b	Sens. analysis	Dis-counting ^c	Nature study ^d
Bergner 1988	301	3	yes	home care	COPD	out	survival SIP general well-being walking tolerance FEV ₁ , FEV ₁ /FVC	- - - -	direct 1,2,3	no	irrelevant	CEA
Scheffler 1981	-	-	-	aerosol bronchod. delivery methods	-	in	-	-	direct 1	no	irrelevant	CMA
Jasper 1987	36	2	yes	aerosol bronchod. delivery methods	asthma COPD	in	FEV ₁ , FVC hospitalization days	- -	direct 1	no	irrelevant	CEA
Gay 1991	18	cross-over	yes	aerosol bronchod. delivery methods	mechan. ventilated patients with airways obstruction	in	iso recoil flow	-	direct 1	no	irrelevant	CEA
Hay 1991	-	-	-	alpha-1 antitrypsin replacement	congenital COPD	in	life years gained	+*	direct 1,2,3	yes	2% 5% 7%	CEA

a Results per outcome measure:

- no significant difference between the groups
- + a significant difference between the groups
- * the authors report that there is no difference between the groups, but this was not statistically tested
- +* the authors report that there is a difference between the groups, but this was not statistically tested

b Costs:

1. program costs (direct)
2. costs of other health care utilization (direct)
3. out of pocket expenses of patients and their families (direct)
4. costs of productivity losses due to absence from work (indirect)

c Discounting:

- no if no discounting and the study period > 1 year
- not relevant if the study period ≤ 1 year

d Nature of the study:

- Full economic evaluation:
- CMA Cost Minimization Analysis
 - CBA Cost Benefit Analysis
 - CEA Cost Effectiveness Analysis
 - CUA Cost Utility Analysis
- Partial evaluation:
- CA Cost Analysis

2.5 Alternatives compared

Since economics is about choice, the point of departure for an economic evaluation is to specify the alternatives. In all studies, a comparison is made between two or more alternatives, although they are not always explicitly given. In 7 studies the alternative was "doing nothing", which can be described more precisely as "continuing conventional or usual care".

2.5.1 Health education for adults

Health educational programmes are often compared with "doing nothing". Here "doing nothing" is considered to be a reasonable alternative since patients participate in such programmes alongside their usual, mainly pharmacotherapeutic treatment without fear of an acute situation if the programme is omitted. Since adherence to long term pharmacotherapy is known to be low in asthma and COPD patients, Windsor *et al.*,²² developed a standardized health education programme to increase therapy adherence. At the 12 month follow-up, the experimental group of patients (n=132) showed a significantly higher level of improvement in correct inhaler use, inhaler adherence, medication adherence (self reported and checked with a theophylline level test) and total adherence compared to the do-nothing control group patients (n=135). Excluding intervention development costs, the programme costs \$32.03 per patient.

Deter⁸ also compared the cost-effectiveness of health education combined with group coping therapy for working adult asthmatics with a do-nothing control group. He found a significant reduction in the number of working days lost and a non-significant reduction in the number of working days hospitalized and number of visits to the general practitioner, resulting in estimated savings of DM 1,951 per patient.

Few educational efforts directed at adults seem to have been evaluated economically. In an attempt to change this situation Wilson-Pessano *et al.*²¹ report their plans for a cost-effectiveness study in which 300 patients are to be randomly divided into 4 treatment groups: an individual self-management group, a group self-management group, a self-management information/attention control group and a data only control group. The emphasis will be on self-management practices, morbidity, functional impairment, medical regimen and the utilization and costs of services.

2.5.2 Health education for children

Educational programmes in children have been studied more widely. Three of these studies also considered cost-effectiveness.^{7,10,14} All three, by comparing programme costs with savings due to a decrease in emergency room visits, hospitalizations and drug use, reached the conclusion that educational programmes for children are cost-effective. When compared to a 'do-nothing' control group Fireman *et al.*¹⁰ have shown that health education directed at both the child and the parents resulted in a significant reduction in asthma attacks, school absenteeism, emergency room visits and hospitalizations. Compliance with drug therapy was also better. The costs of additional drug use and the costs of the educational programme provided by a nurse-educator were half the savings due to reduced hospitalizations and emergency room visits, savings of \$2,920 and \$225 per child per year respectively. In a randomised trial, Lewis *et al.*¹⁴ compared a detailed, small-group asthma care training programme directed at both children and their parents (n=48) to a control therapy consisting of a few large group sessions (n=28). They estimated savings of approximately \$180 per child per year for those in the experimental group. Besides a significant reduction in emergency room visits and hospitalizations, they also found a significant reduction in parents' reports of children's dependency on adults for care and of children's trouble with asthma as well as improved family communication. In the study by Clark *et al.*,⁷ no significant reductions in hospitalizations and emergency room visits were found when the complete experimental and 'do-nothing' control groups were compared. However when the comparison was restricted to two subgroups, one of children with at least one hospitalization in the year preceding the intervention, the other to children with at least 1 hospitalization and 6 emergency room visits in the previous year, significant reductions and savings were found. The authors might have drawn the conclusion that health education may only be cost-effective when directed at severely afflicted patients with at least one hospitalization a year. However they did not do so, because positive effects in terms of improved self management skills, school performance and school adjustment reported elsewhere have shown health education to be effective for wider groups.

2.5.3 Pulmonary rehabilitation

The continuation of the situation before the programme is the alternative to which Wright *et al.*²³ compared their pulmonary rehabilitation programme. Pulmonary rehabilitation can be defined as an individually tailored, multidisciplinary programme which, through accurate diagnosis, therapy, emotional support and education, stabilizes or reverses the progress of the disease and seeks to return the patient to the highest possible functional capacity.²⁹ Wright *et al.*²³ estimated the savings achieved by pulmonary rehabilitation due to the reduction in

hospital days in the first year following the programme to be \$217,610 for a group of 57 patients for whom follow-up data were available. Exercise ability also increased. The costs of the programme amounted to \$66,462, resulting in net savings of \$151,148 per year for these 57 patients.

Providing exercise programmes and prescribing daily exercise alone is not enough. As with drugs, non-compliance with exercise is a common problem. Toevs *et al.*²⁰ studied the effects of using behavioural programmes to improve compliance with an exercise programme. The control group was assigned an exercise programme but did not receive a behavioural programme. Adding the behavioural programme to the exercise programme produced one additional quality adjusted life year at a cost of \$24,256 per patient. Assuming that the effects will last for 2.5 years, the cost per QALY drops to \$18,455.

2.5.4 Domiciliary oxygen therapy

"Doing nothing" is not a realistic alternative for oxygen therapy. Repeated hospitalizations or in the long run even admission to a nursing home could become necessary in the absence of this therapy. A study comparing oxygen therapy with the above mentioned alternatives is unlikely to be designed and conducted because of the overwhelming acceptance of oxygen therapy.³⁰ Its efficacy in terms of reduced mortality and improved neuropsychological function has been established by several clinical trials. Since it does not seem necessary to ask whether domiciliary oxygen is worthwhile *per se*, studies of oxygen therapy are concerned with the alternative ways of delivering the oxygen. In three studies^{15,16,19} concentrator oxygen - a concentrator is an electrically driven machine which extracts oxygen from the air- has been compared to gas cylinder oxygen in order to determine the most efficient way of delivering oxygen at home. The first of these studies also added liquid oxygen to the comparison. Lawson *et al.*¹⁵ explicitly state that the alternatives under consideration seem to be equally effective. This is confirmed by the finding of Ström *et al.*¹⁹ that scores on the Sickness Impact Profile did not significantly differ between the groups treated with gas cylinder oxygen and concentrator oxygen. Although quite different in their methodology and calculation of costs, all three studies reached the overall conclusion that in most situations the concentrator was the least expensive method of delivering oxygen. In addition Ström *et al.* found that concentrators were easier and safer to handle and can be managed by the patient him or herself in most cases.

Unlike cylinders, concentrators cannot always be hired, and the oxygen delivery and maintenance cannot thus be provided by the manufacturer. In that case large capital outlays may be needed from the health care sector at the start of a concentrator oxygen project. This is not only due to the high purchase price of the equipment but also to the need to establish a workshop facility from which the concentrators are maintained.¹⁵ Where these large capital costs apply, concentrator

oxygen therapy is more cost-effective the larger the number of patients served by this unit. Renting or purchasing concentrators seems to make no difference to maintenance costs.¹⁶

2.5.5 Home care

Home care is usually compared to alternatives which may well be substitutes for it, such as institutional care or doctor's office care.^{5,6,9,17} Two studies^{5,9} found no significant reductions in health care costs due to home care. Bergner *et al.*⁵ randomly assigned homebound COPD patients aged 40-75 to one of three groups: respiratory home care (n=99), standard home programme (n=102) or doctor's office care without additional home nursing services (n=100). They found no significant difference between the three groups in survival, pulmonary function, walking rate, activities of daily living, and scores on the General Well Being and the Sickness Impact Profile. The home care programmes even lead to additional health service utilization rather than to the substitution of ambulatory care for inpatient care. Also, the imputed costs for unpaid help for the respiratory and standard home care groups did not differ significantly from those for the office care group. The reason no cost reduction was found may be that Bergner *et al.* studied community-based home care programmes. Providing these programmes, as compared to hospital-based home care programmes, may imply less communication between community based nurses and the treating physician about services that could be substitutes for regular check ups at the hospital.

Dranove,⁹ by comparing two hospitals, one with and one without a home care programme, showed that home care significantly substitutes for the last few days of a hospital stay and outpatient visits. Home care reduced the initial length of stay of patients older than 65 years suffering from cerebrovascular disease (CVD) or COPD by 3.4 days and the number of clinic visits by 6.2 per patient. No reduction in the number of emergency room visits was found. Despite this substitution Dranove found no significant reduction in overall hospital costs because of the relatively large costs of the home care programme.

Probably home care is more cost-effective in highly selected patient populations. For example, Campbell *et al.*⁶ studied the effects of a hospital based home care programme in severe asthma, COPD or bronchiectasis patients showing substantial co-morbidity and frequent hospitalizations due to pulmonary disease. An additional inclusion criterion was the presence of a significant other in the home or community willing to assume the role of caretaker. When pre-programme costs were compared to programme costs, the overall health care savings were \$328 per patient per month. Similarly, for respirator-dependent patients needing around-the-clock home care, Motwani *et al.*¹⁷ found the costs of home care to be lower than the cost of care in the respiratory centre which was considered the only possible alternative for these patients.

2.5.6 Aerosol delivery of bronchodilator drugs

Two studies reported on the cost-effectiveness of an updraft nebulizer compared to a metered dose inhaler as alternative methods to deliver aerosol bronchodilator drugs in hospitalized patients.^{11,13} Both studies included the same indirect costs e.g. labour costs, medication costs and the cost of the devices. The randomized study by Jasper *et al.*¹³ involved patients with exacerbated obstructive lung disease, most of whom were transferred to the pulmonary ward from the emergency room or intensive care units. They received bronchodilation treatment with metaprotenerol via therapist administered updraft nebulization or via metered dose inhaler. The randomized cross-over study by Gay *et al.*¹¹ involved intubated mechanically ventilated patients, some of whom may not have had obstructive lung disease.¹⁰ In the case of the intubated patients the nebulizer and the metered dose inhaler were attached to endotracheal tubes. Whereas Gay *et al.* performed a single treatment comparison, Jasper *et al.* assessed not only immediate response to bronchodilator drugs but also response during hospitalization and response at discharge from the hospital. In the first study no differences were found between the delivery methods so far as iso-recoil flows or side effects in terms of an increased heart rates. Jasper *et al.* found no differences between the delivery methods with respect to initial post-bronchodilator and discharge post-bronchodilator FEV₁ (Forced Expiratory Volume in one second) and FVC (Forced Vital Capacity) and duration of hospital stay. In addition, both studies found that the metered dose inhaler was more cost-effective than the updraft nebuliser, and may be substituted for it. According to Jasper *et al.* such a substitution would not be appropriate for patients in the intensive care unit or the emergency room. However, the 1990 study by Gay *et al.* showed that the metered dose inhaler was also more cost-effective for these patients.

The 1981 Office of Technology Assessment report by Scheffler *et al.*²⁴ contained estimates of the cost of delivering aerosol medication in hospitalized patients using three different devices: an intermittent positive pressure breathing machine (IPPB), an ultrasonic nebulizer and a simple aerosol generator. Equal effectiveness was assumed and supported by referring to several studies. The direct costs of labour, equipment, disposables and electricity were considered. Again simple aerosol treatment proved to be the least costly both as initial treatment and follow-up treatment.

2.5.7 Mechanical ventilation

When the IPPB is used for mechanical ventilation purposes where a patient's breathing has totally ceased owing to the suppression of brain activity, paralysis of the respiratory muscles, chest wall injury or acute and potentially reversible lung disease, its effectiveness is well accepted, and is similar to that of other mechanical aids to support the ventilatory function. In such cases "doing nothing" would inevitably lead to immediate death. Although the method is well-accepted, it remains a problem to select the patients most likely to benefit from prolonged mechanical ventilation. Schmidt *et al.*¹⁸ showed in their study that the cost of extending life for one year with mechanical ventilation varied from \$460 per year for patients who had respiratory failure as a complication of asthma and \$2,155 for COPD to \$8,026 per year of extended life for cardiac patients. Cost-effectiveness ratios were considerably lower for patients with asthma and COPD largely because the mortality during hospitalization was lower. Only the total non-physician charges of hospital care were included and a portion of the costs of the non-survivors in a diagnostic group was added to the costs of survivors in that group.

2.5.8 Alpha-1 Antitrypsin Replacement Therapy

Congenital Alpha-1 Antitrypsin (AAT) deficiency results in COPD. AAT can be replaced intravenously and this replacement therapy - the costs of which are estimated to be \$30,000 per year - has been shown to improve levels of AAT in serum and epithelial tissue. However, clinical efficacy in slowing down COPD progression is unknown. Conducting a controlled trial would take a very long time due to the small number of potential beneficiaries. Therefore Hay *et al.*¹² took available data on congenital COPD incidence, prevalence, survival and costs of COPD treatment and applied it to a cost-effectiveness model of AAT replacement therapy in homozygous AAT-deficient COPD patients. A range of hypothetical efficacy values (expressed as a percentage of the survival rate of the normal population) was compared to the cost of AAT replacement therapy and other lifetime health care costs after the diagnosis of COPD. At an efficacy of 70%, the costs per life year saved would be between \$28,000 and \$72,000 depending on patient age, sex and smoking status. Quality of life adjustment does not alter these results drastically.

2.6 Outcomes

Sixteen of the 20 studies examined both the costs and the outcomes, using a great variety of outcome measures which can be classified into four main categories. More detailed information on the outcome parameters used and the significance of the effects measured by them can be found in the column 'main outcome measures' and 'results per outcome measure' of Table 1.

2.6.1 Clinical measures

Clinical measures include physiological measures such as lung function parameters, physical measures such as exercise tolerance and walking distance, and laboratory assessments. Lung function parameters are essential for monitoring the course of asthma and COPD, because of their known relationship to disease prognosis.³¹ However, interventions such as health education, pulmonary rehabilitation and home care^{7,8,32} do not seem to alter spirometric data. Therefore, most of the studies dealing with these interventions do not include lung function parameters. Relevant effects of health education are usually measured in terms of increased knowledge, improved skills, improved attitude towards asthma and reduced health care utilization. The effects of pulmonary rehabilitation are usually measured in terms of improved oxygen consumption and physical measures such as walking distance and exercise tolerance.²³

An evaluation confined to a fairly narrow assessment of physiological status, as in Jasper *et al.*¹³ gives little information about patients' overall functioning. Moreover, measures such as walking distance look meaningful and easy to analyze, but may bear only a loose relation to the patient's underlying condition. It has been shown that a therapy which produces changes in clinical status may not produce corresponding changes in quality of life.³³⁻⁴⁰ Other factors such as the patient's attitude, expectations, emotions and ability to adapt, and sociological factors may intervene. Since it is likely that clinical and quality of life measures address different aspects of the disease, a meaningful outcome assessment should include both.

2.6.2 Disease-specific quality of life measures

The studies which did use quality of life measures of outcome such as Lewis *et al.*¹⁴ and Ström *et al.*,¹⁹ who even developed a study-specific questionnaire, generally present a pattern of different aspects of quality of life directly related to having the particular disease. Such a heterogeneous description of changes in disease-specific quality of life may be sufficient, particularly when the results all point in the same direction and no comparison with interventions in other

patient populations is envisaged. The rationale for using such disease-specific outcome measures lies in their potential for increased responsiveness. Although disease-specific measures may be more sensitive to disease-specific changes, they lack general reference points which enable across-programme comparisons.

2.6.3 Generic quality of life measures

In order to allow such comparisons and in order to facilitate the interpretation of the heterogeneous results one may want to combine them into a few subdimension scores or into one overall summary score, called a health profile and a health index respectively. Both Ström *et al.*¹⁹ and Bergner *et al.*⁵ attempted to quantitatively assess the relative importance of their information using a measure of the overall quality of life such as the SIP. The SIP is a generic measure of health-related dysfunction in 12 areas of activity: ambulation, mobility, body care and movement, social interaction, emotional behaviour, alertness behaviour, communication, work, eating, sleep and rest, household management and recreation and pastimes. Besides 12 category scores, it provides an overall score, a psychosocial dimension score and a physical dimension score.

2.6.4 Utilities and QALYs

The primary objective of indices and profiles is an overall summary description of health status. One may go one step beyond the description by explicitly valuing the summarized descriptions of the changes in health status. This is particularly necessary when measures of different dimensions of the quality of life are of different size or sign. By means of utility measurement such positive and negative effects may be evaluated, balanced and integrated in an overall value for the quality of life, the so called utility. Toevs *et al.*,²⁰ who used the Quality of Well-Being Scale developed by Kaplan and Bush,^{41,42} converted the quality of life gains into a single index of so-called "well-years" or QALYs. At baseline and follow-up each patient's health state was classified into 1 of 43 mutually exclusive levels of functioning. The levels were obtained by combining different levels of mobility, physical activity, social activity and different symptomatic disturbances. Then utilities were assigned to each patient's health state. These utilities, which reflect the preference weights for the various health states, were elicited from the general population. Finally each patient's utility was multiplied by the number of years the effects were expected to last. This technique is relatively new and still rarely used in the domain of chronic diseases, because it is expected to be insensitive to slight improvements (or reductions) in health status.⁴³ But, particularly because economic evaluation studies aim at providing information for overall comparisons between interventions, QALYs are promising outcome measures, even

though a number of methodological problems need to be solved. However, the study by Toevs *et al.* shows this approach to be feasible in evaluating treatment for a common chronic condition such as COPD.

Sometimes the outcome assessment is restricted solely to the intermediate effects of treatment on reductions in health services use. Although these outcomes can be measured solely in volume terms, they are usually measured in monetary units and incorporated in the numerator of a cost-effectiveness ratio, by the calculation of net incremental costs. The same is true for absence from work due to illness, although the costs associated with sick leave are often presented separately. Therefore we describe them more thoroughly in the next section. Like absence from work in adults, school absenteeism is an important outcome measure in studies involving children. Only Fireman *et al.*⁹ studied the impact of health education on school absenteeism, finding that it was reduced tenfold.

2.7 Costs

In the reviewed studies four main categories of costs could be identified:

2.7.1 Costs of organizing and operating a particular health care programme

The studies dealing with alternative ways of delivering aerosol bronchodilation^{11,13,24} all adopt an administrator's perspective in which only programme costs are considered. All found that delivering medication through a metered dose inhaler is the least expensive, mainly because of lower labour costs but also because of lower equipment costs. The comparison of alternative ways of delivering oxygen is also restricted to programme costs.^{15,16,19} Ström *et al.*¹⁹ even restricted their analysis to annual running cost per patient -which were twice as high for gas cylinders as for concentrators- without considering set-up costs. However set-up costs can be very important, as can be seen from the studies of domiciliary oxygen therapy, by McKeon *et al.*¹⁶ and Lawson *et al.*¹⁵, which were described above.

In evaluating alternative ways of delivering aerosol bronchodilation or oxygen, a restriction to programme costs may be justified, since equal effectiveness between the delivery methods is assumed and no change in the use of other health services is to be expected. However in the study by Windsor *et al.*²² of health education this may be too limited a scope. If health education is as effective in increasing medication compliance as Windsor *et al.* showed, than a decrease in the use of

other health care services such as visits to doctors' surgeries may be expected and should have been measured.

When calculating costs associated with providing a particular programme, resources that are used solely for the purpose of the evaluation study itself have to be identified and subtracted from the programme costs, but few studies do so. Clark *et al.*⁷ and Lewis *et al.*¹⁴ are exceptions, in separating the costs of evaluating the educational programmes from routinely incurred costs. Unfortunately, these studies also excluded programme development costs, which should have been included.

2.7.2 Costs of utilization of other health services

With the exception of the study by Windsor *et al.*,²² all the studies concerning health education^{7,8,10,14,21} or home care^{5,6,9,17} do include the costs of other health services. They are particularly interested in savings as a result of reduced emergency room visits, clinic visits, and hospitalization days. The latter was of prime importance in the two studies dealing with pulmonary rehabilitation.^{20,23}

It frequently occurs that the mean cost of hospitalization is very much affected by a small number of expensive cases. As a result, differences in the net costs of two programmes may be influenced by only a few cases. Since this reduces statistical power, testing differences in hospitalization costs is difficult.

Savings in health care resource use may be either compared with the programme costs or deducted from the programme costs in order to calculate the net cost of the intervention. Clark *et al.*⁷ used the first approach and found that for children who had been hospitalized during the year preceding the programme, health education reduced emergency room visits and hospitalizations, saving \$11.22 for every \$1.00 spent on health education. Toevs *et al.*²⁰ used the second approach and calculated the net costs of the intervention by deducting the hospital costs which were saved from the direct costs of the behavioural intervention programme.

2.7.3 Out-of-pocket expenses of patients and their families

The most comprehensive calculation of COPD health care costs is provided by Bergner *et al.*³ They include, from a societal viewpoint, all relevant inpatient and outpatient costs, which they obtained from several sources such as patients, bills, receipts, third party payers, pharmacists etc. Like Deter,⁸ they also included out-of-pocket payments by patients, mainly travelling expenses. A third study also included costs falling on patients, but a separate identification of these costs is not provided.¹⁵

2.7.4 The loss of productivity due to absence from work, imputed volunteer time or family time.

The first three categories of costs are often referred to as direct costs. In economic evaluation jargon the term indirect costs refers primarily to production losses due to absence from work because of illness. Measuring the reduction in days of absence due to illness is especially relevant in asthmatic patients and patients with mild airflow limitation. Measuring the impact of treatment on employment status may not be very relevant for patients with severe chronic airflow limitation, since this mostly concerns relatively older patients who have only a few work years remaining and who would probably continue to decrease in work ability during those years because of their lung disease. Asthma and COPD interventions can only make an impact in terms of returning people to the work force for patients whose lung disease is identified at a younger age, when the disease is less progressive. Although relevant for less severely afflicted patients, measuring the impact of treatment on employment status may be difficult. Factors such as the national unemployment rate, duration of unemployment, age etc. may influence the percentage of patients returning to work. The study by Deter,⁸ which involved only working adult asthmatics, is the only one to consider savings due to a reduced number of working days lost. Loss of productivity due to absence from work on days of group therapy was included in programme costs. Production loss was estimated by means of individual real market wages. Earnings are commonly used but also frequently criticized measures of production losses, since it is questionable whether production is actually lost, especially in times of unemployment. Also, as a consequence of this approach, higher priority in the access to care would be given to high income earners. Perhaps a more realistic but difficult to perform alternative would be to consider the real costs to employers arising from the replacement and re-education and training of employees. Drummond *et al.*²⁶ advise that productivity gains and losses should not be included in calculating the net costs of an intervention because of the problem of the differential valuation of time for those in and those out of the labour force. It certainly seems preferable to count changes in productivity separately from other costs and outcomes. One may also measure restricted activity days for patients who are not in the work force and estimate their productivity loss by, for example, using the wages paid employees would get when performing the same work. Such a proxy is used by Bergner *et al.*,⁵ when they estimated the imputed costs of unpaid help by family or friends using the costs of home health aides.

2.7.5 Costs versus expenditures and charges

In general, costs refer to the sacrifice made when a given resource is consumed, and not to expenditures only. Wright *et al.*²³ should not have excluded the services for which members of the rehabilitation staff donated their time from the programme costs. Although donated time requires no expenditures, it does involve resource use, because these staff members cannot put their time into alternative services.

Although detailed costing methodology is often lacking it can be seen that a majority of studies still use average costs (per hospital day, per home nursing visit, average wages etc.) or charges. This may not reflect the real costs borne by society for a particular treatment. The average daily hospital costs are likely to be a poor guide for the savings in resource use if hospital stays were shortened, since resource use is typically higher in the first part of a patient's hospital stay.⁴⁴ Because home nursing can be a substitute only for the last few days of a hospital stay, Dranove *et al.*,⁹ in studying the impact of a home nursing department on the savings in hospital costs, thought it appropriate to include only the costs of routine care and clinical laboratory work in the average daily hospital cost. In comparing respiratory centre care with home care, Motwani *et al.*¹⁷ showed that the actual costs of ventilator dependent respiratory centre care were much higher than Medicaid reimbursement rates. This was because higher levels of staff were required in the respiratory centre than in the other hospital departments, although the reimbursement rates were the same. Based on charges Motwani *et al.* concluded that home care was more cost-effective than respiratory centre care. When based on actual costs, home care would become even more cost-effective. Sometimes, using the amounts charged as a proxy for costs may not introduce a systematic bias into identifying differences between treatment programmes, although it still risks misstating the magnitude of any cost differences.

An incremental analysis, as for example in Lowson *et al.*,¹⁵ does not consider costs which are common to both alternatives. It deals therefore with differential costs, rather than with the total treatment cost. An incremental analysis may be sufficient to support decisions concerning only the programmes studied. But if a comparison with other alternatives, perhaps outside the medical domain of the intervention is wanted, then total costs are needed.

2.7.6 Discounting and sensitivity analysis

Sound economic methodology also prescribes that costs and benefits occurring at different times should not just to be added or subtracted, but rather expressed in terms of their equivalent values at one point in time. This adjustment procedure is called discounting and usually involves calculating the net present values of future costs and outcomes using the same discount rate. The discounting of health effects

is not uncontroversial (for a recent debate see ^{45,46,47}). In most of the studies, discounting was irrelevant, because no long term consequences were studied. Hay *et al.*,¹² Lowson *et al.*¹⁵ and Toevs *et al.*²⁰ have applied discounting. They also studied the influence of the application of various discount rates (0%, 5%, 10%) in a sensitivity analysis. Both Lowson *et al.* and Toevs *et al.* concluded that the various discount rates have little influence on the cost-effectiveness ratios found in their studies because only a limited time period was considered. Discounting did have a larger effect on the evaluation of Alpha-1 Antitrypsin replacement therapy, because here lifetime treatment costs and outcomes were estimated.¹² In general, programmes that require current investments to achieve benefits far in the future are more strongly affected by the choice of a discount rate. Therefore a screening programme for Alpha-1 Antitrypsin deficiency could be even more affected by the choice of discount rate than is Alpha-1 Antitrypsin replacement therapy itself.

Each study should perform a sensitivity analysis to show the influence of several assumptions on the overall cost-effectiveness ratio. Hay *et al.*¹² have provided an exemplary sensitivity analysis, in which they show the effects of a 40% increase and decrease of not only discount rates but also of annual costs of COPD care, the quality of a restricted activity day, the median survival from onset of COPD and annual costs of AAT treatment. The cost-effectiveness ratio proved to be most sensitive to the assumptions about the latter two factors.

As Guyatt *et al.*²⁷ point out, it is important to analyze the costs (and outcomes) from a broad perspective. This means the inclusion of all four categories of costs, as far as they are relevant. This may also mean the consideration of various viewpoints (government, Ministry of Health, patients, health care providers) if one wants to understand more fully the gains and losses to key actors in the system.

2.8 Quality of the effectiveness evidence

It is argued by Drummond⁴⁸ that economic appraisals should be based on reliable evidence of effectiveness. It is certainly possible to perform a good economic appraisal without actually comparing two or more patient groups. But especially in that case, effectiveness information based on previous good quality studies should be provided. Good examples of such studies are the ones by Lowson *et al.*,¹⁵ Schmidt *et al.*¹⁸ and Scheffler *et al.*²⁴ If no such information is available a range of possible effectiveness levels should be considered to give some information on the sensitivity of the cost-effectiveness to these levels. This was done by Hay *et al.*¹² However, Motwani *et al.*¹⁷ used the opinion of 10 patients and their medical professionals: this would seem to be insufficient evidence to support the assumption that respiratory centre care was the only appropriate substitute for home care.

Ten out of the 20 economic appraisals actually compared two or more patient groups,^{5,7,8,10,13,14,19-22} and in 7 of these 10 studies patients were randomly divided into an experimental and a control group.^{5,7,13,14,20-22} In experimental studies that concentrate on the impact of health education or pulmonary rehabilitation, an observed effect could partly be due to the attention the experimental group received. Giving a placebo to a control group in studies other than those concerning pharmacotherapy is considered to be difficult. However, Wilson *et al.*²¹ have tried to control for the personal attention given to the experimental groups by designing a study which incorporates two control groups, one in which patients get the same written information as the experimental group but no personal attention, and one in which patients get no information or attention at all. The former is called an attention-placebo.²⁷ Lewis *et al.*¹⁴ also provided their control group with the same knowledge as the experimental group. However the knowledge was provided in impersonal large group sessions, whereas the experimental group received more personal attention in small groups.

Three studies were before-and-after comparisons of just one patient group.^{6,16,23} Because of this design, the favourable outcomes which both Campbell *et al.*⁶ and Wright *et al.*²³ found cannot solely be attributed to the intervention. Other factors such as the natural course of the disease or regression towards the mean might have explained at least part of the improvement.

The study by Gay *et al.*¹¹ used a cross-over design, which is used mainly to enhance the power of the statistical comparisons. In the case of different aerosol delivery methods, no carry-over effect (in which the effect of administering albuterol by means of a metered dose inhaler in the first period influences the effect of administering it by means of an updraft nebulizer in the second period) is likely, since a wash-out period of 4 hours was allowed between the first and the second delivery.

Drawing conclusions based on a comparison of two health care institutions is extremely difficult because of the large number of biases that can occur. Dranove⁹ tried to avoid the bias that might have resulted from the comparison of two hospitals -one with and the other without a home care programme- by taking younger patients who were not eligible for home care as a kind of control group. Instead of comparing the mean length of hospitalization for elderly patients in the two hospitals, he compared the difference in the length of hospitalization of younger and older patients in one hospital with the difference in the other hospital. Thus he partly controlled for differences between patient populations. Despite this, he could not completely rule out the possibility that differences between the hospitals affected his study results.

Some studies are of poor quality due to the inadequate sample size. These studies may not have the power to detect clinically significant differences. The fact that Deter⁸ and Gay *et al.*¹¹ found no significant difference between the experimental and control group in terms of hospitalization days and visits to the primary physician may be due to the small sample size. Deter, moreover, divided

patients (n=57) randomly into an experimental and a control group before applying the inclusion criterion that patients had to be employed. Therefore only 22 patients remained, of whom only 7 patients in the experimental group and 8 in the control group finished the whole study. Large numbers of withdrawals can cause serious analysis problems, especially when withdrawal is due to a deterioration only occurring in the control group. When deterioration also leads to higher health care cost, the cost-effectiveness ratio will be overestimated. In the study by Clark *et al.*⁷ more than 20% of the patients were lost to follow-up, and in the study by Bergner *et al.*⁵ 21% were lost. Since neither provided a comparison of these patients with the patients who completed the study it is impossible to say whether this influenced the cost-effectiveness ratio in any way.

Obviously, the most reliable effectiveness evidence is that produced by a well-designed randomized controlled trial. Performing an economic evaluation alongside such a trial entails a number of problems. Even those trials which are directly concerned with assessing the effectiveness rather than the efficacy of the therapy may still be conducted under such rarefied experimental conditions that costs and outcomes will be quite different to those under normal conditions.⁴⁹ There remains a need to assess therapies as they are routinely delivered in actual daily practice.

2.9 Conclusions and discussion

Although central governments and third party payers increasingly stress the need to set priorities on the basis of efficiency considerations, there have been only 20 economic appraisals of asthma or COPD-interventions published in the last 11 years, most of them American. Because of the small number of studies per subject, the considerable differences in the measurement of both costs and outcomes and other methodological issues, only very preliminary conclusions can be drawn.

Health educational programmes seem to be relatively inexpensive and in children they seem to result in savings due to less hospitalizations and emergency room visits. Moreover there seem to be indications for improvements in attitude, self-management skills and school absenteeism. The study by Windsor *et al.*²² is the only economic appraisal dealing with health education in adults offering guidance for a cost-effective improvement in compliance behaviour. Toevs *et al.*²⁰ also found evidence of increased compliance when a behavioural programme was added to an exercise programme. The costs per QALY for this intervention were more than \$24,000. The conclusions that can be drawn from the studies of inpatient aerosol drug delivery methods and oxygen delivery methods are perhaps the most straightforward, the first indicating that the simple metered dose inhaler is the most cost-effective and the latter supporting the concentrator option, provided that the number of patients to be served is not too small. The results of the studies dealing with home care point in different directions. However, there seems to be evidence in

favour of hospital-based home care programmes directed at highly selected, very severely afflicted patient groups, as compared to community-based home care programmes in wide patient populations.

Physicians are responsible not only for providing an optimal therapy from a medical viewpoint, but also for providing an efficient therapy. More cost-effectiveness evidence is required on the various asthma and COPD interventions described above. This is especially true for the relatively expensive treatments such as pulmonary rehabilitation and home care. Moreover, in two particular areas cost-effectiveness evidence seems to be totally lacking. The most striking is the lack of any cost-effectiveness study concerning pharmacotherapy, although this is the main intervention in asthma and COPD management. Of all the bronchodilators that are used in current practice, we do not know which ones are the most efficient for patients exhibiting particular clinical symptoms. We do not know whether the extra benefits of combined bronchodilator and corticosteroid therapy compared to bronchodilator monotherapy outweigh the extra costs. We argue that -in addition to the relatively large number of effectiveness studies- there is a need for subjecting these relatively "low cost" therapies to a thorough economic evaluation, since more than 7% of total drug sales in the Netherlands are related to asthma and COPD management and because it involves so many patients. In the future, cost-effectiveness evidence will increasingly be asked for as a complement to the clinical evidence of the safety and efficacy of new drugs. Such evidence may influence the price-setting of new drugs and support cost-reducing therapeutic substitution for already existing drugs.

A second important area in which information is totally lacking is that of diagnostic technologies. For example, in some countries an economic evaluation of spirometry would be of value in guiding decisions about whether or not spirometry should be made available too for general practitioners. Similarly, some countries are considering screening for Alpha-1 Antitrypsin deficiency. This decision should be based on a thorough, perhaps theoretical cost-effectiveness analysis, in particular since the number of patients with Alpha-1 Antitrypsin deficiency is relatively small and replacement therapy is relatively expensive. For information on the costs and benefits of such technologies to be useful, it should become available before the technologies are diffused throughout the health care system.⁵⁰

Economic appraisals cannot be expected to have an influence on decision-making until the methodological quality of the studies is adequate. We have shown that there is still room for improvement. It is difficult to compare the results of various cost-effectiveness studies, because the analyses are not conducted according to uniform, accepted methodological standards. As these conditions are being met, economic evaluation studies can contribute to improved efficiency in delivering asthma and COPD care.

2.10 Acknowledgements

The authors would like to thank R.M. Kaplan, an anonymous referee and our colleagues G. Boas and V.H. Ploum for their comments on an earlier draft, Mr. J.P.M.A. van Emmerik for his assistance in preparing this manuscript and Glaxo Netherlands for their financial support.

2.11 References

1. American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD and asthma). *American Review of Respiratory Disease* **136**, 225, 1987.
2. Mülken, M.P.M.H. van, Doorslaer E.K.A. van, and Rutten, F.F.H. Chronic Non-specific Lung Disease in Figures. Report of a pilot-study (in Dutch). Institute for Medical Technology Assessment, Maastricht, The Netherlands, 1989.
3. Lako, C.J. Absenteeism in primary school (in Dutch). *Tijdschrift voor Sociale Gezondheidszorg* **61**, 776, 1983.
4. Wever, A.M.J. and Wever-Hess, J. Trends in the frequency of hospitalizations due to asthma and COPD in the Netherlands (in Dutch). 1980-1988. *Nederlands Tijdschrift voor Geneeskunde* **135**, 659, 1991.
5. Bergner, M., Hudson, L.D., Conrad, D.A., Patmont, C.M., McDonald, G.J., Perrin, E.B. and Gilson, B.S. The cost and efficacy of home care for patients with chronic lung disease. *Medical Care* **26**, 566, 1988.
6. Campbell-Haggerty, M., Stockdale-Woolley, R., and Nair, S. Respi-Care. An innovative home care program for the patient with chronic obstructive pulmonary disease. *Chest* **100**, 607, 1991.
7. Clark, N.M., Feldman, C.H., Evans, D., Levison, M.J., Wasilewski, Y. and Mellens, M.D. The impact of health education on frequency and cost of health care use by low income children with asthma. *J. Allergy Clin. Immunology* **78**, 108, 1986.
8. Deter, H.C. Cost-benefit analysis of psychosomatic therapy in asthma. *Journal of Psychosomatic Research* **30**, 173, 1986.
9. Dranove, D. An Empirical Study of a Hospital-Based Home Care Program. *Inquiry* **22**, 59, 1985.
10. Fireman, P., Friday, G.A., Gira, C., Vierthaler, W.A. and Michaels L. Teaching Self-Management Skills to Asthmatic Children and Their Parents in an Ambulatory Care Setting. *Pediatrics* **68**, 341, 1981.
11. Gay, P.C., Patel, H.G., Nelson, S.B., Gilles, B. and Hubmayr, R.D. Metered dose inhalers for bronchodilator delivery in intubated mechanically ventilated patients. *Chest* **99**, 66, 1991.
12. Hay, J.W. and Robin, E.D. Cost-effectiveness of Alpha-1 Antitrypsin replacement therapy in treatment of congenital chronic obstructive pulmonary disease. *American Journal of Public Health* **81**, 427, 1991.
13. Jasper A.C., Mohsenifar, Z., Kahan, S., Golberg, H.S. and Koerner, S.K. Cost-Benefit comparison of Aerosol Bronchodilator Delivery Methods in Hospitalized Patients. *Chest* **91**, 614, 1987.
14. Lewis, C.E., Rachelefsky, G., Lewis, M.A., de la Sota, A. and Kaplan, M. A randomized trial of A.C.T. (Asthma Care Training) for Kids. *Paediatrics* **74**, 478, 1984.

15. Lawson, K.V., Bishop J.M., and Drummond, M.F. Costing new services: long-term domiciliary oxygen therapy. *The Lancet* **1**, 1146, 1981.
16. McKeon, J.L., Saunders, A. and Murree-Allen, K. Domiciliary oxygen: rationalization of supply in the Hunter region from 1982-1986. *Medical Journal of Australia* **146**, 73, 1987.
17. Motwani, K.J. and Herring, G.M., Home care for ventilatory dependent persons: a cost-effective humane public policy. *Health and Social Work* **13**, 20, 1988.
18. Schmidt, C.D., Elliott, C.G., Carmelli, D., Jensen, R.L., Gengiz, M., Schmidt, J.C., Tolman, E.D. and Clemmer, T.P. Prolonged mechanical ventilation for respiratory failure: A cost-benefit analysis. *Critical Care Medicine* **11**, 407, 1983.
19. Ström, K., Boe, J., Herala, M., Boman, G., and Gustavii, A. Assessment of two oxygen treatment alternatives in the home. *International Journal of Technology Assessment in Health Care* **6**, 489, 1990.
20. Toews, C.D., Kaplan, R.M. and Atkins, C.J. The costs and effects of behavioral programs in Chronic Obstructive Pulmonary Disease. *Medical Care* **22**, 1088, 1984.
21. Wilson-Pessano, S.R., Scamagas, P., Arsham, G.M., Chardon, L., Coss, S., German, D.F., Hughes, G.W. An Evaluation of Approaches to Asthma Self-management Education for Adults: The AIR/Kaiser-Permanente Study. *Health Education Quarterly* **14**, 333, 1987.
22. Windsor, R.A., Bailey, W.C., Richards, J.M., Manzella, B., Soong, S.J. and Brooks, M. Evaluation of the efficacy and cost effectiveness of health education methods to increase medication adherence among adults with asthma. *American Journal of Public Health* **80**, 1519, 1990.
23. Wright, R.W., Larsen, D.F., Monie, R.G. and Aldred, R.A. Benefits of a Community-Hospital Pulmonary Rehabilitation Program. *Respiratory Care* **28**, 1474, 1983.
24. Scheffler, R. M. and Delaney, M. The implications of cost-effectiveness analysis of medical technology. Background paper #2: case study of medical technologies. Case study #12: assessing selected respiratory modalities: trends and relative costs in the Washington DC area. U.S. Government Printing Office, Washington D.C., 1981.
25. Williams, S.J. Chronic respiratory illness and disability: a critical review of the psychosocial literature. *Social Science and Medicine* **28**, 791, 1989.
26. Drummond, M.F., Stoddart, G.L. and Torrance, G.W. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford, 1987.
27. Guyatt, G., Drummond, M., Feeny, D., Tugwell, P., Stoddart, G., Haynes, R.B., Bennett, K. and Labelle, R. Guidelines for the clinical and economic evaluation of health care. *Social Science and Medicine* **22**, 393, 1986.
28. Drummond, M.F. (ed.). *Economic Appraisal of Health Technology in the European Community*. Commission of the European Communities. Health Services Research Series No. 2. Oxford University Press, Oxford, 1987.
29. Hodgkin, J.E. Pulmonary Rehabilitation. *Clinics in Chest Medicine* **11**, 447, 1990.
30. Petty, T.L. Home Oxygen Therapy. *Mayo Clin. Proc.* **62**, 841, 1987.
31. Postma, D.S., de Vries, K., Koëter, G.H. and Sluiter H.J. Independent influence of reversibility of airflow obstruction and non-specific hyperreactivity on the long-term course of lung function in chronic airflow obstruction. *Am Rev Respir Dis* **134**, 276, 1986.
32. Atkins, C.J., Kaplan, M., Timms, R.M., Reinsch, S. and Lofback, K. Behavioral exercise programs in the management of Chronic Obstructive Pulmonary Disease. *J Consult Clin Psychol* **52**, 591, 1984.

33. Williams, S.J. and Bury, M.R., Impairment, disability and handicap in chronic respiratory illness. *Social Science and Medicine* 29, 609, 1989.
34. Jones, P.W. Quality of life measurement in chronic lung disease. *J Drug Ther Res* 13, 181, 1988.
35. McSweeney, A.J., Heaton, R.K., Grant, I., Cugell, D., Solliday, N. and Timms, R. Chronic obstructive pulmonary disease; socio-emotional adjustment and life quality. *Chest* 77, 309, 1980.
36. Guyatt, G.H., Berman, L.B., Townsend, M., Pugsley, S. and Chambers, L.W. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 42, 773, 1987.
37. Kaptein, A.A., Dekker, F.W., Dekhuijzen, P.N.R., Wagenaar, J.P.M. and Janssen, P.J. The value of the 12-minutes walking test in patients with chronic obstructive pulmonary disease, II walking in the laboratory and walking in daily life (in Dutch). *Nederlands Tijdschrift voor Geneeskunde* 131, 1717, 1987.
38. Schrier, A.C., Dekker, F., Kaptein, A.A. and Dijkman, J.H. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 89, 894, 1990.
39. Schayck, C.P. van, Rutten, M.P.M.H., Doorslaer, E.K.A. van, Folgering, H. and Weel, C. van. Two-year bronchodilator treatment in patients with mild airflow obstruction: contradictory effects on lung function and quality of life. (submitted).
40. Guyatt, G.H., Townsend M., Berman, L.B. and Pugsley, S.O. Quality of life in patients with chronic airflow limitation. *British Journal for Diseases of the Chest* 81, 45, 1987.
41. Kaplan, R.M., Atkins, C.J. and Timms, R. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. *Journal of Chronic Disease* 37, 85, 1984.
42. Kaplan, R.M., Bush, J.W. and Berry C.C. Health Status: Types of Validity and the Index of Well-being. *Health Services Research* 11, 478, 1976.
43. Donaldson, C., Atkinson, A., Bond, J. and Wright, K. Should Qalys be programme-specific? *Journal of Health Economics* 7, 239, 1988
44. Drummond, M.F., Ludbrook, A., Lowson, K. and Steele, A. *Studies in Economic Appraisal in Health Care*. Volume 2. Oxford University Press, Oxford, 1986.
45. Martens, L.L.M., Doorslaer, E.K.A. van. Dealing with discounting. *International Journal of Technology Assessment in Health Care* 6, 139, 1990
46. Johanneson, M., Levin, L.A. Comments on "Dealing with discounting". (Letter to the editor) *International Journal of Technology Assessment in Health Care* 7, 649, 1991
47. Doorslaer, E.K.A. van, Martens, L.L.M. On the discounting of gains of life expectancy. Reply to Johanneson and Levin, *International Journal of Technology Assessment in Health Care* 7, 653, 1991
48. Drummond, M.F. *Principles of Economic Appraisal in Health Care*. Oxford University Press, Oxford, 1980.
49. Drummond, M.F. and Stoddart, G.L. Economic Analysis and Clinical Trials. *Controlled Clinical Trials* 5, 115, 1984.
50. Rutten, F.F.H., Greep, J.M. and Haan, G.H.M.G. Public Control of the Diffusion of Health Technology. In *The Economics of Medical Technology. Proceedings of an International Conference on Economics of Medical Technology*. (Edited by Rutten, F.F.H. et al.), Springer-Verlag, Berlin & Heidelberg, 1988.

3

Cost-effectiveness of inhaled corticosteroid plus bronchodilator therapy versus bronchodilator monotherapy in children with asthma

*Maureen P.M.H. Rutten-van Mölken¹, Eddy K.A. van Doorslaer²,
Margreet C.C. Jansen¹, Elisabeth E. Van Essen-Zandvliet³, Frans F.H. Rutten².*

1 Department of Health Economics, University of Limburg.

2 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

3 Department of Paediatrics, subdivision Paediatric Respiratory Medicine, University Hospital Rotterdam/Sophia Childrens Hospital.

Originally published as: Cost-effectiveness of inhaled corticosteroid plus bronchodilator therapy versus bronchodilator monotherapy in children with asthma. PharmacoEconomics 1993; 4:257-270. Reprinted with permission of the publisher.

3.1 Abstract

In an incremental cost-effectiveness analysis, combined inhaled β_2 -agonist plus inhaled corticosteroid therapy (BA+CS) was compared with inhaled β_2 -agonist plus placebo (BA+PL) in 116 asthmatic children aged 7-16 years. Clinical data have been reported previously. To account for the selective withdrawal rate due to pulmonary problems that occurred in the group receiving BA+PL, costs were calculated using 2 approaches: 1) the cumulative cost approach and 2) the patient-year approach.

Besides improvements in forced expiratory volume in 1 second (FEV_1) and airway responsiveness expressed as the provocative dose of histamine required to give a 20% fall in FEV_1 (PD_{20}), the frequency of asthma symptoms and school absenteeism were significantly reduced in the BA+CS group. Annual drug acquisition costs for the group receiving BA+CS were NLG480 higher than for the BA+PL (\$US1 = NLG2.12, 1989 prices). Based on conservative calculations using the cumulative cost approach, annual savings due to reduced health care utilization, excluding the costs of study drugs, by the group receiving BA+CS compared with BA+PL were estimated to be about NLG273 per patient. The incremental cost-effectiveness of BA+CS was estimated to be about NLG175 per 10% increase in FEV_1 , or somewhat less than NLG10 per symptom-free day gained. The patient-year approach estimated savings due to corticosteroids of about 43% of the costs of BA+PL (95% confidence intervals, 21 to 58%). Savings were larger when the indirect costs that a family incurred during school absenteeism were considered.

Addition of an inhaled corticosteroid to an inhaled β_2 -agonist is a cost-effective treatment option that might even result in net healthcare savings.

3.2 Introduction

Childhood asthma can be a costly disease that imposes a particular burden on families and communities each year. The prevalence of asthma among children in The Netherlands is estimated to range from 9 to 14%.¹ Among the chronic childhood diseases, asthma is the leading cause of school absenteeism.² The use of health care services by children with this disease can be extensive. In the US, nearly half of the emergency room visits for asthma and one-third of both the annual hospitalizations for asthma and the number of asthma-related doctor's office visits for ambulatory care were for children under 18 years of age.³ The annual mean number of general practitioner contacts in the Netherlands is estimated to be 5.9 per asthmatic child aged 0 to 19 years compared to 2.4 per non-asthmatic child within the same age category. The number of hospitalizations is estimated to be 26 per 1000 asthmatic children under 19 years of age.¹ In some countries the

percentage of hospitalizations for asthma among children has increased during the past decade.⁵⁻⁷ Such an increase was not observed in The Netherlands, except for children aged 0 to 4 years.⁸

Although there is no complete cure for asthma yet, the medical treatment of asthma has improved considerably, such that appropriate medical regimens should provide both prevention and appropriate control of symptoms. Children with asthma commonly receive health education as well as drug therapy. Pharmacotherapy often consists of maintenance treatment with a combination of drugs for a prolonged period of time as well as initiation of additional drug therapy by the patient or his or her parent or guardian.

Considering the burden that childhood asthma poses on families and society, and the increasing need to set priorities for resource allocation decisions on the basis of efficiency considerations, one would expect an increasing interest in the cost-effectiveness of diagnostic and therapeutic interventions in asthma. Between 1980 and 1991 several cost-effectiveness studies in this area have been published, but only 3 of them involved children and not a single one investigated the cost-effectiveness of anti-asthma drugs.⁹ Yet such studies would be relevant, since new drugs are continuously added to the large number of already available, mostly fairly effective drugs, often without evidence of additional benefit and often without encouragement of cost-reducing therapeutic substitution.

In The Netherlands β_2 -agonists are the most common drugs used to treat childhood asthma, because they provide relatively rapid relief of asthma symptoms. The number of prescriptions for these drugs in children under 12 years of age has increased considerably between 1980 and 1988 and is now almost 120 per 1000.⁸ More recently, there have been a few studies indicating that the long term or continuous use of bronchodilator drugs might be associated with adverse effects on airway obstruction and hyperresponsiveness in adults or even with excess mortality.¹⁰⁻¹³ However, these findings are still under debate. To control the inflammatory process in asthma, which results in an increase of symptoms and bronchial hyperresponsiveness, recent treatment guidelines for moderate to severe asthma recommend either inhaled corticosteroids, sodium cromoglycate or nedocromil sodium.¹⁴ Sodium cromoglycate is recommended as primary treatment for childhood asthma. However, evidence of the efficacy of cromoglycate is conflicting.¹⁵ It has been shown to inhibit the early and late allergic response to a histamine challenge,¹⁶⁻¹⁷ but the effect on bronchial hyperresponsiveness is still subject to debate.¹⁷⁻¹⁹ The clinical study to which this cost-effectiveness analysis is linked is the first randomised controlled study to establish the efficacy of adding inhaled corticosteroid therapy to β_2 -agonist therapy in asthmatic children. In a recent review of 12 uncontrolled studies in which inhaled corticosteroids were given to asthmatic children for at least 12 months, Selroos concluded that pulmonary function was good, symptoms were well controlled and growth rate was not affected.²⁰ Moreover, a number of studies in asthmatic children have shown that inhaled corticosteroids decreased bronchial hyperresponsiveness.²¹⁻²³ The maximum

follow-up in these studies was 6 months. Previous studies in adults have shown that corticosteroids achieve short-term and long-term improvements in lung function, bronchial hyperresponsiveness, symptoms and exacerbations of asthmatic patients.^{17,24-29}

This article reports an investigation into the cost-effectiveness of combined inhaled bronchodilator and anti-inflammatory therapy consisting of a β_2 -agonist plus an inhaled corticosteroid (BA+CS) compared with an inhaled bronchodilator monotherapy consisting of a β_2 -agonist combined with a placebo (BA+PL). Whether BA+CS in asthmatic children results in additional long-term benefits when compared to BA+PL and whether these benefits outweigh or justify the additional costs are the questions which are being addressed in this study. Special attention was paid to statistical testing of cost differentials in order to meet the criticism that cost data are often more subjective and more speculative than clinical data.³⁰

3.3 Methods

The cost-effectiveness analysis was performed alongside a randomised double-blind multicentre trial which is described in detail elsewhere.³¹ A concise summary is given here. Between October 1987 and April 1989, 116 asthmatic children aged between 7 and 16 years were randomised into the study. These individuals had to demonstrate airway obstruction, reversibility of airway obstruction in response to inhaled salbutamol (albuterol), and hyperresponsiveness to a histamine challenge. Airway obstruction was defined as an FEV₁ (Forced Expiratory Volume in one second) of 55 to 90% of that predicted on the basis of gender, age and height³² and/or a ratio of FEV₁ to FVC (Forced Vital Capacity) of 50 to 75%. Reversibility was defined as an increase in FEV₁ of $\geq 15\%$ of baseline after 0.2 mg of inhaled salbutamol. Hyperresponsiveness was measured as the provocative dose of histamine required to cause a 20% fall in FEV₁ (PD₂₀). The children were enrolled from the outpatient clinics of 2 university children's hospitals and one general children's hospital. Candidates were excluded if they had serious concomitant disease, required oral corticosteroids or were known to depend on inhaled corticosteroids to control their asthma.

Immediately after a second baseline assessment, patients were randomly allocated into an experimental group BA+CS receiving an inhaled β_2 -agonist (salbutamol 0.2 mg 2 puffs thrice daily) combined with an inhaled corticosteroid (budesonide 0.2 mg 2 puffs thrice daily), and a control group BA+PL receiving the same inhaled β_2 -agonist combined with an inhaled placebo. Randomization was stratified by gender, age, treatment centre and prior use of inhaled corticosteroids and/or cromoglycate. Inhaled corticosteroids and cromoglycate were tapered off and completely stopped 2 weeks before entry into the study. No concomitant medication was allowed, except for rescue therapy for children with increasing pulmonary

complaints. This comprised additional inhaled fenoterol with a maximum of 4 daily doses of 0.2 mg, a course of oral prednisolone (30 mg on the first day, tapering off to zero in one week, according to bodyweight) in the event of an exacerbation of asthma symptoms and a course of antibiotics in case of a bacterial infection.

Patients were seen at the outpatient clinic every 2 months for monitoring, data collection and distribution of new drug canisters during a planned follow-up of 3 years. They were maintained on study drugs, unless they withdrew from the study on their own initiative or were withdrawn because of a lack of treatment efficacy or non-compliance with the protocol, or because they required more than 2 courses of prednisolone over a 3-months period or more than 4 courses of prednisolone over a year.

3.3.1 Clinical endpoints

The primary clinical endpoints were FEV₁ expressed as a percentage of the predicted value (FEV₁% predicted) as an indicator of airway obstruction and PD₂₀ as an indicator of hyperresponsiveness. FEV₁% predicted was measured at baseline and every 2 months during follow-up. PD₂₀ was measured by means of a histamine provocation test at baseline and every 4 months after randomization. Both measures were performed only when the patient's asthma was clinically stable and not within 4 weeks after termination of a course of prednisolone. All study medication was discontinued 8 hours before these tests.

Important additional measures of effectiveness that were incorporated in the cost-effectiveness analysis were the number of days free of symptoms and the number of days of absence from school. Two weeks prior to each outpatient visit, patients recorded on a diary card the severity of symptoms on a 4-point scale. This information was then linearly extrapolated to a period of 2 months. A symptom-free day was defined as a day for which no symptoms were reported. School absenteeism was measured by means of a diary notebook, which was used during the whole study period. In this book the children registered absence from school and illness periods, under parental supervision.

3.3.2 Economic endpoints

The aims of this incremental cost-effectiveness analysis were to assess the difference in costs and effects between the treatment groups, and to determine incremental cost-effectiveness ratios. Thus, the costs of regular outpatient visits and lung function tests which were scheduled every 2 months for both the experimental and the control group were excluded. Also excluded were the costs of special visits and tests that were required for the study protocol only and not for patient care. Both direct costs of health care utilization and indirect costs of parents' production losses were tabulated.

Two types of direct costs relevant to the evaluation of the 2 pharmacotherapies were considered in this study: 1) the costs of providing the treatment, that is the acquisition costs of the study drugs and 2) the economic consequences of using the study drugs in terms of changes in health care utilization. The second cost category included: a) costs of oral prednisolone rescue therapy, b) costs of inhaled fenoterol rescue therapy, c) costs of antibiotics, d) costs of additional asthma-related drug use other than the allowed rescue therapy and antibiotics (e.g. mucolytics, antihistamines, theophyllines, homeopathic drugs, drugs against common cold and cough), e) costs of additional non-scheduled outpatient visits, tests and telephone contacts, f) costs of hospitalizations.

Information on the use of oral prednisolone, antibiotics and additional non-rescue drugs was collected every 2 months during the outpatient visits by means of a standardised interview of the patient by the pulmonary physician. Information on additional outpatient visits, telephone contacts, hospitalizations or potential additional services was obtained from the patients' records. We included hospitalizations, additional outpatient contacts and services due to asthma alone and not due to other illnesses. The patients recorded the use of additional fenoterol on a diary card during the 2 weeks before each visit. This information was extrapolated to a period of 2 months. The costs of drugs were based on 1989 retail prices including pharmacy delivery costs. Other direct costs were calculated from a third party perspective, using charges as proxy for costs¹⁾. All costs are reported in Dutch Guilders (\$US1 = NLG2.12 at 1989 mean exchange rate).

Indirect costs refer to the value of the production lost to society due to illness.³³ We considered the indirect costs that a family incurs during school absenteeism. The number of school days a child was absent from school was used to represent the number of days lost from work or from housekeeping by one of the parents. Indirect costs were calculated using the human capital approach.³⁴ The production loss of mothers performing unpaid labour was explicitly included and

1) Charges for outpatient visits, telephone contacts and lung function tests do not differ between hospitals in the Netherlands. The costs of hospitalizations were calculated by averaging the charges per hospital day at the participating centres. This average cost per day was multiplied by the length of stay to obtain the costs per hospitalization.

the 1989 average hourly wage was used to estimate the value of the production loss for each individual^{2), 35}

In this article we present school absenteeism both as an effect parameter in terms of the number of days of school absenteeism prevented and as a cost parameter in terms of the indirect costs of a parent's production loss. To avoid double counting we did not include indirect costs in our final calculation of cost-effectiveness.

3.3.3 Statistical analysis

Differences in withdrawal rates between therapeutic groups were analysed using the logrank test and differences in proportions using z-tests.³⁶ Differences in FEV₁% predicted, PD₂₀, days of absence from school and symptom free days were determined using unpaired Student's t-tests. All PD₂₀ calculations were performed using the base-2 logarithm of PD₂₀ as this reflects doubling of dosages.

Anticipating the selective withdrawal in the BA+PL group, which is reported in the results section, we have used two different approaches to *calculate* costs: 1) *the cumulative cost approach* and 2) *the patient-year approach*. In the first approach, average costs per patient were calculated for each 2-month period. Since the number of patients in the study declined, these averages were based on different numbers of patients. The averages were then cumulated to calculate the average costs per patient during the entire study. The assumption underlying this approach is that the (non-observed) average costs per 2-month period of patients after their withdrawal from the study were the same as the (observed) average costs per 2 months of the patients in the respective treatment groups who did not withdraw³⁾. The cumulative cost approach.

Since the costs of patients who withdrew might have been higher than the average in their group if they had continued, we also applied the patient-year approach. In this approach, the costs per patient-year are calculated assuming that the (non-observed) costs of withdrawals equalled the average (observed) costs they

-
- 2) Production loss was valued by 42,5% of the average net hourly income of the Dutch working population in 1989. This percentage is based on the advice of the Dutch Committee for the Development of Policy Analyses who recommended to value lost working time at 50% of the real average income and lost private time at 35% of the real average income.
 - 3) Suppose that the costs of 58 patients in the BA+PL group during the first 2 months were on average 132 pp. 11 of the 58 patients drop out within the next two months and their costs during these months are unknown. Further suppose that the costs of the remaining 47 patients who did not drop out during the second 2 monthly period were 66 pp. Thus the cumulative mean costs per patient in the BA+PL group are estimated to be 198 (132+66) for each of the 58 patients in the BA+PL group. The same approach is applied to the BA+CS group.

had before they withdrew⁴⁾. In other words, the observed costs before withdrawal were linearly extrapolated to a 1-year period, resulting in a higher estimate of the cost-differences between the treatment groups than was produced by the cumulative cost approach, because of the selective withdrawal process. The patient-year approach resulted in an estimate of the lower, more favourable limit of the cost-effectiveness, whereas the cumulative cost approach resulted in a less favourable estimate of the cost-effectiveness.

A distinction was made between calculating (as explained above) and *testing* cost differences between the treatment groups. The cumulative cost approach does not allow any testing of cost-differences for statistical significance, because it gives only group averages, not individual observations. In contrast, the patient-year approach does allow statistical testing of cost-differences, which was performed using analysis of variance on logarithmically transformed costs. Costs were logarithmically transformed in order to reduce the skewness of the population data (caused by a few patients whose healthcare costs were very high). First, differences between treatment groups in total direct health care costs were examined. Further analyses examined differences in several categories of costs per patient-year in order to identify components of costs which partially accounted for the difference in total direct healthcare cost. Finally, an ordinary least squares regression analysis was performed to test the difference in transformed costs between treatment groups, thereby adjusting for the effects of various baseline patient characteristics on the logarithmically transformed mean costs per patient-year.

3.4 Results

3.4.1 Baseline characteristics

One hundred and sixteen patients (74% male) with a mean age of 10.9 years entered the study. The mean FEV₁ was 77.8% of that predicted for gender, age and height. The mean log₂PD₂₀ was 4.24 µg. No statistically significant differences between the treatment groups in baseline FEV₁ and log₂PD₂₀ were found. Also, the 2 groups were similar with regard to age, gender, prior use of inhaled corticosteroids or inhaled cromoglycate and number of days with symptoms. More details on baseline characteristics can be found in Van Essen-Zandvliet *et al.* 1992.

4) Suppose that the costs of a single patient during the first two months of study are 535. Within the next two months this patient drops out and his costs during these months are unknown. Thus the cost per patient-year for this patient is estimated to be $12/2 \times 535 = 3210$.

3.4.2 Withdrawals

Figure 1 shows the number of withdrawals in the BA+PL group to accumulate over the study period, starting immediately after randomization. Among the control group only 32 of the 58 patients completed the study. 26 patients (45%) dropped out, 24 because of pulmonary problems. In contrast, among the experimental group only 3 of the 58 patients withdrew (5%), one of them because of an increase in asthma symptoms and the other 2 because of poor motivation. This difference in withdrawal rates was highly significant ($p < 0.0001$).³¹ Withdrawals in the BA+PL group tended to have poorer airway calibre, greater hyperresponsiveness and more frequent symptoms. They were also more likely to be prior users of inhaled corticosteroids and less likely to be prior users of inhaled cromoglycate. Other baseline characteristics did not differ significantly between patients who withdrew from the study and those who did not withdraw.³¹

Besides the difference in withdrawal rates, interim analyses, which were undertaken at 6-monthly intervals by unblinded but independent scientists, indicated such significant differences in terms of the primary clinical endpoints that it was considered unethical to continue the trial. It was therefore terminated earlier than originally planned. The length of follow-up on randomised treatment ranged from 10 to 28 months (median 22 months). All of the 87 patients who did not withdraw from the study reached a follow-up of at least 10 months, 39 of them reached a follow-up of 22 months and 1 reached a follow-up of 28 months. Our analyses are based on information up until withdrawal. Thus, as follow-up time increases, the patients available for follow-up in the BA+PL group became an increasingly less representative sample of the original participants. All withdrawals receiving BA+PL were given BA+CS after their withdrawal, but they were no longer followed.

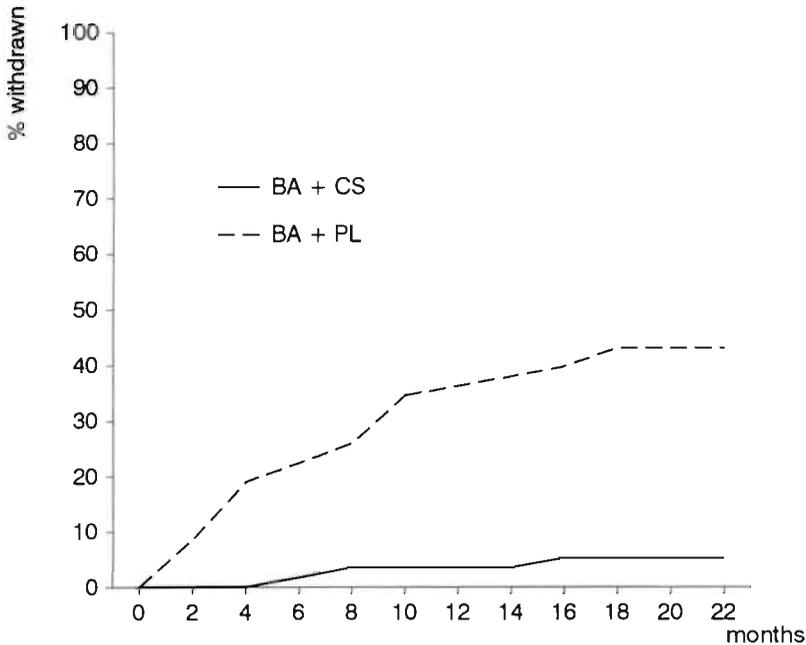


Figure 1. Cumulative rates of withdrawal from therapy with inhaled β_2 -agonist plus inhaled corticosteroid therapy (BA+CS) compared with regular inhaled β_2 -agonist plus placebo therapy (BA+PL)

3.4.3 Effectiveness

A description and discussion of the physiological results is given by Van Essen-Zandvliet *et al.* 1992.³¹ It suffices to note here that both FEV₁ and PD₂₀ show significantly more improvement in patients receiving BA+CS than in patients receiving BA+PL. The mean difference in changes in FEV₁% predicted from baseline between these groups was 11.85% during the first year ($p < 0.001$). This difference was reached during the first 4 months of follow-up and it increased slightly throughout the study. The mean difference between BA+CS and BA+PL in the change in PD₂₀ from baseline was 1.53 doubling dose steps in the first year ($p < 0.0001$). This difference increased to a mean of 2.02 dose steps during the second year.

Table 1 shows that the mean asthma-related absence from school was 3.7 days per year in the BA+PL group, whereas the patients in the BA+CS group were absent an average of 1.7 days per year ($p < 0.05$). The table also shows the number

of symptom free days per patient-year which was 212 in the BA+CS group compared with 174 in the BA+PL group ($p < 0.0001$).

Table 1. Number of days of absence from school due to asthma and number of symptom-free days per patient-year by treatment group

	BA+CS	BA+PL
no. patients at baseline	58	58
no. patients at 22 months follow-up	25	14
<i>Schoolabsenteeism</i>		
no. patients with ≥ 1 day of school absenteeism	17	25
mean (se) no. of days of school absenteeism per patient-year	1.7 (0.5)*	3.7 (0.9)
<i>Symptom free days</i>		
no. patients with ≥ 1 symptom free day	57	49
mean (se) no. symptom free days per patient-year	211.7 (2.9)**	173.5 (4.5)

Abbreviations: BA+CS = β_2 -agonist + corticosteroid; BA+PL = β_2 -agonist + placebo

* $p < 0.05$

** $p < 0.01$

3.4.4 Costs

Drug acquisitions costs for BA+CS and BA+PL were NLG 648.26 and NLG 167.79 per patient per year, a difference of NLG 480.47. Figure 2 shows the cumulative direct costs of health care utilization per patient based on the *cumulative cost approach*, including hospitalizations but excluding the acquisition costs of the study drugs. The area between the two lines indicates the savings due to inhaled corticosteroid therapy. When therapy with BA+CS was compared with therapy with BA+PL, savings in terms of lower healthcare utilization were estimated to be about NLG 273.42 per patient after the first year of follow-up.

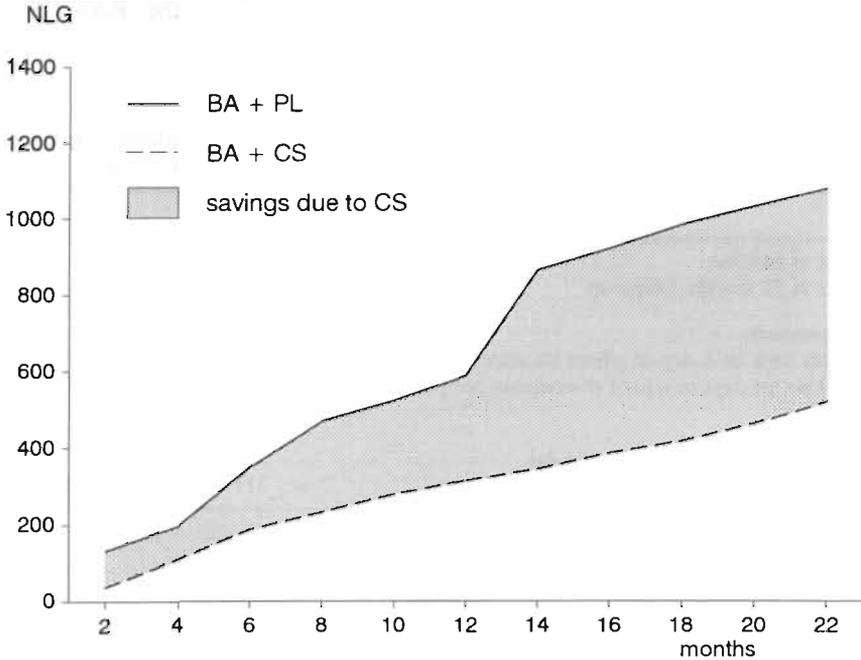


Figure 2. Cumulative direct health care costs of asthma for patients receiving BA+CS compared with BA+PL, including hospitalizations, but excluding the acquisition costs of the study drugs.

On the basis of the *patient-year approach*, the annualized direct costs per patient excluding the study drug acquisition costs were estimated to be NLG 289 in the BA+CS group compared with NLG 959 in the BA+PL group, which represents a saving of about 70% of the costs of BA+PL. This difference was found to be highly significant ($p < 0.001$). At each follow-up measurement, the direct costs per patient were significantly lower in the BA+CS group than in the BA+PL group ($p < 0.05$). At each follow-up measurement, the average costs incurred by future patients withdrawing were significantly higher than the average costs of the patients not withdrawing ($p < 0.005$). The costs of the former were generally highest in the 2 months immediately preceding withdrawal.

The various components of the direct costs as measured by the patient-year approach are given in table 2. Although not all subcategories of cost differences reached conventional levels of significance, they all differed in the same direction as the total direct costs. The costs of the prednisolone and fenoterol rescue therapy and the costs of additional outpatient contacts are significantly lower in the BA+CS group than in the BA+PL group. Despite the fact that the difference in the cost of hospitalizations did not reach the level of statistical significance, it contributes

considerably to the total healthcare cost difference. During the entire study 2 patients receiving BA+CS were hospitalised (NLG 4564), whereas 5 hospitalizations occurred in the BA+PL group (NLG 17 356)⁵⁾. One patient in the control group was hospitalised twice.

The estimated indirect costs in terms of the production loss of the parents due to the asthma-related school absenteeism of the children amount to about NLG 78 per patient-year in the BA+CS group compared with NLG 173 per patient-year in the BA+PL group.

Table 2. Mean (se) direct medical costs per patient-year, excluding acquisition costs of study drugs, by cost category and treatment group (1989 Dutch guilders)

cost category	BA+CS	BA+PL
prednisolone rescue therapy	2.04 (0.74)**	12.06 (2.72)
fenoterol rescue therapy	37.99 (6.38)**	84.89 (12.68)
antibiotics rescue therapy	5.02 (1.57)	5.88 (2.21)
other additional drug use	164.22 (6.95)	177.60 (10.96)
additional outpatient contacts and services	27.09 (6.78)*	86.08 (20.32)
hospitalizations	52.47 (37.21)	592.00 (315.37)
total direct health care costs	288.83 (43.41)**	958.52 (321.61)
indirect costs	78.45 (22.20)*	173.15 (39.53)

Abbreviations: BA+CS = β_2 -agonist + corticosteroid; BA+PL = β_2 -agonist + placebo

* p < 0.05

** p < 0.01

Results of the ordinary least squares regression analysis with logarithmically transformed costs per patient-year (excluding study drug acquisition costs) as dependent variable are shown in table 3. Other things being equal, patients receiving BA+CS had significantly lower health care costs than those receiving BA+PL, and patients with higher levels of FEV₁ had somewhat lower healthcare costs than patients with lower levels of FEV₁. For the total sample, this model

5) In the article by Van Essen-Zandvliet *et al.* only 3 hospitalizations were reported, all three occurred in the BA+PL group. These figures were based on the study diaries of the patients, which were checked by the physicians every two months. The figures in our article were based on both these diaries and a review of patient records. Hospitalizations were substantially underreported in the diaries.

implies that the costs of the patients in the BA+CS group are about 43% ($(e^{-0.556}-1) \times 100\%$) lower than the costs of the patients receiving BA+PL. This percentage, which has a 95% confidence interval ranging from 21 to 58% is different from the 70% reported earlier. This is a result of both the logarithmic transformation, which reduces the impact of outlier values (especially hospitalization costs), and the correction for confounding by differences in patient characteristics that occurred as a result of the selective withdrawal prior to the first follow-up visit.

Table 3. Effects of independent variables on direct medical costs per patient-year[†] (linear regression model)

explanatory variables [‡]	total sample		subsample of non-withdrawals	
	regression coefficient [¥] (se)	p-value	regression coefficient [¥] (se)	p-value
study drugs	-0.556 (0.156)	0.001	-0.292 (0.130)	0.027
gender	0.005 (0.183)	0.980	-0.078 (0.142)	0.587
smoking parent(s)	-0.137 (0.157)	0.385	-0.175 (0.123)	0.160
prior use of corticosteroids	0.164 (0.186)	0.381	0.118 (0.142)	0.410
prior use of cromoglycate	-0.182 (0.163)	0.267	0.068 (0.128)	0.596
age	-0.067 (0.040)	0.095	-0.023 (0.031)	0.451
FEV ₁	-0.016 (0.008)	0.052	-0.009 (0.006)	0.142
log ₂ PD ₂₀	0.028 (0.062)	0.657	-0.018 (0.050)	0.726
intercept	8.029 (0.888)	<0.001	6.854 (0.685)	<0.001
adj R ²	0.099		0.030	
F-value [§]	2.509	0.0157	1.338	0.2376

Abbreviations: BA+CS = β_2 -agonist + corticosteroid; BA+PL = β_2 -agonist + placebo; FEV₁ = forced expiratory volume in 1 second; PD₂₀ = provocative dose of histamine required to cause a 20% drop in FEV₁; se = standard error of the regression coefficient

† dependent variable: $\ln(\text{costs per patient-year} + 1)$

‡ explanatory variables at baseline:

study drugs: 1 = BA + CS, 0 = BA + PL

gender: 1 = male, 0 = female

smoking parent(s): 1 = smoking parent(s), 0 = non-smoking parent(s)

prior use of corticosteroids: 1 = prior use, 0 = no prior use

prior use of cromoglycate: 1 = prior use, 0 = no prior use

¥ slope of linear regression of costs per patient-year vs. an independent variable, controlled for other independent variables

§ Mean square regression divided by mean square residual

3.5 Incremental cost-effectiveness

The incremental cost-effectiveness of combined BA+CS therapy was calculated as the quotient of the net incremental direct costs (that is the additional study drug acquisition costs minus the savings in all other direct health care costs) and the net improvement in each of the effect parameters. Figure 3 shows the incremental cost-effectiveness ratios of combined bronchodilator and corticosteroid treatment, using the cumulative cost approach to calculate the net incremental costs. The calculations of net incremental direct costs were based on data obtained for the first year of follow-up, to limit systematic underestimation of the effects and savings due to corticosteroids, resulting from withdrawal of patients from the study.

Additional annual drug acquisition costs of BA+CS compared with BA+PL amounted to NLG 480 per patient, whereas the cumulative savings in other direct health care costs were estimated to be NLG 273 per patient during the first year. The FEV₁ increased 11.85% more in patients receiving BA+CS than in patients receiving BA+PL. Thus, the cost-effectiveness ratio for this parameter was estimated to be about NLG 175 per 10% increase in FEV₁.

The costs per 2 doubling doses increase in PD₂₀ amounted to about NLG 271 per year. The costs per day of school absenteeism prevented were about NLG 92 per patient per year and the costs per symptom-free day gained were less than NLG 10 per patient per year. In other words, in terms of the *cumulative cost approach*, the health care savings did not outweigh the higher costs of this treatment, but the incremental health care costs are relatively low.

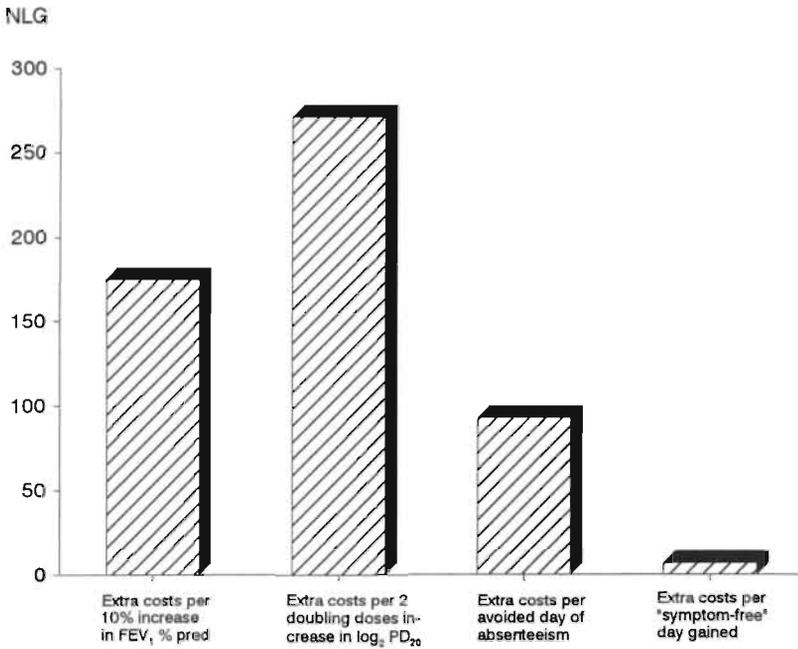


Figure 3. Incremental cost-effectiveness ratios for treatment with BA+CS compared with BA+PL based on the cumulative cost approach. Study drug acquisition costs are included but indirect costs excluded

The above mentioned calculations provide a conservative estimate of the incremental cost-effectiveness of BA+CS. Less conservative is the calculation of the cost-effectiveness based on the *patient-year approach*. On the basis of this approach, the annualised savings in direct healthcare costs, excluding study drug acquisition costs, of BA+CS compared with BA+PL were estimated to be NLG 670 per patient. Comparing this figure with the additional annual study drug acquisition costs of NLG 480, net savings of about NLG 190 per patient-year result. Thus, when net incremental direct cost was calculated by means of the *patient-year approach*, the savings even outweighed the extra costs of inhaled corticosteroids by about NLG 190 per patient-year.

3.6 Discussion

Although questions about the generalisability of our results to other countries and other clinical settings can only be answered by replications of our study, our findings suggest that combined inhaled bronchodilator and inhaled corticosteroid therapy for childhood asthma is cost-effective compared with β_2 -agonist monotherapy. Two approaches to calculating average costs per patient were used, the cumulative cost approach and the patient-year approach. The latter resulted in the most favourable estimate of the incremental cost-effectiveness of BA+CS over BA+PL.

Because the mean costs per patient were significantly higher for those who withdrew from the trial than for those who did not withdraw, the large withdrawal rate in the BA+PL group might cause a systematic underestimation of costs for this group when the cumulative cost approach is used. In this approach, mean costs of patients who withdrew from the trial were assumed to equal the mean costs of non-withdrawals. As a result of selective withdrawal, only the patients whose asthma was stable and less severe, and for whom healthcare utilization was low, remained in the bronchodilator treatment group. This led to an underestimation of the savings (and clinical effects) attributable to corticosteroids. Since the costs in the BA+PL group might have been higher if the withdrawals had continued, we also calculated the savings by means of the patient-year approach. In this approach, the relatively high costs for patients who withdrew during the period before their withdrawal were extrapolated to subsequent periods.

On the basis of the more conservative cumulative cost approach, combined inhaled bronchodilator plus inhaled corticosteroid therapy was found to be cost-effective compared with BA+PL, with ratios ranging from about NLG 271 per 2 doubling doses increase in PD₂₀, to NLG 175 per 10% increase in FEV₁ to less than NLG 10 per symptom-free day gained.

With the patient-year approach, the savings in health care costs exceeded the annual additional costs of inhaled corticosteroids by about NLG 190 per patient-year. For every additional NLG spent, NLG 1.40 was realized in healthcare cost savings.

Most previous cost-effectiveness studies have been restricted to straightforward calculations of costs as presented above. In our study special attention was paid to statistical testing of the differences in costs. Only the patient-year approach enabled such a statistical analysis. Since the mean healthcare costs per patient-year depend largely on a small number of patients who have been hospitalised and whose health care costs are correspondingly high, the standard errors associated with these costs are very large and the total cost distribution becomes very skewed (in the BA+CS group 2 hospitalizations occurred compared to 5 in the BA+PL group). For this reason the annualised costs per patient were logarithmically transformed in order to reduce skewness and thus to allow statistical testing with greater precision. As a result of this transformation, the impact of hospitalization costs on the difference in

total costs was reduced, resulting in an estimated cost reduction due to BA+CS of about 43% of the costs of BA+PL, with a 95% confidence interval ranging from a 21% reduction to a 58% reduction.

These analyses lead us to conclude that, in children with moderate asthma, BA+CS therapy has additional benefits that are worth the relatively small additional costs, and that may be accompanied by net savings.

The total savings to society are much greater if the indirect costs that a family incurs during school absenteeism are considered. In this study the estimation of savings in terms of potential productivity losses was based on the significantly lower number of days of absence from school in the experimental treatment group. It was assumed that in case of a child's illness, one of the parents would not be able to go to work or to do his or her usual activities. It was also assumed that a certain percentage of the average wage was a useful estimate of production losses. Real labour costs might have been a better approximation, but would have required an extensive separate study that went beyond the scope of this trial. Since there is considerable debate over the contribution of indirect cost estimates to the economic evaluation of health care programmes, the indirect costs were not included in the cost-effectiveness ratios.

Theoretically, it is possible that the total costs after withdrawal of the 3 patients who withdrew from therapy with BA+CS were higher than the total costs after withdrawal of the 26 patients who withdrew from therapy with BA+PL, for example because of adverse events that require expensive treatment or because of serious pulmonary problems that require frequent hospitalization. However, this seems very unlikely, since all adverse events were minor (e.g. tremble, headache, nausea, cough) and were never a reason for terminating randomised treatment. Moreover, only one of the 3 withdrawals in the BA+CS group withdrew because of an increase in asthma symptoms compared to 24 of the 26 in the BA+PL group.

All health care utilization costs were based on charges. Charges have the disadvantage that they often do not reflect actual medical resource use. Despite this, taking charges to estimate direct costs does not bias the differences in direct costs between the two treatment groups systematically, although the magnitude of the differences in these costs may be misstated. The marginal savings from the reduced number of hospitalization days may be lower than those estimated using charges, since resource use is typically lower in the later part of a patient's hospital stay and asthma or chronic obstructive pulmonary disease related hospitalization days may be less resource intensive than for example hospital days following major surgery.

This cost-effectiveness study did not pay attention to costs borne by the parents of these children. Practical limitations of a 2-year study made this impossible. For example, the use of over-the-counter drugs might have been substantial in the patients for whom bronchodilator monotherapy was prescribed. However, the costs of these drugs would be relatively small compared with the costs of prescribed drugs, outpatient visits and hospitalizations. Also, in children with a recent onset of the disease the costs of modifications to the home

environment may be substantial. However, these costs are related to having asthma, independent of the drug therapy that is prescribed.

Our study used clinical disease-specific outcome measures. Thus, the cost-effectiveness ratios obtained cannot be readily compared with those obtained for interventions other than those for childhood asthma. It would have been relevant to take into account various aspects of the quality of life of these children, including a valuation of changes in these aspects. In practice, it is very difficult to measure quality of life in children. For example, it is not clear who should make the judgements about a child's quality of life and which dimensions of quality of life should be measured in children; for example whether developmental, behavioral or learning capacity are more relevant dimensions than functional capacity. Moreover, very few quality of life instruments for children are available.

However, the improved lung function and hyperresponsiveness, the lower incidence of school absenteeism and the significantly higher number of symptom free days per patient-year in the BA+CS group suggest that at least some improvement in overall quality of life was reached. If the improvements achieved in these 4 outcome measures, led to only a very small increase on a utility scale, BA+CS therapy would still be very cost-effective. If we conservatively assume a 0.05 increase in quality of life measured on a 0 (death) to 1 (perfect health) scale (utility), then the costs per Quality Adjusted Life Year (QALY) gained of combined β_2 -agonist plus corticosteroid therapy would be somewhat more than NLG 4000 (\$US 1887) per child per year. If we assume an even smaller increase of 0.01, then the costs per QALY gained would be about NLG 21000 (\$US 9905). No comparable figures of costs per QALY gained in asthmatic children are available, but the above mentioned figures compare favourably with the costs per QALY of other evaluated interventions in Dutch health care.³⁷ The only comparable figure in the area of chronic airways disease is that reported by Tovey *et al.* in 1984, who estimated the costs per QALY of a behavioral intervention program to enhance compliance to an exercise program for adult COPD patients to range between \$US 10834 and \$US 36897 in 1984.

Under 'real world' clinical conditions, the use of BA+CS therapy may yield cost-effectiveness ratios different from the ones measured here. On the one hand, broader indications for BA+CS or reduced compliance might worsen the cost-effectiveness. On the other hand, a small decrease in the acquisition costs of the corticosteroids could easily lead to savings that are even larger than estimated by the patient-year approach. It is also likely that in regular clinical practice, monitoring of patients would be less careful than in a clinical trial situation, as a result of which the withdrawal rate in the bronchodilator treatment group might be lower than in this trial.

Recent studies have suggested that β_2 -agonist therapy given on demand is preferable to regular β_2 -agonist therapy.^{10,11} It is not yet clear whether it is preferable for patients who receive regular continuous inhaled corticosteroid therapy to receive additional β_2 -agonist therapy on demand or regularly. Inhaled

corticosteroid therapy is currently not recommended as treatment of first choice for children with moderate asthma, but sodium cromoglycate is. This recommendation may partly be based on the limited side effects of cromoglycate and the possibility of adverse effects of continuous inhaled corticosteroid therapy such as reduced growth³⁹ or adrenal suppression. Our clinical study and those of others found few adverse effects and no effects on growth. Since both cromoglycate and inhaled corticosteroids are directed at suppressing inflammation it seems necessary that future studies pay attention to the relative cost-effectiveness of inhaled corticosteroids versus cromoglycate, including the costs of possible side effects. It is clear however, that inhaled corticosteroid therapy is a much better option than inhaled bronchodilator monotherapy, which is one of the most commonly used treatments for asthmatic children. This conclusion was reached on the basis of a relatively short term study with a follow-up of 28 months at the most. In principle, it may be possible for inhaled corticosteroid therapy to decrease the loss of lung function by controlling inflammation and thereby altering the natural history of childhood asthma to such an extent that continuous pharmacotherapy is no longer required.

We have shown previously that FEV₁ normalised in 42% of the patients receiving BA+CS compared with 29% of those receiving BA+PL, and PD₂₀ normalised in 26% of the patients receiving BA+CS compared with 6% of the patients receiving BA+PL.³¹ Early intervention with BA+CS may modify the natural history of asthma such that the severity of the disease and the healthcare utilization at older age will be lowered. However, these speculations have not been confirmed by clinical evidence.

To summarise, data from the first cost-effectiveness study of pharmacotherapy in asthmatic children suggest that combined inhaled bronchodilator plus corticosteroid therapy is both clinically efficacious and cost-effective. If this combination therapy does indeed rank among the few interventions that have been shown to create additional benefits accompanied by net health care savings, there should be little doubt that this combination pharmacotherapy represents a wise use of scarce resource.

3.7 Acknowledgements

The clinical study was supported by grants from the Netherlands' Health Research Promotion Program (SGO), Glaxo, Astra Pharmaceuticals and Boehringer Ingelheim. The cost-effectiveness analysis was supported by grants from the Dutch Asthma Foundation and Glaxo. We wish to thank the Dutch CNSLD group* and all participating centres for providing data and for their support when reviewing the patients' records. We would also like to thank Ben van Hout and René van Vliet for useful suggestions and Jan van Emmerik for computer assistance.

* The Dutch Chronic Non-Specific Lung Disease (CNSLD) study group consists of a steering committee (K.F. Kerrebijn, Ph.H. Quanjer and H.J. Sluiter), of members from the Departments of Pulmonology of the University Hospital of Amsterdam (E.M. Pouw, D.F.M.E. Schoonbrood, C.M. Roos, H.M. Jansen), Groningen (P.L.P. Brand, A. de Gooyer, H.A.M. Kerstjens, D.S. Postma, Th.W. van der Mark, H.J. Sluiter, G.H. Koëter), Leiden (P.M. de Jong, P.J. Sterk, A.M.J. Wever, J.H. Dijkman), Nijmegen (P.N.R. Dekhuijzen, H.T.M. Folgering, C.L.A. van Herwaarden), Rotterdam (S.E. Overbeek, J.M. Bogaard, C. Hilvering) and Utrecht (S.J. Gans, H.J.J. Mengelers, B.A.H.A. van der Bruggen-Bogaarts, J. Kreukniet), from the Departments of Pediatric Pulmonology at the Sophia Children's Hospital, Rotterdam (E.E.M. van Essen-Zandvliet, K.F. Kerrebijn), the Juliana Children's Hospital, the Hague (E.J. Duiverman, J.M. Kouwenberg, J.E. Prinsen), University Hospital of Groningen (H.J. Waalkens, J. Gerritsen, K. Knol), from the Department of Allergy, University Hospital of Groningen (J.G.R. de Monchy), from the Department of General Practice, University of Leiden (A.A. Kaptein, F.W. Dekker), and from the Department of Physiology, University of Leiden (P.J.F.M. Merkus, Ph.H. Quanjer). Scientific counsel: S.J. Pocock, M.D. Hughes, (London, UK), E.R. Bleeker, D.A. Meyers (Baltimore, USA).

3.8 References

1. STG (Stuurgroep Toekomstscenario's Gezondheidszorg). Chronic diseases in the year 2005. Part 2: Scenario's for chronic non-specific lung disease 1990-2005. Bohn Stafleu van Loghum, Houten/Antwerpen, 1990 (in Dutch)
2. Lako CJ. Absenteeism in primary school. *Tijdschrift voor Sociale Gezondheidszorg* 61: 776-783, 1983 (in Dutch)
3. Weiss KB, Gergen PJ, Hodgson A. An economic evaluation of asthma in the United States. *N Eng J Med* 326: 862-866, 1992
4. Anderson HR. Increase in hospital admissions for childhood asthma: trends in referral, severity, and readmissions from 1970 to 1985 in a health region of the United Kingdom. *Thorax* 44: 614-619, 1989
5. Gergen PJ, Weiss KB. Changing patterns of asthma hospitalizations among children: 1979 to 1987. *JAMA* 264: 1688-1692, 1990
6. Jackson RT, Mitchell EA. Trends in hospital admission rates and drug treatment of asthma in New Zealand. *NZ Med J* 96: 727-729, 1983
7. Mitchell EA. International trends in hospital admission rates for asthma. *Arch Dis Child* 60: 376-378, 1985

8. Wever-Hess J, Wever AMJ, Yntema JL. Mortality and morbidity from respiratory diseases in childhood asthma in the Netherlands, 1980-1987. *Eur Respir J* 9: 429-433, 1991
9. Rutten-van Mölken MPMH, Van Doorslaer EKA, Rutten FFH. Economic appraisal of asthma and COPD care: a literature review 1980-1991. *Soc Sci Med* 35: 161-175, 1992
10. Sears MR, Taylor DR, Print G *et al.* Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* 336: 1391-1396, 1990
11. Schayck Van CP, Dompeling E, Herwaarden Van CLA, Folgering H, Verbeek ALM, Hoogen Van der HJM, Weel Van C. Bronchodilator treatment in moderate asthma or chronic bronchitis: continuous or on demand? *Br Med J* 303: 1426-1431, 1991
12. Crane J, Glatt A, Jackson R, Ball M, Pearce N, Burgess C, Kwong T, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981-1983: case control study. *Lancet* 8644: 917-922, 1989
13. Spitzer WO, Suissa S, Ernst P *et al.* The use of β -agonist and the risk of death and near death from asthma. *N Engl J Med* 326: 501-506, 1992
14. National Heart Lung and Blood Institute (National Institutes of Health) expert panel. Guidelines for the diagnosis and management of asthma. *J Allergy Clin Immunol* 88 (part 2, suppl): 425-492, 1991
15. Essen Van-Zandvliet EE, Kerrebijn KF. The effect of anti-asthma drugs on bronchial hyperresponsiveness. *Immunol Allerg Clin NA* 10: 483-501, 1990
16. Cockcroft DW, Murdock KY. Comparative effects of inhaled salbutamol, sodium cromoglycate and beclomethasone dipropionate on allergen-induced early asthmatic responses, late asthmatic responses and increased bronchial responsiveness to histamine. *J Allergy Clin Immunol* 79: 734-740, 1987
17. Molema J, Herwaarden Van CLA, Folgering HThM. Effects of long-term treatment with inhaled cromoglycate and budesonide on bronchial responsiveness in patients with allergic asthma. *Eur Respir J* 2: 308-316, 1989
18. Svendsen UG, Frolund L, Madsen F, Nielsen NH, Holstein-Rathlou N-H, Weeke B. A comparison of the effects of sodium cromoglycate and beclomethasone dipropionate on pulmonary function and bronchial hyperreactivity in subjects with asthma. *J Allergy Clin Immunol* 80: 68-74, 1987
19. Furukawa CT, Shapiro GG, Bierman CW *et al.* A double blind study comparing the effectiveness of cromolyn sodium and sustained release theophylline in childhood asthma. *Pediatrics* 74: 453-459, 1984
20. Selroos O. The effects of inhaled corticosteroids on the natural history of obstructive lung diseases. *Eur Respir Rev* 1: 354-365, 1991
21. Waalkens HJ, Gerritsen J, Koëter GH, Krouwels FH, Van Aalderen WMC, Knol K. Budesonide and terbutaline or terbutaline alone in children with mild asthma: effects on bronchial hyperresponsiveness and diurnal variation in peak flow. *Thorax* 46: 499-503, 1991
22. Bennati D, Piacentini GL, Peroni DG, Sette L, Testi R, Boner AL. Changes in bronchial reactivity in asthmatic children after treatment with beclomethasone alone or in association with salbutamol. *J Asthma* 26: 359-364, 1989
23. Kerrebijn KF, Van Essen-Zandvliet EEM, Neijens HJ. Effects of long-term treatment with inhaled corticosteroids and beta-agonists on the bronchial responsiveness in asthmatic children. *J Allergy Clin Immunol* 79: 653-659, 1987
24. Barnes PJ. Effect of corticosteroids on airway hyperresponsiveness. *Am Rev Respir Dis* 141: S70-S76, 1990

25. Juniper EF, Kline PA, Vanzielegem MA, Ramsdale EH, O'Bryne PM, Hargreave FE. Effect of long-term treatment with an inhaled corticosteroid (budesonide) on airway hyperresponsiveness and clinical asthma in nonsteroid-dependent asthmatics. *Am Rev Respir Dis* 142: 832-836, 1990
26. Juniper EF, Kline PA, Vanzielegem MA, Ramsdale EH, O'Bryne PM, Hargreave FE. Long-term effects of budesonide on airway responsiveness and clinical asthma severity in inhaled steroid-dependent asthmatics. *Eur Respir J* 3: 1122-1127, 1990
27. Haahtela T, Jarvinen M, Kava T *et al.* Comparison of a beta2-agonist, terbutaline, with an inhaled corticosteroid, budesonide, in newly detected asthma. *N Eng J Med* 325: 388-392, 1991
28. Bel EH, Timmers MC, Hermans J, Dijkman JH, Sterk PJ. The long-term effects of nedocromil sodium and beclomethasone dipropionate on bronchial responsiveness to methacholine in nonatopic asthmatic subjects. *Am Rev Respir Dis* 141: 21-28, 1990
29. Vathenen AS, Knox AJ, Wisniewski A, Tattersfield AE. Time course of change in bronchial reactivity with an inhaled corticosteroid in asthma. *Am Rev Respir Dis* 143: 1317-1321, 1991
30. Buxton MJ. Problems in the economic appraisal of new technology: the evaluation of heart transplants in the UK. In: Drummond MF (ed). *Economic Appraisal of Health Technology in the European Community*. Oxford University Press, 1987
31. Essen Van-Zandvliet EE, Hughes MD, Waalkens HJ, Duiverman EJ, Pocock SJ, Kerrebijn KF and the Dutch Chronic Non-Specific Lung Disease Study Group. Effects of 22 months treatment with inhaled corticosteroids and/or beta-2-agonists on lung function, airway responsiveness and symptoms in children with asthma. *Am Rev Respir Dis* 146: 547-554, 1992
32. Miller MR, Pincock AC. Predicted values; how should we use them? *Thorax* 43: 265-267, 1988
33. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, 1987
34. Rice DP, Hodgson TA, Kopsten AN. The economic costs of illness: a replication and update. *Health Care Financing* 7: 61-80, 1985
35. Luce BR, Elixhauser A. Estimating costs in the economic evaluation of medical technologies. *Int J Techn Ass Health Care* 6: 57-75, 1990
36. Armitage P, Berry G. *Statistical methods in medical research*. Blackwell Scientific Publications, Oxford, 1987
37. Doorslaer EKA van, Rutten FFH, Haan G. Economic evaluation of preventive health care. Is one ounce prevention worth one pound of therapy? In: Lapré RM, Rutten FFH, Hagen JH (ed). *The economics of preventive health care*. Conference report of the Society for Health Economics, De Tijdstroom, Lochem, Gent, 1988 (in Dutch)
38. Toevs CD, Kaplan M, Atkins CJ. The costs and effects of behavioral programs in chronic obstructive pulmonary disease. *Medical Care* 22: 1088-1100, 1984
39. Wolthers OD, Pedersen S. Controlled study of linear growth in asthmatic children during treatment with inhaled glucocorticoids. *Pediatrics* 89: 839-842, 1992

4

Costs and effects of inhaled corticosteroids and bronchodilators in asthma and COPD

*Maureen P.M.H. Rutten-van Mölken¹, Eddy K.A. Van Doorslaer²,
Margreet C.C. Jansen¹, Huib A.M. Kerstjens³, Frans F.H. Rutten².*

1 Department of Health Economics, University of Limburg.

2 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

3 Department of Pulmonology, University Hospital Groningen.

Forthcoming as: Costs and effects of inhaled corticosteroids and bronchodilators in asthma and COPD. American Journal of Respiratory and Critical Care Medicine. Reprinted with permission of the publisher.

4.1 Abstract

The objective of this study was to determine the costs and effects of combined bronchodilator and anti-inflammatory therapy.

In a 2.5 year randomized controlled study combined β_2 -agonist/corticosteroid therapy (BA+CS) and combined β_2 -agonist/anticholinergic therapy (BA+AC) were compared to β_2 -agonist/placebo therapy (BA+PL). Included in the study were 274 patients with moderately severe obstructive airways disease aged 18-60 years. The main clinical endpoints were lung function, hyperresponsiveness, restricted activity days and symptom-free days. The economic endpoints were the costs of health care utilization.

Compared to BA+PL, BA+CS led to significant improvements in FEV₁, PC₂₀ and symptom-free days. BA+AC did not differ from BA+PL in this respect. The respective annual acquisition costs of BA+CS, BA+AC and BA+PL were 532 US\$, 277 US\$ and 156 US\$. Thus, BA+CS costs 376 US\$ more than BA+PL. However, compared to BA+PL therapy, BA+CS led to statistically significant savings in other health care costs of about 175 US\$ (95% CI from 46 to 303 US\$). Thus more than half of the additional costs of adding the inhaled corticosteroid are compensated for by a reduction in the costs of other health care services. Overall, inhaled corticosteroids lead to a small but net increase in health care costs of 201 US\$ per patient per year. The incremental cost-effectiveness ratio of BA+CS compared to BA+PL ranges from 200 US\$ per 10% increase in FEV₁ to 5 US\$ per symptom-free day gained. In order to reach net societal savings the economic benefits of increased productivity due to inhaled corticosteroids have to be valued higher than 42 US\$ per day. No significant differences in health care costs were found between the BA+AC and BA+PL groups.

It can be concluded that the addition of an inhaled corticosteroid to a β_2 -agonist leads to significant benefits in respiratory function and restricted activity days which seem to be worth the relatively low additional health care costs, whereas addition of an anticholinergic agent appears expensive and of no long-term value.

4.2 Introduction

Recently, several epidemiologic and economic publications have established the considerable cost of illness associated with asthma,^{1,3} chronic obstructive pulmonary disease (COPD)^{4,5} or both.⁶ The costs of health care utilization due to asthma in the United States have been estimated to exceed \$3.6 billion in 1990.¹ In the Netherlands the costs of asthma and COPD exceeded US \$0.5 billion in 1989.⁶ Pharmacotherapy is the principal element of this treatment and entails high costs: more than 30% of the direct medical care expenditure on asthma in the United

States was spent on pharmacotherapy,¹ and more than 7% of total drug sales in the Netherlands are related to asthma and COPD management. This knowledge and the increasing need to set priorities for resource allocation decisions on the basis of efficiency considerations, led to an increased interest in the cost-effectiveness of diagnostic and therapeutic interventions in asthma and COPD. Despite the interest and several publications about the cost-effectiveness of other interventions in this area,⁷ no prospective investigation into the cost-effectiveness of drugs for asthma or COPD has as yet been reported. Yet such studies would be relevant. Guidelines for the management of patients with asthma have recently been published, stating an important role for inhaled corticosteroids.⁸⁻¹⁰ Cost-effectiveness research may contribute to an evaluation of the appropriateness of these guidelines. Moreover, new drugs are continuously being added to the large number of already available, mostly fairly effective drugs. There is not always evidence of additional benefits and the new drugs often do not encourage cost-reducing therapeutic substitutions.

This article reports on an investigation into the costs and effects of combined inhaled bronchodilator and anti-inflammatory therapy consisting of a β_2 -agonist plus an inhaled corticosteroid, in comparison with two different bronchodilator therapies consisting of a β_2 -agonist combined with either an anticholinergic or a placebo. As far as we know, this is the first prospective longitudinal economic evaluation of long-term respiratory drug use. Recent studies have shown that corticosteroids achieve beneficial effects in asthmatic patients,¹¹⁻¹⁷ but there is less unanimity about the effects of corticosteroids in patients with COPD.¹⁸⁻²¹ No long-term effects of anticholinergics are known. The long-term effects of β_2 -agonists are under heavy debate.²²⁻²⁵

The question addressed in the present study is whether the additional treatment costs of the combination therapies are outweighed or justified by additional clinical benefits and reduced utilization of other health care services. Special efforts have been made to ascertain whether the cost differences have statistical significance. This question was specifically added to meet a concern expressed by commentators on cost-effectiveness studies: "There is, I fear, a tendency on all sides to be satisfied with much looser, more subjective, and more speculative assessments for costs ... than would be acceptable for clinical information".²⁶ To meet this concern, the observed and systematically recorded economic data was subjected -where possible- to the rigours of statistical testing, as was the clinical data.

4.3 Methods

The economic analysis was performed alongside a randomized double-blind multi-centre trial, which is described in more detail elsewhere.^{27,28} A brief summary of its design is given here. The study included 274 patients with obstructive airways disease aged between 18 and 60 who exhibited both airway obstruction and hyperresponsiveness. Airway obstruction was defined as an FEV₁ (Forced Expiratory Volume in one second) or an FEV₁/FVC (Forced Vital Capacity) below the 95% confidence interval of normality. Hyperresponsiveness was defined by the concentration of histamine which caused a 20% fall in FEV₁ (PC₂₀). Patients were recruited from 6 university pulmonary outpatient clinics in the Netherlands. Candidates were excluded if they had serious concomitant disease or required drugs likely to interfere with the purpose of the study. The use of inhaled corticosteroids was tapered off and discontinued completely at least 4 weeks before randomization. Informed consent was obtained from all patients and the study was approved by the medical ethics committees.

Immediately after the baseline assessment, patients were randomly assigned to one of three different drug regimes: 1) BA+CS: an inhaled β_2 -agonist (terbutaline 250 μg 2 puffs q.i.d.) combined with an inhaled corticosteroid (beclomethasone 100 μg 2 puffs q.i.d.), 2) BA+AC: an inhaled β_2 -agonist combined with an inhaled anticholinergic (ipratropium bromide 20 μg 2 puffs q.i.d.), 3) BA+PL: an inhaled β_2 -agonist combined with an inhaled placebo. Randomization was stratified by age, sex, FEV₁ expressed as a percentage of the predicted value (FEV₁ % predicted), reversibility, PC₂₀, allergy, smoking habits, prior use of inhaled steroids and treatment centre.²⁷ Stratification was performed using the minimization method²⁹ on a personal computer by an independent 24 hours service telephone centre. No concomitant medication was allowed, except for rescue therapy comprising additional inhaled salbutamol for increasing pulmonary complaints at the patients' discretion, a course of oral prednisolone in the event of an exacerbation and a course of antibiotics in case of bacterial infection.

At the end of the study, data from a standardized history were used to identify different syndromes, thereby adhering to the 1987 criteria of the American Thoracic Society. Patients reporting attacks of breathlessness and wheeze (asthma attacks) without chronic (i.e. for more than 3 months per year) cough or sputum production were identified as having asthma. Current or former smokers without a history of asthma attacks, reporting either chronic cough with or without sputum production or dyspnea when walking quietly on level ground, or both, were included in the COPD group. Patients with both asthmatic attacks or recurrent wheeze and chronic cough and sputum production were labelled asthmatic bronchitis. Subjects with insufficient data to establish a diagnosis from history taking were included in an undefined diagnosis group.²⁸

Patients were seen at the outpatient clinics every three months for monitoring, data-collection and the distribution of new canisters during a follow-up of 2.5 years.

They were maintained on the drugs being evaluated unless they dropped out on their own initiative or were withdrawn due to unacceptable increases in pulmonary symptoms, adverse effects, non-compliance with the protocol or the fact that they required more than 2 courses of prednisolone over a 3 months period or more than 4 courses of prednisolone over a year.²⁸ After their withdrawal patients were no longer followed, so no clinical and economic data after withdrawal are available.

4.3.1 Clinical endpoints

The primary physiological endpoints were FEV₁% predicted, used as an indicator of airway obstruction and PC₂₀ histamine as an indicator of responsiveness. FEV₁% predicted was measured at baseline and every three months during the follow-up. PC₂₀ histamine was assessed at baseline and every six months. Both of these measures were only performed when patients were clinically stable and not within 4 weeks after termination of a prednisolone course. Eight hours before these tests, all study medication was discontinued.

Important additional measures of effectiveness were the number of symptom-free days and the number of restricted activity days (including lost work days) prevented. Symptom-free days were measured two weeks prior to each outpatient visit. During these weeks, patients recorded in a diary the presence and severity of four symptoms -wheezing, coughing, shortness of breath and mucus secretion- on a four-point scale. The information was extrapolated to a period of three months. A symptom-free day was defined as a day on which all four symptoms were recorded to be totally absent. Restricted activity days were measured every three months asking the following question: "How many days during the last 3 months have you been unable to go to work or to do your housekeeping or usual activities due to your asthma or COPD?" By asking the question this way the restricted activity days of people performing unpaid labour were explicitly included.

4.3.2 Economic endpoints

The study can be characterised as an incremental costs and effect analysis. Its aim is to assess the difference in costs of health care utilization and effects between the treatment groups. Consequently, we excluded the costs of the regular outpatient visits and lung function tests that were scheduled every three months for all treatment groups because of the trial. Also excluded were the costs of special visits and tests that were required for the study protocol only and not for patient care. An attempt was made to compute health care costs from a societal viewpoint, which means that we estimated the social value of the resources devoted to care instead of using charges.

The costs of health care utilization included the costs of: 1) study drugs, 2) oral prednisolone rescue therapy, 3) antibiotics rescue therapy, 4) inhaled salbutamol rescue therapy, 5) additional drug use related to asthma and COPD other than the allowed rescue therapy (e.g. mucolytics, antihistamines, theophyllines, homeopathic drugs, drugs against cold and cough), 6) additional non-scheduled outpatient visits, tests and telephone contacts related to asthma or COPD, 7) hospitalizations due to asthma or COPD. Information on the use of oral prednisolone, antibiotics and additional non-rescue drugs was collected every three months during the outpatient visits by means of a semi-structured interview by the pulmonary physician. During this interview, patients were also asked whether they had had any additional contact with the outpatient clinic or the hospital since the last scheduled visit. If so, more detailed information on these additional non-scheduled outpatient visits, telephone contacts, hospitalizations or potential additional services was obtained from the patients' records. To record the number of salbutamol rotacaps (dry powder) used, patients kept a diary during the two weeks prior to each visit. This information was extrapolated to a period of 3 months. All costs were based on 1989 Dutch prices which were expressed in US Dollars. The 1989 exchange rate was 1 US Dollar to 2.12 Dutch Guilders. The costs of drugs included a mark-up for pharmacy expenses. Estimates of the costs of additional non-scheduled outpatient visits and lung function tests were based on real resource cost calculations. Detailed cost-accounting studies were performed in two of the six participating centres and mean values were taken as cost estimates. Only the estimates of the costs of hospitalizations and a few, rarely performed tests were based on charges. The costs of hospitalizations were estimated by averaging the per diem fees of the participating hospitals. This average fee per day was multiplied by the length of stay to obtain the costs per hospitalization.

4.3.3 Statistical analyses

Due to the mentioned absence of data after withdrawal, a pure intention-to-treat-analysis of the differences between the treatment groups was not possible. For all patients, both those who completed the study and those who did not, we calculated the mean costs per patient-year by dividing each patient's total health care costs by the number of quarters the patient was in the trial, the result of which was multiplied by 4. This way of calculating annual health care costs per patient accounts for the different durations of follow-up. Data on symptom free days and restricted activity days were handled in the same way.

In all analyses each of the two combination therapies is compared to BA+PL. Unless otherwise stated, means and associated standard errors are reported. All PC_{20} calculations were performed using the base-2 logarithmic of PC_{20} as this reflects doubling concentrations.²⁸ Differences in withdrawal rates were tested using

z-tests. Differences between the therapeutic groups in terms of FEV₁% predicted, PC₂₀, restricted activity days and symptom-free days were tested using t-tests.

Because of a few patients contributing large costs, the distribution of costs was heavily skewed to the right. As a result, common parametric analysis for testing differences in costs were not appropriate. We therefore performed an adjusted statistical analysis derived from the methods applied in the Rand Health Insurance Study.^{30,31} This analysis provides unbiased, reliable and more precise estimates of the differences in costs. It consists of the following four steps. First, costs were logarithmically transformed in order to reduce the skewness and thus better approximate a normal distribution.³² As a result of this transformation, the impact of high cost events is greatly reduced. Furthermore, after this transformation the standard errors of the mean were no longer significantly different between the treatment groups. Second, multiple regression analysis was performed to test the difference in transformed costs between the treatment groups, thereby adjusting for the effects of various baseline patient characteristics on the logarithmically transformed mean costs per patient-year. Third, using the regression equation we predicted the transformed costs for each treatment group, and standardized the predictions for differences in baseline characteristics. Finally, the transformed costs which were predicted by means of the regression equation were retransformed into money terms using a smearing factor, which has been shown to be the most appropriate method of retransformation.³³ The exact procedure and equations that were used for transformation, standardization and retransformation are described in detail in chapter 8.

4.4 Results

4.4.1 Baseline characteristics

Two-hundred and seventy-four patients were randomly assigned to one of the 3 treatment groups. Table 1 shows that the treatment groups did not differ significantly at baseline with respect to age, gender, current smoking, prior use of inhaled corticosteroids, diagnosis, allergy, FEV₁% predicted and hyper-responsiveness. More details on baseline characteristics have been presented previously.^{27,28}

Table 1. Patient characteristics by treatment group.

	BA+CS [†]	BA+AC [†]	BA+PL [†]
No. of patients	91	92	91
Age: mean (sd)	40.2 (12.3)	38.9 (12.7)	39.6 (11.6)
Gender: no (%) male	59 (65%)	59 (64%)	58 (64%)
Currently smoking: no (%)	33 (36%)	31 (34%)	34 (37%)
Prior use of inhaled corticosteroids: no (%)	52 (57%)	53 (58%)	55 (60%)
Diagnoses based on symptoms: no(%)			
asthma	32 (35%)	37 (40%)	30 (33%)
COPD	31 (34%)	11 (12%)	21 (23%)
asthmatic bronchitis	19 (21%)	33 (36%)	24 (26%)
undefined	9 (10%)	11 (12%)	16 (18%)
Allergic: no (%)	51 (56%)	51 (55%)	51 (56%)
FEV ₁ % predicted: mean (sd)	64.6 (15.4)	63.3 (15.0)	63.3 (15.7)
log ₂ PC ₂₀ histamine(mg/ml): mean (sd)	-1.62 (2.23)	-2.01 (2.31)	-2.21 (2.33)

† BA+CS = β_2 -agonist plus corticosteroid
 BA+AC = β_2 -agonist plus anticholinergic
 BA+PL = β_2 -agonist plus placebo

4.4.2 Withdrawals

The study was terminated earlier than originally planned, because interim analysis had signalled highly selective, treatment-related withdrawal and significant differences in primary physiological endpoints. Figure 1 shows the number of withdrawals, accumulated over the study period, starting immediately after randomization. Withdrawal rates were significantly higher in the groups receiving treatments BA+AC (49%) and BA+PL (48%) than in the group receiving treatment BA+CS (13%). Seventy of the total of 101 withdrawals (5 in BA+CS; 35 in BA+AC; 30 in BA+PL) dropped out because of an increase in pulmonary problems. Twenty patients (1 in BA+CS; 6 in BA+AC and 13 in BA+PL) dropped out because they were no longer willing to participate. Three patients dropped out because of adverse effects: 1 because of hoarseness (BA+CS), 1 because of tremor (BA+CS) and 1 because of a sore throat after each inhalation (BA+PL). Eight others (4 in BA+CS; 4 in BA+PL) dropped out for reasons unrelated to the disease or the treatment.²⁸ Compared with those who completed the trial, those who withdrew were more likely to be female and their baseline lung function and hyperresponsiveness was worse. Other baseline characteristics were not significantly different from the characteristics of those who did not withdraw.²⁸

Compared with those who completed the trial, those who withdrew were more likely to be female and their baseline lung function and hyperresponsiveness was worse. Other baseline characteristics were not significantly different from the characteristics of those who did not withdraw.²⁸

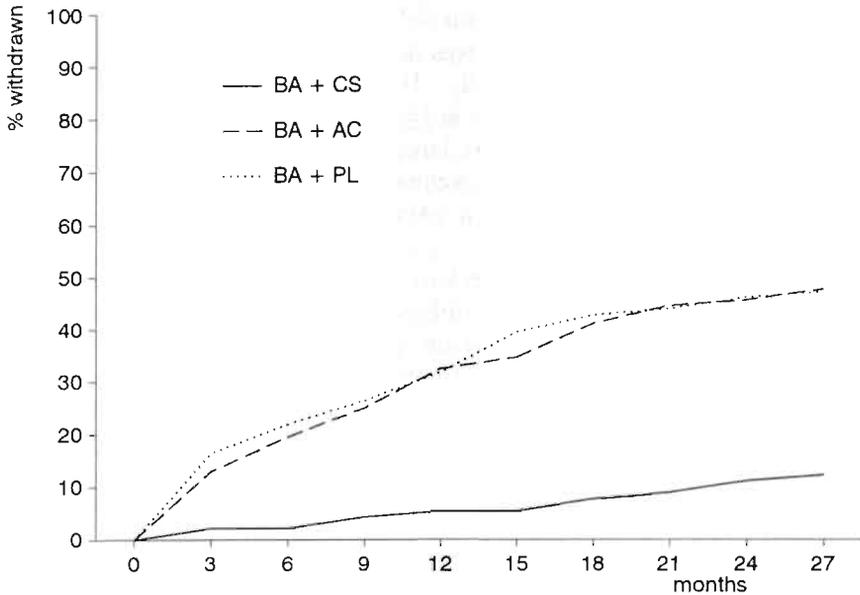


Figure 1. Cumulative withdrawal rates

Withdrawal was associated with higher health care costs. In all treatment groups, the mean of the observed costs per quarter were at least twice as high in withdrawals as in non-withdrawals. The costs of the former were generally highest in the 3 months immediately preceding withdrawal. All 173 patients who did not withdraw reached a follow-up of at least 21 months, 129 of them reached the full follow-up of 2.5 years.

4.4.3 Efficacy

A description and discussion of the physiological results has been published by Kerstjens *et al.*²⁸ In summary, FEV₁ and PC₂₀ showed significantly more improvement in patients receiving BA+CS than in those receiving BA+PL or BA+AC. The mean difference in changes in FEV₁% predicted from baseline between these groups was 10.03 (se 2.7) during the first year ($p < 0.001$). This difference was already reached at 3 months follow-up and remained approximately constant over the 2.5 year period. There was no significant difference in the change in FEV₁ between BA+AC and BA+PL. The mean difference after one year between BA+CS and BA+PL in change in PC₂₀ from baseline was 1.43 (se 0.43) doubling concentrations ($p < 0.0001$). The largest part of the increase in PC₂₀ was reached after 6 months, but further improvements occurred until a mean difference of 2 doubling concentrations was reached. Again no difference between BA+AC and BA+PL was found.

Table 2 shows that patients in the BA+CS group reported on average 5 days of restricted activity per patient-year due to breathing problems, whereas patients in the BA+AC and BA+PL group reported on average 13 and 10 restricted activity days per patient-year respectively. The differences between BA+CS and BA+AC as compared to BA+PL were not statistically significant. The table also shows the number of symptom-free days per patient-year to be 121 in the BA+CS group compared to 84 in the BA+PL group and 68 in the BA+AC group. The mean (se) difference of 37.5 (17.8) days between BA+CS and BA+PL was statistically significant ($p = 0.032$).

Table 2. Number of restricted activity days due to obstructive lung disease and number of symptom-free days per patient-year per treatment group

	BA+CS [†]	BA+AC [†]	BA+PL [†]
no. patients at baseline	91	92	91
<i>Absenteeism/restricted activity days</i>			
no. patients with ≥ 1 restricted activity day	41	45	38
total number of restricted activity days	428	849	543
mean (se) no. restricted activity days per patient-year	5.2 (1.1)	13.0 (2.6)	10.0 (2.7)
<i>Symptom-free days</i>			
no. patients with ≥ 1 symptom-free day	65	47	50
total number of symptom-free days	10650	5028	5886
mean (se) no. symptom-free days per patient-year [‡]	121.3 (12.9)	68.2 (10.2)	83.7 (11.7)

† BA+CS = β_2 -agonist plus corticosteroid
 BA+AC = β_2 -agonist plus anticholinergic
 BA+PL = β_2 -agonist plus placebo

‡ $p = 0.032$ for the difference between BA+CS and BA+PL; the difference between BA+AC and BA+PL was not significant

4.4.4 Costs

Annual drug acquisition costs of combined BA and CS therapy were 532 US\$ per patient. This is 376 US\$ more than BA+PL therapy. The respective annual drug acquisition costs of BA+AC and BA+PL were 277 US\$ and 156 US\$. The relative price ratios of these drugs have not changed since 1989.

On the one hand BA+CS was about 3.5 times as expensive as BA+PL. On the other hand, table 3 shows that BA+CS led to a reduction in the costs of all other health care services. When compared to BA+PL, BA+CS therapy reduced the costs of rescue medication, other pulmonary drugs, outpatient contacts and hospitalizations. The costs of hospitalizations in particular contributed to the lower other health care cost of patients receiving BA+CS compared to patients receiving BA+AC or BA+PL. In the BA+CS group only one patient was hospitalized for 6 days with associated costs of 3,347 US\$, whereas 8 hospitalizations occurred in the BA+AC group with lengths of stay varying from 4 to 13 days and associated total costs of 20,535 US\$. In the BA+PL group the total costs of 6 hospitalizations with lengths of stay varying from 6 to 18 days amounted to 20,274 US\$. Overall, -when compared to BA+PL- BA+CS led to annual savings in other than the costs of the study drugs of about 454 US\$ per patient, thus outweighing the additional treatment costs of 376 US\$ by 78 US\$ per patient (Table 4). The additional treatment costs of BA+AC compared to BA+PL approximately equalled the savings due to BA+AC.

Table 3. Mean (se) observed costs per patient-year in US\$ by cost category and treatment group

cost category	BA+CS [†]	BA+AC [†]	BA+PL [†]
<i>Treatment costs</i>			
study drug acquisition costs	532.4	277.4	156.0
<i>Other health care costs</i>			
prednisolone rescue therapy	1.5 (0.3)	4.2 (0.7)	4.2 (0.8)
salbutamol rescue therapy	48.7 (6.9)	77.0 (8.1)	70.9 (8.0)
antibiotics rescue therapy	3.0 (0.7)	6.5 (1.2)	7.7 (1.7)
other additional pulmonary drug use	20.2 (6.2)	15.1 (4.4)	49.0 (14.4)
additional outpatient contacts and services	27.8 (5.3)	57.1 (13.3)	34.8 (7.3)
hospitalizations	18.8 (18.8)	285.1 (136.6)	407.1 (233.4)
total other health care costs	120.0 (21.8)	445.6 (138.5)	573.6 (248.3)
<i>Total health care costs incl. treatment costs</i>	652.5 (21.8)	722.9 (138.5)	729.7 (248.3)

† BA+CS = β_2 -agonist plus corticosteroid
 BA+AC = β_2 -agonist plus anticholinergic
 BA+PL = β_2 -agonist plus placebo

However, these simple arithmetical averages of costs and savings do not provide acceptable precision for analyzing the differences in health care costs. The large standard errors in table 3 indicate that the figures reported in this table are not very precise. This imprecision results from rare high cost events such as hospitalizations. In order to get more precise estimates of costs by treatment group, costs were logarithmically transformed, thereby reducing the influence of high costs events (i.e. reducing the skewness). Then, an ordinary least square regression analysis was performed to test for the effect of treatment on the transformed costs per patient-year, thereby adjusting for differences in patient characteristics between the groups. The dependent variable "costs per patient-year" did not include the study drug acquisition costs because these costs were not observed on an individual patient's basis. The results of the regression analysis are shown in table 4.

Table 4. Ordinary least squares regression analysis of mean costs per patient-year[†]

explanatory variables [‡]	total sample		subsample of non-withdrawals	
	b (se)	p-value	b (se)	p-value
BA+CS	-0.671 (0.231)	0.004	-0.570 (0.255)	0.027
BA+AC	0.109 (0.236)	0.643	-0.054 (0.284)	0.848
sex	0.027 (0.212)	0.900	0.070 (0.239)	0.770
smoking	-0.201 (0.215)	0.350	0.288 (0.235)	0.222
allergy	0.062 (0.224)	0.782	0.217 (0.251)	0.389
prior use of corticosteroids	0.641 (0.192)	0.001	0.428 (0.216)	0.049
age	-0.003 (0.009)	0.739	-0.011 (0.010)	0.251
FEV ₁ % pred.	-0.021 (0.007)	0.002	-0.019 (0.007)	0.009
log ₂ PC ₂₀	-0.056 (0.047)	0.242	0.036 (0.053)	0.494
reversibility	-0.005 (0.013)	0.674	-0.004 (0.015)	0.791
intercept	6.545 (0.741)	0.000	6.491 (0.841)	0.000
adjusted R ²	0.120		0.052	
F-value	4.313	0.000	1.941	0.043

[†] dependent variable: $\log(\text{costs per patient-year} + 10)$
 study drug acquisition costs were not included in the costs per patient-year because these costs were not observed on an individual patient's basis

[‡] explanatory variables at baseline:

BA+CS: 1=BA+CS, 0=BA+AC or BA+PL

BA+AC: 1=BA+AC, 0=BA+CS or BA+PL

sex: 1=male, 0=female

smoking: 1=smoker, 0 non-smoker

allergy: 1=allergic, 0=non-allergic

prior use of steroids: 1=prior use, 0=no prior use

FEV₁: Forced Expiratory Volume in 1 second as a % of the predicted value

PC₂₀: Provocative Concentration of histamine causing a 20% fall in FEV₁

Reversibility: reversibility of airways obstruction after inhaling terbutaline and ipratropium

Other things equal, patients receiving BA+CS had significantly lower health care costs than patients receiving BA+PL, patients with higher levels of FEV₁ had significantly lower health care costs than patients with lower levels of FEV₁ and patients who had not used inhaled corticosteroids prior to the start of the trial had significantly lower health care costs than patients who had used inhaled corticosteroids. There is no significant difference in costs between BA+AC and BA+PL therapy.

The regression equation was used to predict the expected transformed costs by treatment, thereby standardizing for any differences in baseline characteristics. The transformed costs estimated this way were then retransformed into US\$ by using a smearing factor. The results of applying this technique are given in figure 2.

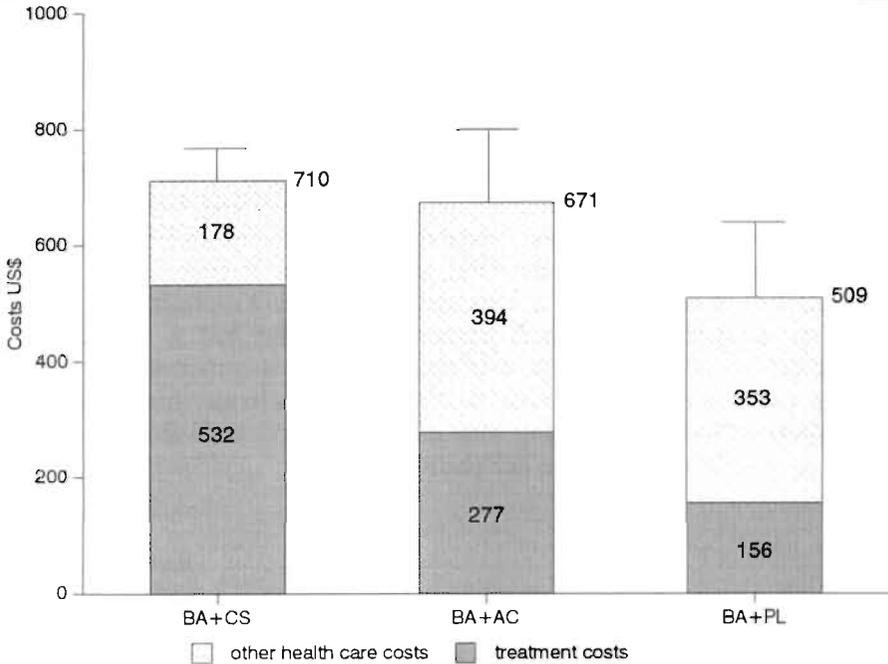


Figure 2. Estimated adjusted mean costs per patient-year
Error bars: upper limit of the 95% confidence interval

Annual non-treatment health care costs were estimated at 178 US\$ (se 28) for patients given BA+CS, 394 US\$ (se 65) for patients given BA+AC and 353 US\$ (se 60) for patients given BA+PL. When compared to BA+PL, the savings due to BA+CS therapy in other than the treatment costs were estimated to be about 175 US\$ with 95% confidence intervals ranging from 46 to 303 US\$. These estimated savings are smaller than the actually observed savings which were derived

from table 4. This is a result of both the logarithmic transformation, which reduces the imprecision that resulted from the impact of outlier values, and the correction for confounding by differences in patient characteristics. But, although the estimated difference in costs between BA+CS and BA+PL is smaller than the actually observed difference, the regression analysis has shown that it is highly significant.

Overall, of the additional annual treatment costs due to inhaled corticosteroids of 376 US\$ per patient, about 175 US\$ is compensated for by a reduction in the use of other health care services. Hence adding an inhaled corticosteroid to an inhaled β_2 -agonist for this type of patients results in net additional health care costs of 201 US\$ per patient per year.

4.4.5 Subgroup analyses

Subgroup analysis of cost differences between BA+CS and BA+PL were undertaken for patients stratified by age, allergy, baseline lung function, baseline hyperresponsiveness, prior use of corticosteroids, reversibility, smoking and diagnosis group (Table 5). A significant interaction was found between treatment and smoking, hyperresponsiveness, reversibility and symptom-based diagnosis. When BA+CS was compared to BA+PL, the reduction in costs due to BA+CS was more marked in patients who did not smoke, who were more hyperresponsive, whose airway obstruction was more reversible and who had a symptom-based diagnosis of asthma (Table 5). Thus, it is clear that inhaled corticosteroids are most beneficial in patients with the "classic" asthma profile (allergic, highly reversible, severely hyperreactive, non-smoking and mostly young), both from the clinical perspective as well as the economic perspective.

Table 5. Interaction effects: comparison of the log-costs of CS and BA+PL according to subgroup

	log-costs		p-value
	BA+CS [†]	BA+PL [†]	
<i>Age</i>			
≤ 40 years	4.52	5.24	0.405
> 40 years	4.69	4.99	
<i>Fev₁ % pred.</i>			
≤ 65 %	4.98	5.46	0.848
> 65 %	4.29	4.68	
<i>PC₂₀</i>			
≤ 0.1 mg/ml	4.30	5.63	0.015
> 0.1 to ≤ 0.6 mg/ml	4.60	5.17	
> 0.6 mg/ml	4.83	4.30	
<i>Smoking status</i>			
Non-smoker	4.23	5.35	0.001
Smoker	5.24	4.71	
<i>Allergy</i>			
Non-allergic	4.76	4.74	0.064
Allergic	4.49	5.41	
<i>Prior use of corticosteroids</i>			
No prior use	4.32	4.58	0.438
Prior use	4.83	5.47	
<i>Reversibility</i>			
≤ 9% of predicted FEV ₁	4.76	4.62	0.028
> 9% of predicted FEV ₁	4.47	5.45	
<i>Symptom based diagnosis</i>			
asthma	3.81	5.37	0.020
asthmatic bronchitis	4.83	5.10	
COPD	5.35	5.25	

† BA+CS = β_2 -agonist plus corticosteroid

BA+PL = β_2 -agonist plus placebo

Other abbreviations see table 5.

4.4.6 Cost-effectiveness ratio

The incremental cost-effectiveness of combined BA+CS therapy is calculated as the ratio of the cost difference between BA+CS and BA+PL to the difference in effect parameters between BA+CS and BA+PL. By doing so, we assumed that the effects as measured in this study are a good approximation of the community effectiveness, which is the effectiveness of the therapy in the real daily clinical

practice. On the one hand, when compared to BA+PL, BA+CS led to a net increase in health care costs of 201 US\$ per patient-year. But on the other hand the mean difference in change in FEV₁ between the two groups was 10.03% (se 2.7) in favor of BA+CS. In other words, the incremental cost-effectiveness ratio is about 200 US\$ per 10% increase in FEV₁, with an approximated 95% confidence interval ranging from 57 US\$ to 450 US\$ per 10% increase in FEV₁. We also calculated the costs per symptom-free day gained. Dividing the costs difference of 201 US\$ by the 37.6 symptom-free days gained due to inhaled corticosteroids (see table 2), results in a ratio of 5.35 US\$ per symptom-free day gained. The approximated 95% confidence interval around this point estimate ranges from somewhat more than 1 US\$ to 127 US\$ per symptom-free day gained.

It is not relevant to calculate such a ratio for the cost-effectiveness of BA+AC compared to BA+PL, because the costs of BA+AC therapy are higher but no long-term effects in addition to those produced by BA+PL therapy could be found.

4.5 Conclusion and discussion

This study shows that the addition of an inhaled corticosteroid to maintenance treatment of patients with moderately severe obstructive airway disease with a β_2 -agonist results in a small but significant net increase in health care costs of 201 US\$ per patient. This increase in costs results from the fact that the annual drug acquisition costs of corticosteroid of 532 US\$ are 201 US\$ higher than the savings in other health care costs of 376 US\$ due to corticosteroid.

Whether or not this net increase in costs is worthwhile largely depends on the relative effectiveness of the therapies. The increase due to adding the corticosteroid may easily be justified by the fact that BA+CS therapy is very beneficiary compared to bronchodilator monotherapy with respect to improvements in lung function, hyperresponsiveness, symptom-free days and restricted activity days. Besides, more than half of the additional costs of adding the inhaled corticosteroid were compensated for by a reduction in the costs of other health care services. Furthermore, we did not take into account the economic benefits of the increased productivity (usually referred to as indirect benefits)³⁵ that result from a reduction in the number of restricted activity days. The number of restricted activity days (including absence from work) was taken into account as an effect parameter, not as an indicator of indirect costs. This effect parameter was also found to favour combined bronchodilator and corticosteroid therapy, but the difference in restricted activity days between the treatment groups did not reach the conventional levels of statistical significance. The reason why we did not make an attempt to provide a dollar estimate of these indirect benefits is threefold. First, we wanted to avoid double counting. Second, the difference was not statistically significant and third, the calculation of indirect costs in economic evaluation studies is not

uncontroversial.^{35,36} There is considerable debate about how to value the production gains and how to handle the production of people performing unpaid labour. In our study, the impact of including indirect costs would be great. It is obvious that in case we would have valued the production gains for example by an average gross wage or by friction costs,³⁷ combined bronchodilator and corticosteroid therapy could easily lead to net total savings to society. The threshold value of the production gains on a restricted activity day prevented is about 42 US\$, above which BA + CS therapy leads to total net savings.

Thus, the recent treatment recommendations that bronchodilator therapy should be prescribed in combination with anti-inflammatory drugs only⁸⁻¹⁰ is not only supported by the efficacy component of this study but seems justifiable also from an economic perspective. This study however found no justification for the extra costs of adding an anticholinergic drug to the β -agonist, in terms of either additional long-term effects or in savings.

When calculating the cost-effectiveness ratio we assumed that the effects as measured in this study are a good proxy for the community effectiveness reached in the real world. A pure efficacy trial aims at assessing the potential effects of a drug under optimally controlled circumstances. In such a trial only those patients would be analyzed who completed the study fully in accordance with the protocol. However, we did not exclude any patients from the analysis, but included all available information of the withdrawals. Moreover, since our study population was selected on the basis of objective functional criteria, it covers a broad range of airways obstruction, airways hyperresponsiveness and age. It includes patients with asthma, COPD and patients having characteristics of both. In our opinion, this mixed study population more closely reflects the population to which in routine clinical practice inhaled corticosteroids are administered. Nevertheless, there may still be factors, such as diagnostic accuracy, patients' and health professionals' compliance³⁸ that cause this study to differ from routine clinical practice, but a study with a duration of treatment up to 2.5 years may be expected to better approximate real world compliance than a short-term study.

This study also gave due consideration to the problems associated with the appropriate application of cost-effectiveness analysis in the event of high treatment-related withdrawal and highly skewed cost-distributions, two problems that are quite common in economic evaluation studies. Undoubtedly, a pure intention-to-treat analysis would have been the first best option. However, given that large numbers of patients were lost to follow-up, such an analysis was impossible. Recognizing this, it was our intention to estimate what the differences in costs between the treatment groups would have been under the assumption that all patients would have continued the trial on their initial drug regime. This implies that withdrawals could not be excluded from the analysis. Exclusion of all the cost observations of patients who drop out, inevitably leads to a more serious underestimation of costs in BA+AC and BA+PL than in BA+CS and thus to a serious underestimation of the savings due to inhaled corticosteroids. To avoid this we used all the available

information prior to withdrawal to calculate mean costs per patient-year for all participants. Only for patients with less than one year of follow-up, we carried forward the mean of all available observations to a one-year period. This seems plausible, since it is likely that the deteriorated clinical condition and the high use of health care services observed prior to withdrawal would have sustained or even increased if patients had stayed in the trial on their initial drug regime. Since we had no evidence for a progressive rise we conservatively performed a linear extrapolation. Nevertheless, in order to avoid any overestimation of the costs in withdrawals and minimize the problems associated with highly skewed cost-distributions, the annualized costs per patient were logarithmically transformed in order to allow appropriate statistical testing. Given the impossibility of an intention-to-treat analysis, we recognize that our approach is a second best alternative.

To calculate incremental cost-effectiveness ratios we have used disease-specific outcome measures. Our cost-effectiveness ratio of 200 US\$ per 10% increase in FEV₁ and 5 US\$ per symptom-free day gained can be compared to the cost-effectiveness of other interventions in asthma or COPD. At the present such comparisons are not available. A disadvantage -especially of the costs per 10% increase in FEV₁ ratio- is that these ratios cannot be compared with ratios for pharmaceutical or other interventions in areas other than asthma and COPD. Therefore, it may be more appropriate for future cost-effectiveness studies to relate the costs to various quality of life outcomes rather than to FEV₁. There is a considerable amount of literature indicating that respiratory outcome measures and even symptom score poorly mirror the impact of the disease on quality of life.³⁹⁻⁴⁴ The weak correlations between lung function and quality of life suggests that these may be two completely different aspects of a disease. From an economic perspective, a generic non-disease specific quality of life measure (e.g. utility measures of quality adjusted life years) would be the most appropriate outcome measure, as this could allow a broad comparison across medical domains. It is very reasonable to assume that the significant improvements due to BA+CS in respiratory function and symptom-free days could lead to an improvement of the quality of life. However, confirmative evidence is needed.

An economic evaluation study such as the present study can be used to assess the economic impact of a particular pharmacotherapy, but it does not address the more difficult problem of allocating limited resources. The fact that we found that combined bronchodilator and corticosteroid therapy is efficacious might suggest that this therapy should be made available for up to 1 million Dutch people⁵ who might potentially benefit from it. If the annual additional health care cost of the general diffusion of this therapy are indeed about 200 US\$ per patient, direct health care costs would be increased by 200 million US\$. Any such decision would have to be weighed against a range of competing alternative uses of health care resources for which additional cost-effectiveness studies are urgently needed. However, in the case of combined inhaled bronchodilator and corticosteroid maintenance therapy this

amount of money would buy considerable health benefits for a large population of chronically ill.

4.6 Acknowledgements

We would like to thank the Dutch CNSLD group* and all participating centres for providing data, for their support when reviewing the patients' records and for their assistance during the cost-accounting studies. We would also like to thank Ben van Hout and René van Vliet for useful suggestions, Dirkje Postma for critical and constructive comments and Jan van Emmerik for computer assistance.

This study was supported by grants from the Netherlands' Health Research Promotion Program (SGO), the Dutch Asthma Foundation, Glaxo, Astra Pharmaceuticals and Boehringer Ingelheim.

* The Dutch Chronic Non-Specific Lung Disease (CNSLD) study group consists of a steering committee (K.F. Kerrebijn, Ph.H. Quanjer and H.J. Sluiter), of members from the Departments of Pulmonology of the University Hospital of Amsterdam (E.M. Pouw, D.F.M.E. Schoonbrood, C.M. Roos, H.M. Jansen), Groningen (P.L.P. Brand, A. de Gooyer, H.A.M. Kerstjens, D.S. Postma, Th.W. van der Mark, H.J. Sluiter, G.H. Koëter), Leiden (P.M. de Jong, P.J. Sterk, A.M.J. Wever, J.H. Dijkman), Nijmegen (P.N.R. Dekhuijzen, H.T.M. Folgering, C.L.A. van Herwaarden), Rotterdam (S.E. Overbeek, J.M. Bogaard, C. Hilvering) and Utrecht (S.J. Gans, H.J.J. Mengelers, B.A.H.A. van der Bruggen-Bogaarts, J. Kreukniet), from the Departments of Pediatric Pulmonology at the Sophia Children's Hospital, Rotterdam (E.E.M. van Essen-Zandvliet, K.F. Kerrebijn), the Juliana Children's Hospital, the Hague (E.J. Duiverman, J.M. Kouwenberg, J.E. Prinsen), University Hospital of Groningen (H.J. Waalkens, J. Gerritsen, K. Knol), from the Department of Allergology, University Hospital of Groningen (J.G.R. de Monchy), from the Department of General Practice, University of Leiden (A.A. Kaptein, F.W. Dekker), and from the Department of Physiology, University of Leiden (P.J.F.M. Merkus, Ph.H. Quanjer). Scientific counsel: S.J. Pocock, M.D. Hughes, (London, UK), E.R. Bleecker, D.A. Meyers (Baltimore, USA).

4.7 References

1. Weiss KB, Gergen PJ, Hodgson A. An economic evaluation of asthma in the United States. *N Engl J Med* 1992;326:862-6.
2. Mellis CM, Jennifer KP, Woolcock AJ. The Cost of Asthma. Can it be Reduced? *PharmacoEconomics* 1993;3:205-19.
3. Weiss KB, Dullivan SD. The Economic Costs of Asthma. A review and Conceptual model. *PharmacoEconomics* 1993;4:14-30
4. Strauss MH, Conrad D, LoGerfo JP, Hudson LD, Bergner M. Costs and outcome for patients with chronic obstructive lung disease. *Medical Care* 1986;24:915-24
5. Oster G, Golditz GA, Kelly NL. The economic costs of smoking and benefits of quitting. Lexington MA: Lexington Books, 1984.
6. STG (Stuurgroep Toekomstscenario's Gezondheidszorg). Chronic diseases in the year 2005. Part 2: Scenarios for chronic non-specific lung disease 1990-2005. Bohn Stafleu van Loghum, Houten/Antwerp, 1990 (in Dutch).
7. Rutten-van Mólken MPMH, Van Doorslaer EKA, Rutten FFH. Economic appraisal of asthma and COPD care: a literature review 1980-1991. *Soc Sci Med* 1992;35:161-75.
8. British Thoracic Society Research Unit of the Royal College of Physicians of London, King's Fund Centre, NA. Guidelines for the management of asthma in adults: I-chronic persistent asthma. *Br Med J* 1990;301:651-3.
9. NIH/NHLBI expert panel. Guidelines for the diagnosis and management of asthma. *J Allergy Clin Immunol* 1991;88(suppl 3):425-534.
10. NHLBI/NIH International Consensus report on diagnosis and management of asthma. U.S. Department of Health and Human Services. Public Health Service, National Institutes of Health. Publication No. 92-3091, Bethesda Maryland, 1992.
11. Barnes PJ. Effect of corticosteroids on airway hyperresponsiveness. *Am Rev Respir Dis* 1990;141:S70-S6.
12. Bel EH, Timmers MC, Hermans J, Dijkman JH, Sterk PJ. The long-term effects of nedocromil sodium and beclomethasone dipropionate on bronchial responsiveness to methacholine in nonatopic asthmatic subjects. *Am Rev Respir Dis* 1990;141:21-8.
13. Molema J, Herwaarden Van CLA, Folgering HThM. Effects of long-term treatment with inhaled cromoglycate and budesonide on bronchial responsiveness in patients with allergic asthma. *Eur Respir J* 1989;2:308-16.
14. Vathenen AS, Knox AJ, Wisniewski A, Tattersfield AE. Time course of change in bronchial reactivity with an inhaled corticosteroid in asthma. *Am Rev Respir Dis* 1991;143:1317-21.
15. Juniper EF, Kline PA, Vanzieleghem MA, Ramsdale EH, O'Byrne PM, Hargreave FE. Effect of long-term treatment with an inhaled corticosteroid (budesonide) on airway hyperresponsiveness and clinical asthma in nonsteroid-dependent asthmatics. *Am Rev Respir Dis* 1990;142:832-6.
16. Juniper EF, Kline PA, Vanzieleghem MA, Ramsdale EH, O'Byrne PM, Hargreave FE. Long-term effects of budesonide on airway responsiveness and clinical asthma severity in inhaled steroid-dependent asthmatics. *Eur Respir J* 1990;3:1122-7.
17. Haahtela T, Jarvinen M, Kava T, Kiviranta K, Koskinen S, Lehtonen K, Nikander K, Persson T, Reinikainen K, Selroos O, Sovijardi A, Stenius-Aarniala B, Svahn T, Tammivaara R, Laitinen L. Comparison of a beta2-agonist, terbutaline, with an inhaled corticosteroid, budesonide, in newly detected asthma. *N Eng J Med* 1991;325:388-92.

18. Eliasson R, Hoffman J, Trueb D, Frederick D, McCornick JR. Corticosteroids in COPD. A clinical trial and reassessment of the literature. *Chest* 1986;89:484-90.
19. Postma DS, Peters I, Steenhuis EJ, Sluiter HJ. Moderately severe chronic airflow obstruction. Can corticosteroids slow down progression? *Eur Respir J* 1988;1:22-6.
20. Weir DC, Grove RL, Robertson AS, Burge PS. Corticosteroid trials in non-asthmatic chronic airflow obstruction: a comparison of oral prednisolone and inhaled beclomethasone dipropionate. *Thorax* 1991;45:112-7.
21. Auffarth B, Postma DS, Monchy De JGR, Mark Van der ThW, Boorsma M, Koëter GH. Effects of inhaled budesonide on spirometric values, reversibility, airway responsiveness and cough threshold in smokers with chronic obstructive lung disease. *Thorax* 1991;45:372-7.
22. Sears MR, Taylor DR, Print G, Lake DC, Qingqing L, Flannery EM, Yates DM, Lucas MK, Herbison GP. Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* 1990;336:1391-6.
23. Schayck Van CP, Dompeling E, Herwaarden Van CLA, Folgering H, Verbeek ALM, Hoogen Van der HJM, Weel Van C. Bronchodilator treatment in moderate asthma or chronic bronchitis: continuous or on demand? *Br Med J* 1991;303:1426-31.
24. Crane J, Glat A, Jackson R, Ball M, Pearce N, Burgess C, Kwong T, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981-1983: case control study. *Lancet* 1989;8644:917-22.
25. Spitzer WO, Suissa S, Ernst P, Horwitz RI, Habbick B, Cockcroft D, Boivin J-F, McNutt M, Buist AS, Rebuck AS. The use of β -agonist and the risk of death and near death from asthma. *N Engl J Med* 1992;326:501-6.
26. Buxton MJ. Problems in the economic appraisal of new technology: the evaluation of heart transplants in the UK. In: Drummond MF; ed. *Economic Appraisal of Health Technology in the European Community*. Oxford University Press; 1987.
27. Brand PLP, Kerstjens HAM, Postma DS, Sterk PJ, Quanjer PhH, Sluiter HJ, Dijkman HJ, van Herwaarden CLA, Hilvering C, Jansen HM, Koeter GH, Kreukniet J, and the Dutch CNSLD study group. Long-term multicentre trial in chronic non-specific lung disease: methodology and baseline assessment in adult patients. *Eur Respir J* 1992;5:21-31.
28. Kerstjens HAM, Brand PLP, Hughes MD, Robinson NJ, Postma DS, Sluiter HJ, Bleecker ER, Dekhuijzen R, de Jong PM, Mengelers HJJ, Overbeek SE, Schoonbrood DFME, and the Dutch CNSLD Study Group. A comparison of bronchodilator therapy with or without inhaled corticosteroid therapy in obstructive airways disease. *N Engl J Med* 1992;327:1413-9.
29. Pocock SJ. *Clinical trials, a practical approach*. Chichester: John Wiley and Sons, 1983, pp 84
30. Newhouse JP, Manning WG, Morris CN, Orr LL, Duan N, Keeler EB, Leibowitz A, Marquis KH, Marquis MS, Phelps CE, Brook RH. Some interim results from a controlled trial of cost-sharing in health insurance. *N Engl J Med* 1981;305:1501-7.
31. Manning WG, Newhouse JP, Duan N, Keeler EB, Leibowitz A, Marquis MS. Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review* 1987;77:251-77.
32. Duan N, Manning WG, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Economic and Business Statistics* 1983;1:115-26.
33. Duan N. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* 1983;78:605-10.
34. Ruten-van Mólken MPMH, Van Doorslaer EKA, Van Vliet RCJA. Statistical analysis of cost outcomes in a randomized controlled trial. *Health Economics* (forthcoming).

35. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; 1987.
36. Koopmanschap MA, Rutten FFH. Indirect costs in economic studies: confronting the confusion. *Pharmacoeconomics* 1993;4:446-454.
37. Koopmanschap MA, Van Ineveld BM. Towards a new approach for estimating indirect costs of disease. *Social Science and Medicine* 1992;34:1005-1010
38. Tugwell P, Bennett KJ, Sackett CL, Haynes RB. The measurement iterative loop: a framework for the critical appraisal of need, benefits and costs of health interventions. *Journal of Chronic Diseases* 1985;38:339-351
39. McSweeney AJ, Heaton, RK, Grant I, Cugell D, Solliday N, Timms R. Chronic obstructive pulmonary disease; socioemotional adjustment and life quality. *Chest* 1980;77:309-311
40. Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987;42:773-778
41. Williams SJ, Bury MR. Impairment, disability and handicap in chronic respiratory illness. *Soc Sci Med* 1989;29:609-616
42. Schrier AC, Dekker F, Kaptein AA, Dijkman JH. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990;89:894-899
43. Van Schayck CP, Rutten-van Mólken MPMH, Van Doorslaer EKA, Folgering H, Van Weel C. Two-year bronchodilator treatment in patients with mild airflow obstruction. Contradictory effects on lung function and quality of life. *Chest* 1992;102:1384-1391
44. Malo J-L, Boulet L-P, Dewitte J-D, Cartier A, L'Archeveque J, Coté J, Bedard G, Boucher S, Champagne F, Tessier G, Contandriopoulos A-P, Juniper EF, Guyatt GH. Quality of life of subjects with occupational asthma. *J Allergy Clin Immunol* 1993;91:1121-1127

5

Two-year bronchodilator treatment in patients with mild airflow obstruction: Contradictory effects on lung function and quality of life

*Constant P. van Schayck¹, Maureen P.M.H. Rutten-van Mölken²,
Eddy K.A. van Doorslaer³, Hans Folgering⁴, Chris van Weel¹.*

- 1 Department of Family Medicine and Pulmonary Diseases, University of Nijmegen.
- 2 Department of Health Economics, University of Limburg.
- 3 Institute for Medical Technology Assessment, Erasmus University Rotterdam.
- 4 Department of Pulmonology, University of Nijmegen.

Originally published as: Two-year bronchodilator treatment in patients with mild airflow obstruction: Contradictory effects on lung function and quality of life. Chest 1992; 102:1384-1391. Reprinted with permission of the publisher.

5.1 Abstract

In a two-year randomized controlled study, we studied the effects of bronchodilator treatment on the lung function and the quality of life of patients with mild airflow obstruction. The patients were randomly divided to receive either continuous or symptomatic bronchodilator treatment. Within these treatment groups they received salbutamol in the first year and ipratropium bromide in the second or vice versa. In addition, the quality of life of the patients was compared to that of the general population.

One hundred and forty-four patients completed the study. When compared to the general population, these patients showed a serious impairment in quality of life. No differences between the two drugs were found, but the results indicated that FEV₁ decline in the continuously treated group was significantly larger than in the symptomatically treated group. However, this was not reflected in a significant deterioration of the quality of life in the continuous group as measured by means of the Nottingham Health Profile and the Inventory of Subjective Health. Decline in FEV₁ showed no correlation with changes in quality of life scores. This may be due to a relatively rapid adjustment of the patients to a decline in FEV₁, as a result of which it has no direct effect on the *experienced* quality of life. Another reason may be that continuous bronchodilation masks the worsening of the disease. This lack of awareness might in turn be caused by the continuous symptom relief of bronchodilators.

5.2 Introduction

The progressive nature of asthma and COPD indicates the relevance of studying the effects of drug therapy in an early stage, that is in patients with mild to moderate airflow obstruction. These patients are commonly treated with inhaled bronchodilator drugs. However, controlled studies of the long-term effects of these drugs in patients with a mild to moderate degree of airflow obstruction are scarce.¹⁻³ There are some retrospective or uncontrolled studies indicating that the continuous use of β_2 -adrenergic drugs may have adverse effects.^{4,5} The aim of this study was to address the long-term effects of continuous versus symptomatic bronchodilator treatment, as well as the long-term effects of the β_2 -adrenergic drug salbutamol versus the anticholinergic drug ipratropium bromide, in patients suffering from mild asthma or mild COPD. In this paper we report on lung function decline (FEV₁) and quality of life as well as on the relation between these two outcome measures. A detailed description of lung function in relation to other clinical and physiological outcome parameters is provided elsewhere.¹

Lung function parameters are essential for monitoring the course of asthma and COPD, because of their known relationship to disease prognosis.⁶ However,

the subjectively experienced quality of life may be equally important to the patients in defining treatment benefit. Since only a low correlation between quality of life and lung function parameters was found in previous studies,⁷⁻¹² one cannot rely on the severity of airflow limitation as an indicator of the impact of chronic airflow limitation on patients' quality of life.¹³ The measurements of the quality of life and of lung function may address different treatment aspects. So far, little data on the quality of life of patients suffering from a mild form of airflow limitation is available.¹² Therefore an additional objective of this paper is to compare the quality of life of patients with that of the general population, in order to determine the extent to which quality of life is affected by a mild degree of airflow obstruction.

5.3 Patients and methods

5.3.1 Patients

This study included both asthma and COPD patients aged 30 and over with a mild to moderate airflow obstruction ($FEV_1 \leq$ predicted value -2 standard deviations, but $FEV_1 \geq 50\%$ predicted value) and/or established bronchial hyperresponsiveness ($PC_{20} \leq 8$ mg/ml histamine). Patients were recruited from twenty-nine family practices.¹⁴ Those who were corticosteroid dependent or suffering from other pulmonary or life-threatening diseases were excluded. After giving informed consent and proving their capability of recording symptoms and peak flows, 223 patients entered the study. There was no selection bias.¹⁴ The study was approved by the Nijmegen University Medical Ethics Committee.

The criteria for the diagnosis of asthma and COPD were based on the standards for diagnosis of the American Thoracic Society.¹⁵ COPD was defined as a continuous airflow obstruction ($FEV_1 < 85\%$ of the predicted value for three measurements in one year) combined with chronic cough or chronic sputum production for at least three months a year, and in at least two successive years. Asthma was defined as a reversible obstruction (increase in $FEV_1 \geq 15\%$ of the initial value, 60 minutes after the inhalation of both 400 μ g salbutamol and 80 μ g ipratropium bromide, for three measurements in one year) in combination with bronchial hyperresponsiveness ($PC_{20} < 8$ mg/ml for three measurements in one year) and dyspnoea, wheezing and/or allergy. The combination of these features was mutually exclusive in the study population.¹

5.3.2 Study design

After a wash-out period of eight weeks, patients were randomly chosen to receive either continuous or symptomatic bronchodilator treatment. Within each of these parallel groups we randomly selected half the patients to receive salbutamol for one year followed by ipratropium bromide for the second year. The other half received the drugs in reverse order (see Fig. 1). This was a single blind study, with a blind observer.

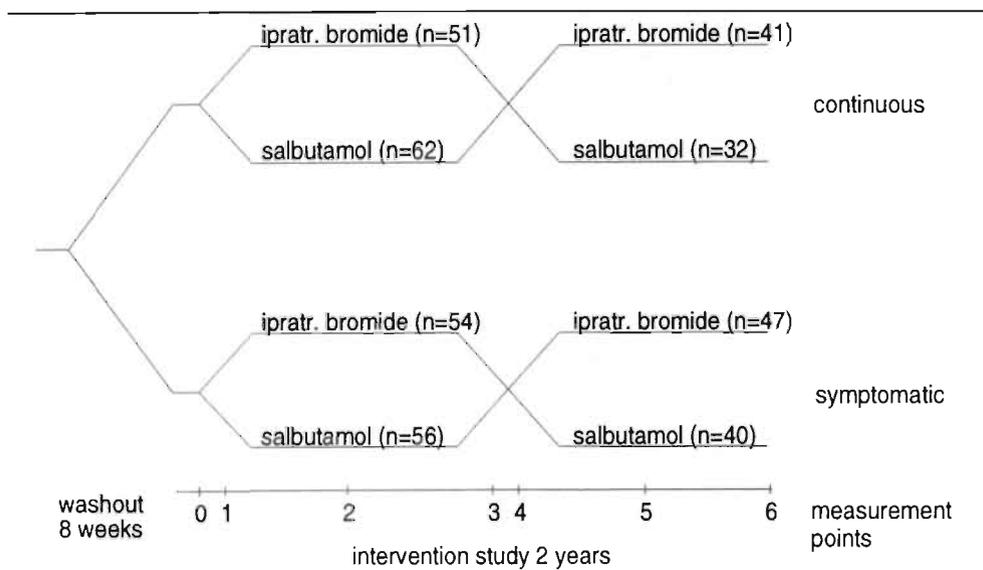


Figure 1. Study design. (n=number of patients at the start and at the end of the study)

In the continuously treated group, patients inhaled 4 dry powder inhalations of 400 μg salbutamol or 4 dry powder inhalations of 40 μg ipratropium bromide daily. Symptomatically treated patients only used medication during exacerbations or periods of dyspnoea, with a maximum of 4 dry powder inhalations per day. Exacerbations were defined according to Boman *et al.*¹⁶ and were treated in a standard way by the general practitioner: an antibiotic (broad spectrum) and a ten-day course of oral prednisone. Otherwise, no corticosteroids or bronchodilators other than the research medication were permitted. Patient compliance was maintained by regular telephone contact and home visits, and checked by counting the unused medication every three months.

5.3.3 Outcome measures

At baseline, lung function (FEV_1 , FEV_1/EVC), reversibility of airflow obstruction, bronchial hyperresponsiveness, allergy and quality of life were measured. Reversibility was measured 60 minutes after inhalation of 400 μg salbutamol followed by 80 μg ipratropium bromide. Bronchial hyperresponsiveness was defined according to Cockcroft.¹⁷ Allergy was tested by means of seven RAST-tests for common allergens. Patients were considered to be allergic if there was at least one positive RAST-test. Outcomes were assessed in terms of annual decline in FEV_1 and quality of life.

FEV_1 was measured at each visit (measurements 0 to 6 in Figure 1), according to a standard procedure, using the integrating flowmeter Microspiro HI-298^R.¹⁸ Data was derived from the curve with the largest sum of FVC and FEV_1 , out of three satisfactory measurements. Medication was discontinued for at least eight hours before the start of the FEV_1 measurement. The measurement was only performed when patients were exacerbation-free.

Quality of Life was measured at baseline and after the first and second year of treatment (measurements 1, 3 and 6 in Figure 1) using a Dutch version of the Nottingham Health Profile (NHP) and the Inventory of Subjective Health (ISH).

The NHP is a generic, self-administered questionnaire designed to measure perceived physical, emotional and social health problems.^{19,20} The emphasis is on the respondent's subjective perception of his or her health status. The NHP asks directly about feelings and emotions, not about changes in behaviour. It consists of two parts, only the first being used here. This contains 38 items relating to 6 dimensions: physical mobility (8 items), pain (8 items), social isolation (5 items), emotional reactions (9 items), energy (3 items) and sleep (5 items). The items were drawn from interviews with patients suffering from various acute or chronic diseases, and from other health questionnaires such as the Sickness Impact Profile. All items have a yes/no answer format. Their weights were derived by McKenna from a sample of both patients and non-patients using Thurstone's method of paired comparisons.²¹ Dimension scores range from 0 to 100: the higher the score the greater the perceived health problems. Separate NHP dimension scores are presented as a profile, not integrated into an overall score.

Whilst the NHP was originally developed as a survey instrument to measure perceived health status in a population, it has been used extensively in evaluation studies and is claimed to be sensitive to change in disease severity.²²⁻²⁴ The NHP has proved to be reliable (4-weeks test-retest ranges from 0.77-0.88) and can easily be administered, with small demands on patient time and effort.¹⁹

The Inventory of Subjective Health, developed by Dirken,²⁵ is a generic, commonly used Dutch scale which contains 21 questions related to subjective physical complaints such as tiredness, chest and heart problems, gastric problems, indigestion, headache etc. Most complaints can be grouped according to the organ system they refer to.²⁶ The remainder relate to overall physical condition. The ISH

items were partly drawn from the Cornell Medical Index, complemented by items drawn from expert interviews about the influence of physical stress on health. Its internal consistency and reliability are good (KR-20 ranges from 0.84-0.91, and 3-6 months test-retest = 0.67), and answers to the ISH do not appear to be influenced by social desirability.^{26,27} The overall ISH score is made up by the number of affirmative answers. The more physical complaints are reported, the higher the score.

5.3.4 Statistical analysis

A within-patient analysis comparing the two drugs could only be performed if it could be shown that there was no period effect or carry-over effect.²⁸ A period effect would occur if patients, on average, improve or deteriorate during the second year regardless of the drug used. A carry-over effect refers to the influence of the drug used in year 1 on the effect of the drug used in year 2. The continuously treated and symptomatically treated groups could only be compared, as groups, if no significant difference was found between salbutamol and ipratropium bromide.

FEV₁ slopes were all calculated by averaging individual slopes. The individual slopes were calculated with a regression analysis of the 7 assessments of FEV₁. A number of covariables, such as the number of pack-years before the study, number of cigarettes smoked during the study, sex, age, height, allergy, initial FEV₁ (as a percentage of the predicted value), initial PC₂₀ and drug use prior to the study were controlled for. Differences in slopes were tested by means of ANCOVA (analysis of covariance).

Spearman coefficients were calculated for the correlations between the FEV₁ and the quality of life at the start of the study and between the change in quality of life and the change in lung function (FEV₁ slope).

Since the quality of life data does not appear to be normally distributed, it was analyzed using non-parametric tests such as the Mann-Whitney U-test for the comparison of two independent groups (asthma versus bronchitis, continuous versus symptomatic treatment) and the Wilcoxon matched pairs signed rank test for within-patient comparisons of salbutamol versus ipratropium bromide.

5.4 Results

5.4.1 Baseline patient characteristics

Of the 223 patients who started the study, 144 patients completed the two-year period with cross-over medication. Seventy-nine patients dropped out, 56 because of pulmonary problems and 23 for reasons unrelated to lung disease or the medication being studied, such as lack of motivation (17), emigration (1), death (2) and the presence of malignant disorders (3). Those who dropped out for these other reasons did not significantly differ from those who completed the study. The 56 patients who dropped out because of pulmonary problems had a greater airflow obstruction, a greater bronchial hyperresponsiveness and were more likely to be asthmatics. Because of their pulmonary problems, these 56 patients were given additional medication (e.g. inhaled corticosteroids).

Table 1. Patient characteristics at baseline

Characteristic	completed the study	dropped out:	
		pulmonary problems	other reasons
Number	144	56	23
Sex (% male)	56	50	74
Asthmatic patients (%)	35	52*	43
Age (years)	52 (12)	54 (13)	55 (12)
(ex)smokers (%)	81	80	87
Allergic (%)	24	37	26
FEV ₁ (l)	2.41 (0.79)	2.12 (0.64) *	2.34 (0.59)
Reversibility obstruc. (%)	16 (18)	21 (20)	14 (10)
Geometric PC ₂₀ (mg hist/ml)	8	6*	8
NHP score energy	23.6 (31.3)	24.6 (33.1)	32.1 (33.5)
NHP score pain	11.3 (22.1)	12.8 (22.3)	23.9 (31.4)
NHP score sleep	21.0 (28.1)	19.2 (25.0)	26.7 (29.6)
NHP score emotional reaction	12.7 (19.4)	12.7 (17.8)	14.8 (17.0)
NHP score social isolation	8.1 (19.7)	6.3 (12.9)	12.0 (22.0)
NHP score physical mobility	9.2 (14.5)	9.0 (13.8)	13.5 (15.8)
ISH score	7.0 (4.5)	7.1 (4.8)	9.0 (5.7)

Averages and standard deviations (in parentheses). Differences between the group that completed the study and the two groups of drop-outs were tested (* $p < 0.05$). The ISH and NHP scores for those who dropped out for other reasons seem to be somewhat higher, although not significantly so. When the 3 patients with malignant disorders and the 2 patients who died are excluded, the mean ISH and NHP scores in this group decrease, although their scores still remain somewhat higher.

As for the characteristics of the 144 patients who completed the study with cross-over medication, there were no significant differences between the four treatment groups, except for smoking history. Continuously treated patients receiving salbutamol followed by ipratropium bromide had a significantly higher number of pack years than symptomatically treated patients receiving salbutamol followed by ipratropium bromide (t value -2.113 , $p=0.038$). This difference was not caused by selective drop-out.

With respect to age, sex, FEV₁, and ISH score, asthma patients did not differ significantly from COPD patients. Of the 93 COPD patients, 87% were smokers or ex-smokers, compared to 71% of the 51 asthmatic patients (χ^2 5.892, $p=0.015$). COPD patients experienced more health problems with respect to the NHP dimensions of sleep (z value -2.3632 , $p=0.0181$) and pain (z value -2.0191 , $p=0.0435$) than patients with asthma. No significant differences between asthma and COPD patients were found for the remaining NHP dimensions.

5.4.2 Baseline quality of life impairment compared to the general population

Baseline scores on the 6 NHP dimensions are summarized in figure 2. No data on quality of life impairments in asthma and COPD measured by means of this instrument have been published before. Therefore we present mean NHP dimension scores for patients alongside general population means, although formal statistical comparisons could not be made. With due reserve, we used NHP reference data from a random sample of 1297 people from the Nottingham area,¹⁹ since for the Netherlands no general population NHP scores were available. As reference material we also present NHP scores of 1366 elderly persons (mean age 74; sd 7.8; 55% female) randomly selected from the registers of family physicians in a Dutch province.²⁰

Mean NHP score

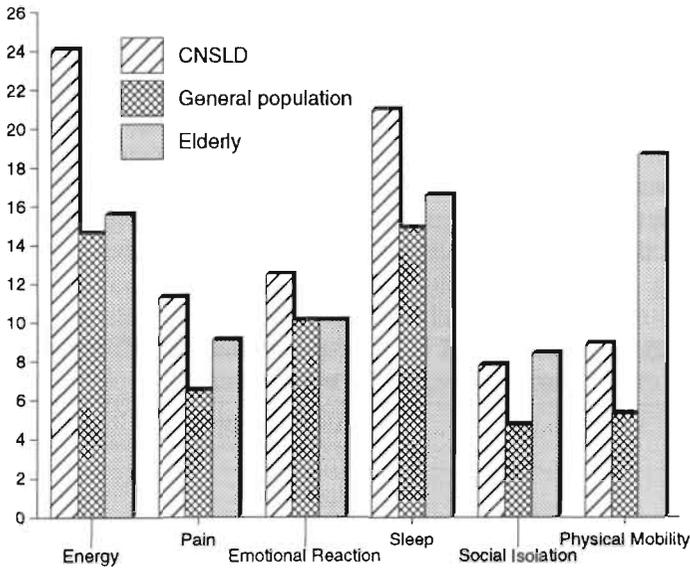


Figure 2. Mean scores on the dimension of the Nottingham Health Profile for the study population ($n=144$), the general population ($n=1297$) and the elderly ($n=1366$)

When compared to a general population, there is an obviously impaired quality of life in the study population. Energy and sleep scores in particular were worse (Figure 2). Patients with mild chronic airflow limitation also experience more health problems on all NHP dimensions than elderly, except for the aspects of social isolation and physical mobility. Eighty per cent of the study population reported one or more problems on the NHP, while in the general population only 46% did so.²⁹ The NHP item most frequently affirmed is "I soon run out of energy". It was affirmed by nearly 60% of our respondents reporting at least one problem, compared to 25% of the general population respondents reporting at least one problem. The next most frequently reported problems are, "I'm waking up in the early hours of the morning" (51% vs 59%), "I'm feeling on edge" (40% vs 17%), "I find it hard to stand for long" (39% vs 17%), "It takes me a long time to get to sleep" (33% vs 27%), "I sleep badly at night" (31% vs 22%), "I lose my temper easily these days" (30% vs 28%) and "I'm tired all the time" (26% vs 16%). The general population reference percentages were obtained by Kind *et al.* from a sample of 1598 residents (49% male) of York and Yorkshire, a sample

which - like our study population - was weighted towards middle and old age.²⁹ The least impairment is reported with regard to physical mobility, a finding which is undoubtedly related to the severity of the items in this dimension.

Much the same picture can be drawn from the ISH score. The mean ISH score of 7.6 (sd of 4.3) for the study population is almost twice the general population mean of 3.3.³⁰ Patients' baseline responses to the eight most frequently affirmed questions of the ISH are presented in Table 2. As can be seen, the questions most frequently affirmed relate to problems of the chest and tiredness. For all items a higher score was obtained in our study population than in a random sample of 22,000 people of the non-institutionalized general population over 16 years of age.²⁶

Table 2. Inventory of Subjective Health: percentage of affirmative responses to the eight most frequently affirmed items

Items	study population		general population	
	% positive responses	rank	% positive responses	rank
Do you often feel tired?	63.2	1	25.6	2
Do you get short of breath easily?	55.9	2	17.2	6
Do your bones or muscles ever ache?	52.8	3	22.8	4
Do you get tired sooner than you would consider normal?	50.0	4	17.0	8
Are you often troubled by back-ache?	44.8	5	29.0	1
Do you often feel sleepy or sluggish?	43.7	6	18.3	5
Do you often feel tight in the chest?	43.1	7	9.4	17
Do you generally get up feeling tired and unrested in the morning?	38.9	8	17.2	6

5.4.3 Effect of treatment on lung function and quality of life: Salbutamol versus ipratropium bromide

Salbutamol and ipratropium bromide had a similar effect on the decline in FEV₁, both in asthma and COPD patients. No period or carry-over effect on the FEV₁ was found. NHP dimension scores and the ISH score were not significantly influenced by a period effect either, except for scores for social isolation. As for social isolation, patients in the continuously treated group on average deteriorated during the second year, irrespective of the drug used ($z = -2.068$, $p = 0.0386$). Since a period effect only occurred in one NHP dimension and no carry-over effect was found, it was considered legitimate to compare salbutamol and ipratropium bromide with regard to the NHP and the ISH. In both the continuous and the symptomatic treatment group, no significant differences in quality of life were

found between salbutamol and ipratropium bromide. This finding holds for both asthma and COPD patients.

5.4.4 Effect of treatment on lung function and quality of life: Continuous versus symptomatic

Figure 3 shows that the decline in FEV_1 in continuously treated patients was 86.10^{-3} l/yr compared to 29.10^{-3} l/yr in symptomatically treated patients. After correction for confounding factors, the FEV_1 declined 72.10^{-3} l/yr in the continuous group compared to 20.10^{-3} in the symptomatic group ($p < 0.05$), irrespective of the drug used. The difference in decline between continuously and symptomatically treated patients is comparable for asthma (92.10^{-3} versus 25.10^{-3} l/yr) and COPD (82.10^{-3} versus 31.10^{-3} l/yr).

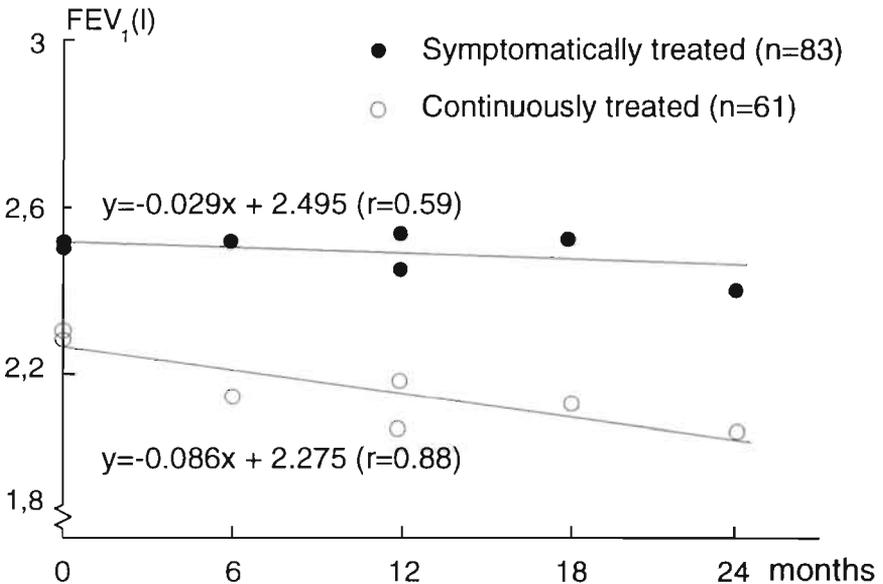


Figure 3. Decline of FEV_1 : comparison between the symptomatic ($n=83$) and continuous treatment group ($n=61$)

In contrast to the observed difference in decline in lung function, there was no significant difference between the continuous and symptomatic treatment groups with respect to the changes in NHP dimension scores and the ISH score after one and after two years of treatment (Table 3), for either asthma or COPD. Although there seems to be a tendency towards improved quality of life in the symptomatic treatment group (a negative change indicates improvement) and a tendency towards deteriorating quality of life in the continuous treatment group, this difference is not significant. Apparently, the clinically significant difference in lung function decline was not reflected in a detectable difference in quality of life impairment. There is one exception: for the change in NHP pain score between baseline and the two-year follow-up period, there is a significant difference between the continuous and the symptomatic treatment group ($z = -2.2162$, $p = 0.0267$). In the continuous treatment group, the patients deteriorated significantly on the pain dimension ($z = -2.9431$, $p = 0.0032$), while in the symptomatic treatment group patients remained unchanged on the pain dimension ($z = -0.8548$, $p = 0.3926$). The same tendency was also found after one year, but it was not significant. The difference between the continuous and symptomatic treatment group on the pain dimension is comparable for asthma and COPD. Although all analyses were made using non-parametric statistics, table 3 gives mean NHP scores to ease the interpretation.

Table 3. Mean change in NHP and ISH scores: comparison between continuous and symptomatic treatment group[†]

	mean change in NHP and ISH scores (2 year follow-up minus baseline)		
	symptomatic	continuous	p-value
NHP: Energy	-0.8 (23.5-24.3)	1.7 (23.6-21.9)	0.5214
Pain	-2.1 (11.9-14.0)	5.9 (13.5-7.7)	0.0267
Emotional reaction	-1.1 (11.8-13.0)	-0.1 (12.1-12.3)	0.1582
Sleep	-2.1 (19.5-21.6)	2.0 (22.4-20.3)	0.3806
Social isolation	-0.2 (8.1-8.3)	0.0 (7.8-7.8)	0.6477
Physical mobility	-0.3 (6.9-6.4)	0.9 (10.5-9.6)	0.6132
ISH:	-0.5 (6.4-6.9)	0.3 (7.4-7.1)	0.1949

[†] Differences between the change in the symptomatic group and the change in the continuous group were tested

5.4.5 Within-patient comparison of lung function and quality of life

Table 4 shows that the correlations between the FEV₁ and the quality of life at the start of the study were very low ($|r| \leq 0.28$). In the symptomatic treatment group, pain and emotional reaction were significantly correlated with lung function. In continuously treated patients none of the items was correlated with lung function.

When the change in quality of life was related to the lung function decline (FEV₁ slope), the correlation was again low ($|r| \leq 0.32$). Neither the change in the ISH score nor the changes in NHP dimension scores correlated significantly with the lung function decline, except for emotional reaction in the continuous treatment group (Table 4). There was no difference between asthma and COPD in this respect.

Table 4. Correlations between (change in) FEV₁ and (change in) quality of life

FEV ₁ with	Correlation between baseline values		Correlation between two year change values	
	r	p	r	p
<i>Symptomatic treatment (n=83)</i>				
NHP: Energy	-0.09	0.44	0.01	0.95
Pain	-0.28	0.01	0.10	0.35
Emotional reaction	-0.27	0.02	0.03	0.78
Sleep	-0.10	0.37	0.05	0.66
Social isolation	-0.16	0.15	-0.12	0.27
Physical mobility	-0.12	0.28	0.01	0.94
ISH:	-0.11	0.32	-0.09	0.43
<i>Continuous treatment (n=61)</i>				
NHP: Energy	-0.10	0.45	0.09	0.51
Pain	-0.03	0.82	0.12	0.37
Emotional reaction	0.11	0.41	0.32	0.02
Sleep	-0.08	0.55	-0.16	0.24
Social isolation	-0.04	0.77	0.10	0.47
Physical mobility	-0.18	0.16	0.04	0.76
ISH:	0.01	0.99	0.03	0.79

5.5 Discussion

When compared to general population reference groups, the NHP and ISH results show that patients suffering from light to moderate airflow obstruction experience substantially more health problems. Patients with this degree of disease severity are commonly treated with an inhaled bronchodilator drug, either symptomatically or continuously. The results of this study show that continuous

bronchodilator treatment does not seem to protect against decline in lung function as has been suggested.^{31,32} On the contrary, the annual decline in FEV₁ for continuously treated patients was found to be significantly higher than for those treated symptomatically, even after correction for confounding variables. The difference in the rate of decline was not caused by any difference in the characteristics of the treatment groups at the start of the study.¹ Increased decline during continuous treatment might be explained by bronchodilators failing to influence the inflammatory processes underlying the disease. A long-term continuous bronchodilator treatment without anti-inflammatory treatment may thus lead to adverse effects.^{2,3}

However, the decline in lung function is not reflected in perceived quality of life as it was measured in our study, except for the pain dimension of the NHP. On eight of the nine items which make up the pain dimension, the number of affirmative responses in the continuous group increased during the trial. The highest and only significant increase of 13.1% was found for "I have pain at night" ($p=0.021$). We checked whether this increase (which was similar for asthma and COPD) was related to an increase in coughing, which in turn might have caused an increase of chest-related pain at night. However, no such increase in coughing was reported on the MRC-questionnaire. Thus, the increase in pain-related complaints in the continuous group may be due to chance, as a result of multiple testing. This is confirmed by the fact that the change in the NHP pain dimension is not related to the change in lung function in either continuously or symptomatically treated patients ($r=0.03$ and 0.15 respectively).

If a decline in lung function leads directly to a lower quality of life, then the deterioration should also have been found for dimensions other than pain, because at baseline about 24.5% of the patients in the continuous group reported no health problem on the NHP at all and only 14.5% affirmed 11 or more items. This indicates that a deterioration, if there was any, would have been easy to measure.

The decision to use generic outcome measures such as the NHP and the ISH may be responsible for our not measuring a deterioration in the quality of life. Williams argues that generic instruments such as the NHP may be too much oriented toward musculo-skeletal problems and disabilities which are often more extreme in nature.³³ Disease-specific quality of life scales, which are often more sensitive to change, might have reflected the declined lung function in this study. But, although important for measuring quality of life impairment directly related to the disease, their focus may be too narrow to capture other, more general aspects of the overall health status. As a result they do not allow comparisons between various health care interventions, which are possible using generic instruments. There is, as a consequence, a growing consensus in family medicine to use standardised generic quality of life measures.³⁴ The inclusion of severe problems in a generic quality of life instrument, however, may be necessary in order to avoid the detection of large numbers of false positives.³³ This does not imply that the more subtle impairments and disabilities from which asthma and COPD patients suffer cannot show up on

the profile, for the NHP was also successful in drawing a profile of migraine sufferers, which gives an indication of its ability to detect impairment, even if motion and mobility are not affected.²⁴ Sensitivity to change was also found to be a concomitant of, for example, the recovery from fractures,³⁵ the physical and emotional changes throughout pregnancy¹⁹ and the effects of minor surgery.³⁶

The ISH is particularly directed at chronic physical problems, since it contains few items related to temporary complaints. A strong relationship between the ISH score and the presence, number and nature of chronic diseases has been found earlier.²⁶ Frequently affirmed questions relate to fatigue, reduced energy, reduced vitality, dyspnoea and other chest problems. These complaints are generally considered to be related to asthma and COPD. Affirmative answers are also frequently given to the pain-related questions of the ISH (see Table 2). This may be explained by the less severe nature of the pain questions rather than by an inclination to complain. For reduced physical activity and vitality caused by fatigue and by dyspnoea itself or by the ever-present fear of dyspnoea lead to a poor physical condition which in turn causes minor pains and discomfort. Therefore, it seems likely that the ISH measures physical, mostly organ-specific problems directly or indirectly related to having a particular disease. But why then do we not find a significant decline in the quality of life?

It may be that because of the rapid bronchodilating response in the day-by-day control of symptoms, continuous bronchodilation masks the ongoing decline in lung function and the deterioration of the disease. Patients with continuous bronchodilator therapy may tolerate precipitating situations better, and may as a result become more exposed to sensitizing agents.^{1,2,37} Continuous bronchodilation without anti-inflammatory treatment may therefore mask the worsening of the disease so that patients are not aware of it. This was confirmed by the finding that, as compared to symptomatic treatment, we could not detect any increase in asthma and COPD-related symptoms as measured by means of the MRC-questionnaire.¹ A time-lag between lung function decline and deterioration in the quality of life may cause a decline in lung function only experienced by the patients in the long run. Such an explanation is consistent with the study by Labuhn, that indicates that physiological factors are more predictive of quality of life for patients with advanced COPD as for patients with mild to moderate COPD.³⁸

In addition, patients may learn to live with their disease, particularly those suffering from chronic diseases such as asthma and COPD. After an adjustment period their limitations seem normal. In that case a significant decline in lung function may have no direct effect on *experienced* quality of life. As we -and many authors before us- have found, the correlation between various measures of pulmonary function and quality of life appears to be relatively weak.⁷⁻¹² Other factors such as psychological and sociological aspects may serve as intervening factors between lung function and quality of life. The patient's attitude, expectations, emotions etc. may not allow declining lung function to be fully expressed in terms of a reduced quality of life.

The effects of continuous versus symptomatic bronchodilator treatment in terms of FEV₁ and quality of life do not point in the same direction. Lung function and overall quality of life seem to be two different and weakly related expressions of asthma and COPD. Therefore we feel that meaningful outcome assessments should include both lung function measures and various disease-specific quality of life questionnaires as well as the overall evaluation of quality of life, for improvement in the latter is one of the important goals of health care. The rationale for disease-specific instruments, each covering one or more aspects of quality of life directly related to having a particular disease, lies in their potential for increased responsiveness. But particularly when the results of these measures do not point in the same direction, an overall evaluation by means of generic quality of life instruments is desirable. To increase confidence in the use of generic quality of life measures in therapeutic studies, their relationship with lung function and the prognosis of the disease in the long run, and their relationship with disease-specific quality of life measures will have to be further explored. It seems warranted therefore to apply both generic and disease-specific quality of life measures alongside physiological measures in future studies.

Since an accelerated decline in lung function is known to be an objective physiological measure of progression of the disease, the main conclusion of this study is that continuous bronchodilator treatment is not as beneficial as is often assumed. Overall quality of life measurements suggest that continuous treatment leads to a lack of awareness of a declining lung function. This may be misleading to both patients and physicians and may have negative effects on future compliance with medication.

5.6 Acknowledgements

The authors wish to thank the Dutch Asthma Foundation, Boehringer Ingelheim Netherlands and Glaxo Netherlands for their financial support. Mrs. L. Bierman and Mrs. A. Raaymakers were of great help in measuring lung function parameters. We are also greatly indebted to all the general practitioners and patients who took part in this study. Finally we would like to thank E. Dompeling, H. Vondeling, M. Essink-Bot and P.N.R. Deckhuijzen for their helpful comments on earlier drafts, P.G.H. Mulder for his statistical advice and R.P. Akkermans and J.P.M.A. van Emmerik for their computer assistance.

5.7 References

- 1 Van Schayck CP, Dompeling E, van Herwaarden CLA, Folgering H, Verbeek ALM, van der Hoogen HJM, van Weel C. Bronchodilator treatment in moderate asthma or chronic bronchitis: continuous or on demand? *Br Med J* 1991; 303:1426-31
- 2 Sears MR, Taylor DR, Print CG, Lake DC, Li Q, Flannery EM, Yates DM, Lucas MK, Herbison GP. Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* 1990; 336:1391-6
- 3 Van Schayck CP, Graafsma SJ, Visch MB, Dompeling E, van Weel C, van Herwaarden CLA. Increased bronchial hyperresponsiveness after inhaling salbutamol during one year is not caused by subsensitization to salbutamol. *J Allergy Clin Immunol* 1990; 86:793-800
- 4 Sears MR, Rea HH, Fenwick J, Gillies AJD. 75 Deaths in asthmatics prescribed home nebulizers. *Br Med J* 1987; 294:477-80
- 5 Crane J, Flatt A, Jackson R, Ball M, Pearce N, Burgess C, Kwong T, Beasley R. Prescribed fenoterol and death from asthma in New Zealand, 1981-83: case-control study. *Lancet* 1989; 8644:917-22
- 6 Postma DS, de Vries K, Koeter GH, Sluiter HJ. Independent influence of reversibility of airflow obstruction and non-specific hyperreactivity on the long-term course of lung function in chronic airflow obstruction. *Am Rev Respir Dis* 1986; 134:276-80
- 7 Williams SJ, Bury MR. Impairment, disability and handicap in chronic respiratory illness. *Soc Sci Med* 1989; 29:609-16
- 8 Jones PW. Quality of life measurement in chronic lung disease. *J Drug Ther Res* 1988; 13: 181-3
- 9 McSweeney AJ, Heaton, RK, Grant I, Cugell D, Solliday N, Timms R. Chronic obstructive pulmonary disease; socioemotional adjustment and life quality. *Chest* 1980; 77:309-11
- 10 Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42:773-8
- 11 Kaptein AA, Dekker FW, Dekhuijzen PNR, Wagenaar JPM, Janssen PJ. The value of the 12-minutes walking test in patients with chronic obstructive pulmonary disease, II walking in the laboratory and walking in daily life (in Dutch). *Ned Tijdschr Geneesk* 1987; 131:1717-21
- 12 Schrier AC, Dekker F, Kaptein AA, Dijkman JH. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990; 89:894-9
- 13 Guyatt GH, Townsend M, Berman L, Pugsley SO. Quality of Life in patients with chronic airflow limitation. *Br J Dis Chest* 1987; 81: 45-54
- 14 Van Schayck CP, van Weel C, Folgering H, Verbeek ALM, van Herwaarden CLA. Treatment of patients with airflow obstruction by general practitioners and chest physicians. *Scand J Prim Health Care* 1989; 7:137-42
- 15 American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *Am Rev Respir Dis* 1987; 136:225-43
- 16 Boman G, Bäcker U, Larsson S, Melander B, Wählander L. Oral acetylcysteine reduces exacerbation rate in chronic bronchitis: report of a trial organized by the Swedish society for pulmonary diseases. *Eur J Respir Dis* 1983; 64:405-15
- 17 Cockcroft DW, Killian DN, Mellon JJA, Hargrave FE. Bronchial reactivity to inhaled histamine; a method and clinical survey. *Clin Allergy* 1977; 7:237-43

- 18 Dompeling E, van Schayck CP, Folgering H, van den Hoogen HJM, van Weel C. Accuracy, precision and linearity of the portable flow-volume meter Microspiro HI-298. *Eur Respir J* 1991; 4:612-5
- 19 Hunt SM, McEwen J, McKenna SP. *Measuring Health Status*, 1st ed. London, Sydney, Dover, New Hampshire: Croom Helm, 1986
- 20 Van Eijk JTHM, Smits AJA, Meyboom W, Mokkink H, van Son J. The Nottingham Health Profile in the Dutch situation (in Dutch). *Allgemeinmedizin* (accepted for publication)
- 21 McKenna AP, Hunt SM, McEwen J. Weighting the seriousness of perceived health problems using Thurstone's method of paired comparisons. *Int J Epidem* 1981; 10:93-7
- 22 McDowell I, Newell C. *Measuring Health. A guide to rating scales and questionnaires*, 1st ed. Oxford, New York: Oxford University Press, 1987; 285-90
- 23 Hunt SM, McEwen J, McKenna SP. *Measuring Health Status: a new tool for clinicians and epidemiologists*. *J Royal Coll Gen Pract* 1985; 35:185-8
- 24 Jenkinson C, Fitzpatrick R, Argyle M. The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. *Soc Sci Med* 1988; 27:1411-4
- 25 Dirken JM. *Arbeid en Stress. Het vaststellen van aanpassingsproblemen in werksituaties*. Thesis. Groningen: J.B. Wolters, 1967
- 26 Van Sonsbeek JLA. *De VOEG: klaaglijt of lijst met gezondheidsklachten? Statistische onderzoeken*, Netherlands Central Bureau of Statistics M37, 1990
- 27 Visser APH. *De betekenis van de VOEG: enkele gegevens over de begripsvaliditeit*. *G & S*, 1983; 3:177-88
- 28 Pocock SJ. *Clinical Trials. A practical approach*. London: John Wiley & Sons, 1984; 110-22
- 29 Kind P, Carr-Hill R. The Nottingham Health Profile: a useful tool for epidemiologists? *Soc Sci Med* 1987; 8:905-10
- 30 Netherlands Central Bureau of Statistics. *Health interview surveys. Trends of health indicators: 1970-1984. Monthly bulletin of health statistics July 1986*: 5:5-15
- 31 Rebeck AS, Chapmann KR. Asthma: 2. Trends in pharmacologic therapy. *Can Med Assoc J* 1987; 136:483-8
- 32 Sinclair BL, Clarke DWJ, Sears MR. Use of anti-asthma drugs in New Zealand. *Thorax* 1987; 42:670-9
- 33 Williams SJ. Chronic respiratory illness and disability: a critical review of the psychosocial literature. *Soc Sci Med* 1989; 8:791-803
- 34 Van Weel C, Rosser WW. Measuring functional status in family practice. *Family Practice* 1991; 8:394-5
- 35 McKenna SP, McEwen J, Hunt SM. Changes in the perceived health of patients recovering from fractures. *Public Health* 1984; 98:97-102
- 36 Hunt SM, McKenna SP, McEwen J. Subjective health assessments and the perceived outcome of minor surgery. *J Psychosom Res* 1984; 28:105-14
- 37 Barnes PJ. Effect of corticosteroids on airway hyperresponsiveness. *Am Rev Respir Dis* 1990; 141(suppl):70-6
- 38 McSweeney AJ, Labuhn KT. Chronic obstructive pulmonary disease. In: Spilker B, ed. *Quality of Life Assessments in Clinical Trials*. 1st ed. New York: Raven Press, 1990; 391-417

6

Comparing the performance of four different instruments in evaluating the effects of salmeterol on asthma quality of life

Maureen P.M.H Rutten-van Mölken¹, Frank Custers², Eddy K.A. Van Doorslaer³, Margreet C.M. Jansen¹, Linda Heurman¹, Francis P.V. Maesen², Sjef J. Smeets², Angeline M. Bommer⁴, Jan A.M. Raaijmakers⁴.

1 Department of Health Economics, University of Limburg.

2 Department of Respiratory Medicine, De Wever Hospital, Heerlen.

3 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

4 Glaxo B.V., Zeist.

Forthcoming as: Comparing the performance of four different instruments in evaluating the effects of salmeterol on asthma quality of life. *European Respiratory Journal. Reprinted with permission of the publisher.*

6.1 Abstract

This study assesses the sensitivity to change of four different quality of life instruments as well as their reliability and construct validity. In a double blind, parallel group study 120 moderate asthma patients aged between 18 and 70 received either inhaled salmeterol 50 μg bid or inhaled salbutamol 400 μg bid. Besides respiratory outcomes, quality of life was measured at 6 weeks' follow-up using: 1) Asthma Quality of Life Questionnaire (AQLQ), 2) Living with Asthma Questionnaire (LWAQ), 3) Sickness Impact Profile (SIP), 4) rating scale (RS) and standard gamble (SG) utilities.

Salmeterol led to significant improvements over salbutamol on virtually all clinical outcomes. Although all the quality of life instruments showed the same trend in favour of salmeterol, only the AQLQ and RS utilities showed significantly greater improvement on salmeterol than on salbutamol. Except for the AQLQ the correlation between change in lung function and change in quality of life was generally low. Whereas the AQLQ correlated well with the patient's overall assessment of efficacy ($r=0.64$; $p < 0.001$), the LWAQ, SIP and utilities failed to show such a correlation. The AQLQ showed the best correlation with symptom scores. The cross-sectional correlation between the AQLQ and the LWAQ was 0.73 whereas the longitudinal correlation was only 0.29. The SG generally showed poor correlation with other measures including the RS. Test-retest reliability was relatively high for the AQLQ, the LWAQ, and the SIP, but lower for the utilities.

In conclusion, patients given salmeterol showed a greater improvement in quality of life compared to patients given salbutamol. Both the AQLQ and the LWAQ are relatively reliable disease-specific questionnaires, with the AQLQ being more responsive to change and showing greater validity. Of the generic instruments, the SIP showed better reliability, but the RS utilities were more responsive. The SG was less reliable and showed poor correlation with other measures.

6.2 Introduction

One of the most important goals of asthma management is to improve a patient's everyday functioning and subjective well-being, often referred to as quality of life. Improvement in quality of life can be achieved through improvement of respiratory function and reduction of symptoms and exacerbations. Whereas the impact of treatment on respiratory functioning can be measured exactly, the assessment of change in quality of life is often based on the physician's intuition about the relevance of lung function changes to a patient's daily functioning. Until recently, the measurement of quality of life by means of questionnaires had not gained widespread acceptance in asthma research. A bibliography of quality of life

measures published in 1990 did not include a single asthma-specific quality of life questionnaire.¹ However, quality of life measures are increasingly being proposed as important efficacy endpoints, especially in studies of drugs.² Quality of life monitoring is important because it reflects patients' concerns and perceptions and it can be one of the predictors of compliance. Moreover, the relatively weak correlations between lung function and quality of life found in a number of earlier studies,³⁻⁸ indicate that quality of life outcomes focus on rather different aspects of a disease than the physiological outcomes. Together, these outcomes may give a complete picture of a patient's health status.

Recently, a number of asthma-specific quality of life measures for use in clinical trials have been proposed.⁹⁻¹⁴ Assessing the measurement properties of these instruments, such as validity, reliability, and responsiveness, is an ongoing process that has to be repeated in different countries, patient populations and treatment settings. Based on the current knowledge of measurement properties, none of the recently developed asthma quality of life questionnaires can be said to be superior in particular circumstances.

In this study, we evaluated the measurement properties of the Dutch versions of four quality of life questionnaires in a clinical trial comparing salmeterol with salbutamol. Two of the questionnaires -the Living with Asthma Questionnaire (LWAQ) and the Asthma Quality of Life Questionnaire (AQLQ)- were disease specific, whereas the other two -the Sickness Impact Profile (SIP) and the Patient Utilities- were generic. The advantage of incorporating both types of instruments into one study is that we can both make a detailed study of those aspects of quality of life that are characteristic for asthma as well as a global assessment of overall quality of life. Such an overall assessment enables the comparison of results across different interventions for different diseases.

In the present study, we were primarily interested in the responsiveness of the four questionnaires to clinical improvements that were expected to result from salmeterol compared to salbutamol. Salmeterol is a new long-acting β_2 -agonist which, when compared to either placebo or salbutamol, has been shown to increase clinic lung function and both morning and evening peak flow, reduce diurnal variation in peak flow, daytime and nocturnal asthma symptoms and exacerbations as well as the need for additional bronchodilator rescue drugs.¹⁵⁻¹⁸ Furthermore, salmeterol has been shown to improve the duration and quality of sleep.¹⁹ This, in turn, may be associated with better daytime functioning and well-being.

All of these clinical effects of salmeterol may be expected to improve quality of life. Thus far, few studies have addressed this question. Ullman *et al.* found a non-significant trend towards a better sense of well-being during salmeterol as compared to salbutamol treatment.¹⁵ A study by Palmer and Hyland showed that the addition of salmeterol instead of a placebo to a patient's standard therapy improved the quality of life, as measured with the LWAQ.²⁰ However, more than 50% of the patients were excluded from this study because they did not complete valid

responses to the questionnaire. Juniper *et al.* showed that scores on the AQLQ were significantly better on salmeterol than on a placebo or salbutamol.²¹

Besides the ability of the four quality of life questionnaires to detect clinically important differences, we also looked into their longitudinal and cross sectional construct validity.²² Both types of validity were addressed, because the questionnaires can both be used to measure within-patient changes as well as to measure differences between patients with various degrees of disease severity. Furthermore the reproducibility in stable subjects was established. Reproducibility is related to responsiveness in that a large variation between replicate measurements (low reliability) may contribute to failure to find an effect.²² Assessing these psychometric properties may aid future selection of measures, assist in prioritizing outcomes and permit more accurate sample size calculations.

6.3 Methods

6.3.1 Patients

Subjects were outpatients with stable asthma, aged between 18 and 70, who consented to participate in the study. On entering the study they were required to have a forced expiratory volume in 1 second (FEV₁) of between 50 and 70% of predicted normal values, not having used bronchodilators for at least 6 hours previously. The FEV₁ had to increase by at least 15% after inhalation of 400 µg salbutamol. Patients were excluded if they had, or had in the past suffered, a serious concomitant disease, an upper or lower respiratory tract infection during the 6 weeks prior to the study, used a booster course of oral steroids during the 6 weeks prior to the study, were likely to change their concurrent medication for diseases other than asthma or had a poor understanding of Dutch. The use of theophyllines, other β-agonists apart from study medication, β-blockers, methylxantines or a booster course of corticosteroids was not allowed. All other maintenance drugs were continued at constant dosage throughout the study.

6.3.2 Study design

The study had a double-blind, randomized, parallel group design. After a 2 week run-in period, 120 patients eligible for treatment were randomized to receive 6 weeks of treatment with either salmeterol 50 µg twice daily or salbutamol 400 µg twice daily. Both drugs were administered as dry powder via the diskhaler device. Rescue salbutamol in a diskhaler device was supplied for symptomatic relief. Patients were seen at the end of the run-in period, after 3 weeks and 6 weeks of treatment and 2 weeks after cessation of treatment. Quality of life measurements

were conducted at the end of the run-in period and after 6 weeks of treatment. The study was approved by the Medical Ethics Committee of the Wever Hospital.

6.3.3 Clinical outcome measures

FEV₁ and forced vital capacity (FVC) were measured at each clinic visit. They were obtained from the best of 3 measurements using a Sensor Medics Pulmograph. Prior to these clinic measurements the patients withheld inhaled bronchodilators for 12 hours. Reversibility was calculated as the change in FEV₁ after inhaling 400 µg divided by the prebronchodilator FEV₁ and expressed as a percentage. Peak Expiratory Flow Rate (PEFR) measurements were obtained using the mini-Wright PEFR meter each day on awakening and in the evening before inhaling the study medication. Patients kept a diary to record the number of daytime and nighttime rescue salbutamol blisters and to record daytime and nighttime asthma symptoms. The daytime symptom score documented the frequency and severity of symptoms on a scale from 0 (no symptoms during the day) to 5 (symptoms so severe that you could not go to work or perform normal daily activities). The nighttime symptom score ranged from 0 (no symptoms during the night) to 4 (symptoms so severe that you did not sleep at all).

An overall assessment of efficacy was obtained at each clinic visit by asking the patients how their asthma/breathing had been since the last visit, using a 5 point response scale ranging from much improved to much worse.

6.3.4 Quality of life instruments

All four quality of life measures used in this study originate from the English language and were translated into Dutch. The advantage of translating existing measures over developing new measures is that it provides a basis for comparison of investigations of quality of life across different countries and cultures.²³ To a large extent the translations were done according to recently published methods for good quality translations.^{23,24}

Asthma Quality of Life Questionnaire

The AQLQ was developed by Juniper *et al.*¹⁴ The questionnaire is interviewer-administered and designed for use in clinical trials involving adult asthma patients. It contains 32 questions in four domains: activity limitation (11 items), symptoms (12 items), emotional function (5 items) and environmental stimuli (4 items). The AQLQ is made up of questions which ask patients to think about the last 2 weeks and quantify, either in terms of frequency, duration, or amount, a number of asthma-related problems. At follow-up patients are shown the answers they have given at baseline. The first 5 questions in the activity limitation domain are 'patient-

specific'. This means that at baseline each patient is asked to state which 5 of his or her regular activities are most troublesome because of asthma. At follow-up the patient is again asked how severe the limitation is in doing these same 5 activities. Answers to each of the 32 items can be scored on a 7-point scale ranging from 1, which indicates maximal impairment, to 7, which indicates no impairment. The results are expressed as the mean score for each domain as well as for overall quality of life. The AQLQ was found to be valid, reproducible, and responsive to change in time.^{8,21,25,26}

The translation procedure of the AQLQ started with forward translations by 2 independent translators who were aware of the objectives of the instruments and the concepts involved. After consensus was reached between the 2 translators and the researchers about the forward translation a backtranslation was performed by a bi-lingual professional translator not involved in health care research. To check the face validity the questionnaires were completed by a number of patients who were encouraged to comment on how they interpreted the questions. If needed, alterations were made in line with the suggestions.

Living with Asthma Questionnaire

The LWAQ was developed by Hyland and colleagues.¹⁰ This questionnaire was translated according to a similar procedure as the AQLQ which was extended by the use of several backtranslations. The LWAQ consists of a list of 68 statements covering 11 domains: social/leisure, sport, holidays, sleep, work/other activities, colds, mobility, effects on others, medication usage, sex, and dysphoric states and attitudes. Factor analysis has shown that the LWAQ also covers 2 distinct constructs which are labelled 'health knowledge' and 'health appraisal'.²⁷ The first construct, which contains 49 items, refers to the patient's knowledge of functional limitation imposed by illness. The second construct, which covers 19 items, refers to the patient's evaluation of how much distress those limitations cause. The LWAQ does not ask patients to consider a well-defined time period. To compensate for the tendency of patients to agree with a statement irrespective of its content, the LWAQ contains both positive and negative statements. Patients are asked to indicate for each statement whether it is untrue, slightly true, very true, or not applicable. For negative statements, not applicable and untrue are scored 0, slightly true is scored 1, and true is scored 2. For positive items, not applicable and very true are scored 0, slightly true is scored 1, and untrue is scored 2. Results can be expressed in an overall score, scores for the 11 domain subscales and scores for the 2 construct subscales. Scores are based on a summation of the patient's responses to all statements divided by the number of valid statements (thus compensating for statements that are 'not applicable' or missing).²⁰ Thus, the overall score and the subscores can range from 0 to 2. The lower the score the better the quality of life. The LWAQ has been shown to be reliable and valid,^{9,10} and there is some indication that it may be responsive to change.²⁷

Sickness Impact Profile

In contrast to the two disease-specific questionnaires described above, the SIP is a comprehensive generic quality of life measure. It was developed by Bergner and colleagues,²⁸ and has been shown to be reliable and valid. A validated Dutch version of the SIP is available.²⁹ The SIP measures health-related dysfunction in 12 domains: ambulation, mobility, body care and movement, social interaction, emotional behaviour, alertness behaviour, communication, work, eating, sleep and rest, household management, and recreation and pastimes. It asks patients to consider their situation on the day they complete the questionnaire. The SIP provides summary scores for physical, psychosocial and overall behavioral dysfunction as well as separate scores for the 12 domains. Scores, which are calculated by addition of predetermined weights, range from 0 to 100 and express the percentage of impairment. The larger the score the greater the impairment. The SIP has been used in many studies, including studies of patients with asthma.^{6,25,26,30,31}

Patient Utilities

The primary objective of the instruments outlined above is to *describe* the (change in) quality of life. Utility measurement goes one step further by explicitly *valuing* the quality of life. A utility is a single number on a scale ranging from perfect health, which has a utility of 1, to death, which has a utility of 0. This number represents the value or preference weight assigned to a particular health state. We have elicited utilities from patients by means of the MUMQ, a slightly adapted Dutch version of the McMaster Utility Measurement Questionnaire.^{32,33} This instrument was translated following exactly the same procedure as followed for the AQLQ. The backtranslation of the McMaster Utility Measurement Questionnaire has been published.³⁴ Patients were first asked to describe their own health state of the past two weeks by indicating the level of dysfunction on 6 domains, i.e. physical state and mobility, self-care, emotions, leisure activities, pain or other complaints, and side-effects of treatment. The actual valuation of their own health state is performed by means of rating scale and standard gamble techniques.³⁵ The rating scale looks like a thermometer with 'perfect health' equal to 100 at the top and 'death' equal to 0 at the bottom. Patients are asked to place their own health state on that scale in such a way that it indicates how much better or worse their health state is relative to death and perfect health. In the standard gamble, patients are offered a choice between the certainty of living in their current health state for the rest of their life or taking a gamble in which they have chance p of returning to perfect health for the rest of their life and chance $1-p$ of dying immediately. Probability p is varied until patients are indifferent between the gamble and living in their current health state. At this indifference point the utility equals p . For example, if a patient is indifferent between continuing to live in his current health state and a gamble offering 70% chance of gaining perfect health and

30% chance of dying, the utility or preference weight the patient assigns to his own health state on the 0 to 1 scale is 0.7.

6.3.5 Analysis

Analyses of differences in baseline characteristics between the two treatment groups were based on t-tests for continuous data, and Mann-Whitney U-tests for ordinal data. Differences in proportions were analyzed using χ^2 tests. A p -value ≤ 0.05 was considered statistically significant. Diary card data were summarized per patient by calculating means and standard deviations for the two week run-in period and the 6 week treatment period.

Responsiveness can be operationalized in three different ways. First, the ability of the questionnaires to detect within-subject changes was compared by means of paired t-tests. Second, we compared the ability of the questionnaires to detect differences in change between the group receiving salmeterol and the group receiving salbutamol. This was done by subjecting the change scores to analysis of covariance (ANCOVA) with the treatment group used as a factor and the baseline scores used as covariates. ANCOVA was also used to assess differences in measures of lung function. Further analyses were performed to investigate whether the effect of treatment was different between patients using corticosteroids and patients not using corticosteroids. Third, we calculated the effect sizes in the salmeterol and salbutamol groups. Effect size was calculated as the mean change from baseline to follow-up in the salmeterol group divided by the pooled within-subject standard deviation of change from both the salmeterol and salbutamol groups.^{22,36}

Cross-sectional construct validity was measured by correlating quality of life scores at each visit with lung function parameters, symptom scores and other quality of life scores. Longitudinal construct validity was assessed by calculating correlations between within-subject changes in quality of life scores and within-subject changes in lung function parameters, symptom scores and other quality of life scores. The strength of the correlation was reported using Spearman Rank Correlation Coefficients. The correlation between change in quality of life and patient's overall assessment of efficacy expressed on a five-point scale was also seen as an indicator of longitudinal construct validity. The scores from the overall assessment were reversed so that a higher score means a better efficacy. The patient's overall assessment of efficacy was seen as a reference measure of whether or not a patient had changed. If the change in quality of life scores reflects this overall assessment of change (i.e. a strong correlation), this is seen as evidence of longitudinal validity.

A quality of life measure is reproducible if it yields the same results when applied repeatedly in stable subjects under the same conditions.³⁷ Assessment of reproducibility is best done by repeated administration of an instrument to subjects

whose status has not changed.³⁷ In our study, assessment of reproducibility was based on data from patients who had indicated no change on the 5 point overall assessment efficacy scale, i.e. who were judged to be stable. Reproducibility was expressed in an intra-class correlation coefficient, which is the ratio of between-subject variance to the total variance including both between- and within-subject variance.³⁶ The higher this ratio, the higher the reproducibility.

6.4 Results

6.4.1 Patients

After randomization, one of the 120 patients had to be excluded because his understanding of the Dutch language was too poor to complete the questionnaires. Eight of the remaining 119 patients did not complete the 6 weeks' follow-up measurement. Three of these patients (1 salmeterol and 2 salbutamol) dropped out because of pulmonary problems. The pulmonary problems of the 2 patients receiving salbutamol were so serious that they required hospitalization. Two patients (1 salmeterol and 1 salbutamol) dropped out because of other illnesses and two patients (1 salmeterol and 1 salbutamol) dropped out due to family circumstances (e.g. death of a close relative). One patient receiving salmeterol was no longer willing to participate. Another 4 patients (2 salmeterol and 2 salbutamol) were not included in the analysis because of protocol violations (they were given antibiotics to treat respiratory infections). Thus, the analysis were performed on 107 patients, 53 receiving salmeterol and 54 receiving salbutamol.

The baseline characteristics of the patients who were not included in the analysis did not differ from those who completed the study. Table 1 shows no statistically significant differences between the treatments groups at baseline. Also, no significant differences were found on social indicators like marital status, educational level, income and employment, except for the fact that a smaller percentage of the patients in the salmeterol group had children (64% vs 84%; $p=0.024$). No differences between the groups were found on baseline quality of life scores.

Table 1. - Baseline characteristics by treatment group.

	Treatment groups		p-value slm vs slb
	salmeterol (slm) n=53	salbutamol (slb) n=54	
Age: mean (sd)	50.7 (15.0)	55.1 (12.9)	0.103
Gender: no. (%) female	26 (49%)	21 (39%)	0.289
Age of onset asthma: mean (sd)	31.4 (19.7)	32.9 (19.5)	0.695
FEV ₁ % predicted: mean (sd)	59.2 (6.6)	59.3 (7.0)	0.958
PEFR prebronchodilator (l/sec): mean (sd)	4.0 (1.2)	4.2 (1.4)	0.473
PEFR postbronchodilator (l/sec): mean (sd)	5.3 (1.7)	5.1 (1.5)	0.546
Reversibility (%): median (quartiles)	27 (17, 35)	22 (17, 31)	0.154
No. (%) using steroids:			
inhaled	37 (69.8%)	33 (61.1%)	0.077
oral	4 (7.5%)	0	
both	3 (5.7%)	4 (7.4%)	
Asthma Quality of Life Quest.: mean (sd)	5.55 (1.01)	5.62 (0.84)	0.702
Living with Asthma Quest.: mean (sd)	0.78 (0.36)	0.77 (0.31)	0.963
Sickness Impact Profile: mean (sd)	6.99 (6.43)	9.03 (7.56)	0.137
Utilities: mean (sd):			
rating scale	69.91 (13.95)	67.13 (17.44)	0.366
standard gamble	0.87 (0.13)	0.87 (0.16)	0.760

Abbreviations: FEV₁ = Forced Expiratory Volume in 1 second, PEFR = Peak Expiratory Flow Rate.

6.4.2 Clinical outcome measures

Table 2 shows that salmeterol led to significant improvements over salbutamol on all respiratory outcomes, except for PEFR in the evening. No significant difference in these improvements was observed between patients taking corticosteroids and those not taking corticosteroids. Thus, an interaction between corticosteroid use and type of bronchodilator was not present. The mean use of supplemental salbutamol during daytime was significantly lower for salmeterol than for salbutamol (0.64 versus 1.24 blisters per day; $p=0.013$). The additional salbutamol use during the night was not significantly different between the groups (0.13 versus 0.25 blisters per day; $p=0.076$). The mean daytime and nighttime symptom scores were very low and not significantly different between the two treatment groups ($p=0.08$ and $p=0.67$ respectively). The percentage of days without any symptoms was significantly higher for salmeterol than for salbutamol (63% versus 48%; $p=0.039$), but the percentage of nights without awakenings was

not significantly different between the groups (80% versus 70%; $p=0.108$). Asthma-related adverse events and pharmacologically predictable adverse events occurred at the same rate in both treatment groups. After 3 weeks of treatment the patients' overall assessments of efficacy clearly favoured salmeterol (table 3). After six weeks of treatment no further changes in the patients' overall assessments were found.

Table 2. Effectiveness in terms of lung function by treatment group

	Mean change in lung function				
	salmeterol (slm)	salbutamol (slb)	diff (slm-slb)	95% CI of diff.	p-value slm vs slb
AM PEFR (l/min)	+32.87	+8.60	24.26	8.94 - 39.59	0.002
PM PEFR (l/min)	+16.76	+4.29	12.47	-1.17 - 26.13	0.073
FEV ₁ % pred.	+9.27	+1.73	7.54	3.20 - 11.88	0.001
FVC (l)	+0.32	+0.04	0.27	0.10 - 0.44	0.002
Reversibility (%)	-16.09	-7.32	-8.77	-12.79 - -4.75	<0.001

Abbreviations: FEV₁ = Forced Expiratory Volume in 1 second, PEFR = Peak Expiratory Flow Rate, FVC = Forced Vital Capacity.

Table 3. Patients' overall assessment of efficacy after 3 weeks of treatment

	no. (%) of patients whose asthma was:				
	much improved	improved	equal	worse	much worse
salmeterol (n=53)	9 (17.0%)	22 (41.5%)	21 (39.6%)	1 (1.9%)	0
salbutamol (n=53)	3 (5.7%)	12 (22.6%)	34 (64.2%)	4 (7.5%)	0

Mann-Whitney test: $p = 0.0011$

6.4.3 Responsiveness

Table 4 shows that the AQLQ, the LWAQ and the rating scale utilities were highly responsive to the within-patient improvements in the salmeterol group. These measures also showed some within-patient improvement in the salbutamol group. Only the standard gamble utilities showed a within-patient deterioration of quality

of life in the salbutamol group. This deterioration was not found in any of the other quality of life measures.

Although all the questionnaires revealed the same trend in favour of salmeterol, only the AQLQ and the rating scale utilities showed a statistically significant higher improvement in quality of life on salmeterol than on salbutamol. The overall AQLQ score and the domain scores for activity and symptoms showed a significantly greater improvement on salmeterol. The activity domain of the AQLQ was found to be most responsive, and the emotional domain was the least responsive to change. The overall LWAQ score and its construct subscores did not show a significant difference between the groups. However, the table does not show the finding that 2 of its 11 domains, sport and sex, did show a difference in favour of salmeterol. The mean difference between the groups as regards change in the dimension of sport was 0.27 (95% CI from 0.09 to 0.44; $p=0.003$) and the mean difference for change in the dimension of sex was 0.24 (95% CI from 0.05 to 0.42; $p=0.012$). No significant interaction effect between treatment and the use of corticosteroids was found for any of the quality of life outcomes.

Table 4. Responsiveness to changes within subjects and differences in changes between salmeterol (n=53) and salbutamol (n=54)

	Change in Quality of Life				
	salmeterol	salbutamol	diff (slm-slb)	95% CI of diff.	p-value slm vs slb
<i>AQLQ</i>	+0.49***	+0.27***	0.22	0.03 - 0.42	0.022
Activities	+0.57***	+0.22*	0.35	0.14 - 0.56	0.002
Symptoms	+0.58***	+0.33**	0.25	-0.001 - 0.49	0.051
Environment	+0.44***	+0.22*	0.22	-0.04 - 0.48	0.100
Emotions	+0.18*	+0.17*	0.01	-0.19 - 0.20	0.960
<i>LWAQ</i>	-0.12***	-0.09*	0.03	-0.04 - 0.09	0.425
Health knowledge	-0.13***	-0.08**	0.05	-0.02 - 0.12	0.146
Health appraisal	-0.06**	-0.10	0.04	-0.12 - 0.03	0.290
<i>SIP</i>	-1.59	-0.93	0.66	-1.11 - 2.43	0.463
Physical	-0.83	-0.20	0.62	-0.73 - 2.00	0.363
Psychosocial	-2.82**	-1.91*	0.91	-1.35 - 3.16	0.429
<i>Utilities</i>					
Rating Scale	+7.35***	+4.08**	3.27	0.08 - 6.46	0.045
Standard Gamble	+0.008	-0.034*	0.042	-0.005 - 0.09	0.077

Covariance analysis: Improvement is indicated by a + for the AQLQ and the Utilities and by a - for the LWAQ and the SIP

Within-patient changes: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Table 5 relates the change in the mean score in the salmeterol group to the within-subject standard deviation of change in both the salmeterol and salbutamol groups. The effect sizes that result from this analysis showed the AQLQ to be most responsive, followed by the LWAQ and rating scale utilities.

Table 5. Effect sizes in salmeterol group by quality of life instrument

Quality of life instrument	within-subject change in salmeterol (Δ S1m)	pooled within-subject sd of change (sd Δ S1m + S1b)	Effect size (Δ S1m / sd Δ S1m + S1b)
<i>AQLQ</i>	0.492	0.600	0.820
Activities	0.573	0.666	0.860
Symptoms	0.581	0.803	0.723
Environment	0.438	0.797	0.550
Emotions	0.176	0.583	0.302
<i>LWAQ</i>	0.118	0.170	0.694
Health knowledge	0.130	0.208	0.625
Health Appraisal	0.064	0.192	0.333
<i>SIP</i>	1.585	4.956	0.320
Physical	0.827	3.950	0.209
Psychosocial	2.815	6.648	0.423
<i>Utilities</i>			
Rating Scale	7.345	10.381	0.708
Standard Gamble	0.008	0.161	0.050

* $p \leq 0.01$

** $p \leq 0.001$

6.4.4 Cross-sectional validity

Cross-sectional construct validity correlations are given in table 6. In general the association between quality of life measures and lung function measures was low. The SIP correlated significantly with all lung function measures except reversibility. The LWAQ and the SIP correlated better with the FVC and the AQLQ correlated better with reversibility than the other quality of life measures.

All quality of life questionnaires, except for the standard gamble utilities, correlated significantly with both daytime and nighttime symptom scores. This correlation was considerably higher for the AQLQ than for the other questionnaires.

The correlations among quality of life questionnaires were much higher than the correlations between quality of life and lung function. The correlation between the 2 disease-specific quality of life measures was strong. The SIP correlates well with the LWAQ and rating scale utilities. Standard gamble utilities showed weak correlations with the other quality of life measures.

Table 6. Cross sectional validity: Spearman rank correlation coefficients

	Quality of Life				
	Asthma Quality of Life	Living with Asthma	Sickness Impact Profile	Rating Scale Utilities	Standard Gamble Utilities
<i>Lung function</i>					
PEFR AM	0.02	-0.13	-0.16*	0.13	0.05
PEFR PM	-0.03#	-0.09	-0.17*	0.11	0.06
FEV ₁ % pred.	0.12	-0.15	-0.17*	0.22**	-0.04#
FVC	0.13	-0.26**	-0.29**	0.10	0.06
Reversibility (%)	-0.28**	0.17*	0.02	-0.09	0.04#
<i>Symptoms</i>					
Daytime score	-0.59**	0.42**	0.39**	-0.26*	-0.17
Nighttime score	-0.70**	0.42**	0.38**	-0.33**	-0.21
<i>Quality of Life</i>					
Asthma Quality of Life		-0.73**	-0.48**	0.47**	0.19*
Living with Asthma			0.56**	-0.43**	-0.13
Sickness Impact Profile				-0.59**	-0.15
Rating Scale Utilities					0.17*

* $p \leq 0.01$ ** $p \leq 0.001$

paradoxical direction

6.4.5 Longitudinal validity

Longitudinal construct validity correlations are given in table 7. Of all quality of life measures, the change in AQLQ correlates best with the change in lung function. The longitudinal correlations between these two measures are stronger than the cross-sectional correlations. The other quality of life measures show hardly any correlation with lung function.

The longitudinal correlation with both daytime and nighttime symptom score is best for the AQLQ, followed by the rating scale utilities. The change in LWAQ correlates with change in daytime symptom score only.

The relatively high correlation between the AQLQ and the patients' overall assessments of efficacy indicates that the patients' perceptions of change were indeed reflected in changes in the quality of life as measured with the AQLQ. The other quality of life measures did not reflect the patients' global assessments of efficacy.

The change in AQLQ shows significant correlation with change in all other quality of life measures. This longitudinal correlation was strongest between the AQLQ and the rating scale utilities. The change in LWAQ correlates with the

change AQLQ and the change in SIP, but not with the change in utilities. The correlation between the two utility measures is non-significant.

Table 7. Longitudinal construct validity: Spearman rank correlation coefficients between change values

Change in	Change in Quality of Life				
	Asthma Quality of Life	Living with Asthma	Sickness Impact Profile	Rating Scale Utilities	Standard Gamble Utilities
<i>Lung function</i>					
PEFR AM	0.36**	-0.04	-0.19	0.19	0.23*
PEFR PM	0.35**	-0.20	-0.28*	0.08	0.16
FEV ₁ % pred.	0.32**	-0.05	-0.11	0.19	0.09
FVC	0.28*	-0.05	-0.12	0.21	0.24*
Reversibility (%)	-0.21	-0.02#	0.01	-0.12	-0.03
<i>Symptoms</i>					
Daytime score	0.51**	-0.24*	-0.20	0.24.	0.08
Nighttime score	0.54**	-0.06	-0.04	0.23*	0.14
<i>Overall assessment of efficacy ~</i>	0.64**	-0.20	-0.29	0.24	0.20
<i>Quality of Life</i>					
Asthma Quality of Life		-0.29*	-0.31**	0.36**	0.29**
Living with Asthma			0.31**	-0.11	-0.09
Sickness Impact Profile				-0.17	-0.18
RS Utilities					0.16

* $p \leq 0.01$

** $p \leq 0.001$

paradoxical direction

~ the patient's overall assessment of efficacy is not a change score

Table 8 looks further into the longitudinal construct validity by correlating the domains of the AQLQ with the domains or constructs of the other questionnaires. Relatively strong correlations are found between the emotional domain and the total and construct scores of the LWAQ. Furthermore, the activities and symptoms domain correlate significantly with the health knowledge construct of the LWAQ and the overall LWAQ score. The association between the domains of the AQLQ and the health appraisal construct of the LWAQ is weak, except for the emotional domain. All domains of the AQLQ show significant correlation with the rating scale utilities. The overall SIP score and the psychosocial domain of the SIP correlate better with the domains of the AQLQ than the physical domain.

Table 8. Longitudinal construct validity of the quality of life domains: Spearman rank correlation coefficients between *change* values

Change in	Change in Asthma Quality of Life Questionnaire			
	Activities	Symptoms	Environment	Emotions
<i>Living with Asthma</i>	-0.28*	-0.25*	-0.18	-0.38**
Health knowledge	-0.31**	-0.24*	-0.21	-0.37**
Health appraisal	-0.03	-0.03	0.06#	-0.30**
<i>Sickness Impact Profile</i>	-0.31**	-0.22	-0.26*	-0.22
Physical	-0.12	-0.05	-0.05	-0.10
Psychosocial	-0.19	-0.15	-0.15	-0.20
<i>Rating Scale Utilities</i>	0.25*	0.28*	0.29*	0.26
<i>Standard Gamble Utilities</i>	0.21	0.23*	0.18	0.26*

* $p \leq 0.01$

** $p \leq 0.001$

paradoxical direction

6.4.6 Reproducibility

Forty patients rated themselves as unchanged on the 5-point overall assessment of efficacy scale, both at 3 and 6 weeks' follow-up. The reproducibility of the two disease-specific questionnaires and the SIP in these stable patients was relatively high. The intra-class correlation coefficient was 0.93 for the AQLQ, 0.85 for the LWAQ and 0.81 for the SIP. The intra-class correlation coefficients for the two utility measures were generally lower: 0.65 for the rating scale and 0.56 for the standard gamble.

6.5 Discussion

Whether a quality of life questionnaire is suitable for use in clinical trials depends largely on how well it detects treatment-related changes in the impact of a disease on quality of life. Of all the quality of life measures we used in this study, only the AQLQ and the rating scale utilities reflected the greater improvement that was found in virtually all clinical outcome measures on treatment with salmeterol as compared to salbutamol. The LWAQ domain scores for sport and sex also showed a significant difference in favour of salmeterol. However, despite the presence of clear clinical differences, the overall LWAQ and its two construct scores failed to detect a difference in quality of life between the treatment groups.

The AQLQ was found to be the most sensitive to improvements in health status. The greater responsiveness of the AQLQ may result from 'informed' administration,³⁸ i.e., that patients read the answers they had given at baseline just before answering the questions at follow-up. This "informed answering" is a part of the administration procedure prescribed for this questionnaire, and contrasts with the procedure for the other questionnaires.

The activity domain of the AQLQ was particularly highly responsive. The responsiveness may be increased because 5 of the 8 items in this domain are 'patient-specific', which means that patients themselves indicate which 5 most important activities should form the items. Thus the activity domain focuses on what really matters to a patient. The most frequently identified individual activities that were impaired because of asthma were housework (e.g. vacuuming, scrubbing the floor) which was mentioned by 37% of the patients, walking uphill or upstairs (33%), sleeping (25%), hurrying (25%), exercising (e.g. running, football, ballet) (22%) and cycling (20%). The disadvantage of using individualized items may be that comparing the activity scores between patients is somewhat difficult since they relate to different activities. Moreover, individualization is not a prerequisite for responsiveness. For, a recent study showed that the non-individualised St. George's Respiratory Questionnaire Impact score was able to detect a difference between cromolyn sodium and placebo, even though the clinical improvements due to this drug seemed to be less than the improvements achieved by salmeterol in our trial.³¹ The emotional and environmental domains of the AQLQ were less responsive than the activities and symptom domains. The environmental domain examines how environmental stimuli such as smoke, weather, perfumes and smells affect the patient's asthma. Probably, this is unlikely to be influenced to any great extent by drugs that primarily aim at bronchodilation. The emotional domain addresses topics such as concern, frustration and fear, which may be highly related to a patient's psychological disposition, and is unlikely to change in 6 weeks.

This explains why the developers of the LWAQ predicted that the health appraisal construct of the LWAQ would be less responsive to change over time caused by drug treatment than the health knowledge construct.³⁹ Since the presence of problems is more closely related to morbidity, the patient's knowledge of these problems is expected to be more sensitive to changes in morbidity than the patient's emotional appraisal of these problems, which is more closely related to personality. This hypothesis is confirmed in our study. It also suggests that for a quality of life instrument to be highly responsive it should focus on functional limitation more than on emotional dysfunction.

Although the health knowledge construct is more responsive than the health appraisal construct, it is not as responsive as the activity domain of the AQLQ, even though it covers a greater range of activity restriction. Along with the patient-specific questions of the AQLQ, this may be due to the process through which the items were reduced. The item reduction process in constructing the LWAQ was based on factor analysis, whereas the item reduction process of the AQLQ was

based on patient responses. Thus, the latter includes only those activities which were frequently reported and judged important.⁴⁰

Of the two utility measurement techniques used in our study, only the rating scale measured a significantly greater improvement on salmeterol than on salbutamol. It was more responsive to change than the standard gamble. Utilities elicited via the standard gamble technique were the only outcomes that showed a deterioration in the salbutamol group. All other measures showed a trend indicating an improvement of quality of life in both groups, but a greater improvement in the salmeterol than in the salbutamol group. This deterioration was not a result of outlier values. It was mainly due to the fact that salbutamol patients who stated in the overall assessment that they had not changed showed a significantly larger reduction in mean utility than salmeterol patients who stated they had not changed. We do not have a good explanation for this.

That the observed trend was not significant in a number of quality of life measures does not necessarily imply that there are no real differences in the domains of these questionnaires. Since salmeterol is clinically effective, it could have led to an increased level of activity as a result of which the burden of symptoms is also increased.⁴¹ Furthermore, the extra effects of salmeterol over salbutamol may be camouflaged by the higher use of rescue salbutamol (especially during the day) in patients given salbutamol.

This difference in the use of rescue salbutamol may also explain why there seems to be no improvement in night-time asthma due to salmeterol. For, the percentage of nights without awakenings did not show a significant difference between the two groups. However, during both the day and the night, patients receiving salbutamol used more supplemental salbutamol than patients receiving salmeterol, although only the daytime use was significantly different between the groups. Furthermore, in both treatment groups the symptom scores were very low at baseline, thus leaving little room for improvement.

It is known that generic quality of life measures such as the SIP are usually less responsive to small changes than disease-specific measures. Although the SIP showed the overall trend for greater improvement in salmeterol than in salbutamol, it failed to identify a significant difference between patients receiving salmeterol and patients receiving salbutamol. The latter is similar to the result of a recent study in mild to moderate asthma patients, in which cromolyn sodium led to significant improvements in asthma severity, nighttime asthma, daytime bronchodilator use and the St. George's Respiratory Questionnaire Impacts score, but not in the SIP score.³¹ Nevertheless, the use of generic quality of life measures has the advantage that it allows comparisons across diseases and patient populations, which is of interest especially when the results of a study are presented to policy makers. Inclusion of generic quality of life measures in addition to disease-specific measures is useful in providing a broader range of domains in which changes may be detected. A new drug such as salmeterol may have unanticipated effects on quality of life that are not detected by disease-specific

questionnaires. However the SIP did not detect any such effects. Our patients' SIP scores were skewed towards the less impaired end of the scale, thus leaving little room for improvement. The same is true for the standard gamble which mean score was already 0.87 at baseline. Patients with the best possible score can still have substantial quality of life impairment. But improvements in these patients simply cannot be detected by the instrument. Such a ceiling effect⁴² decreases the responsiveness of an instrument.

Assessing the statistical significance of changes in quality of life measures is relatively easy, but a meaningful interpretation of the magnitude of the significant changes that have been found is difficult. This is because we have not yet had the opportunity to gain wide experience with these instruments and are still unfamiliar with their units of change. By relating changes in AQLQ scores to patients' overall assessments of change, Juniper *et al.* demonstrated that a within-subject change of 0.5 for both the overall and the AQLQ domain scores can be seen as the minimal important difference.⁴³ This is the smallest difference which patients perceive as beneficial. Thus, in our study, the change in overall AQLQ score and in the activities and symptoms domain scores within the salmeterol group can be judged clinically relevant. No such studies of the minimal important differences have been carried out for the other instruments.

Not only patients given salmeterol, but also patients given salbutamol show some improvement in respiratory function and quality of life after 6 weeks of treatment. These quality of life improvements were large enough to be detected by the AQLQ, the LWAQ, and the rating scale utilities. Thus, salbutamol appears to have a positive effect on quality of life also. To some extent, this effect may result from patients receiving increased attention in a clinical trial situation compared to routine clinical practice. Furthermore, the knowledge that the effects of treatment will be carefully measured may have led to an increased compliance of patients with the drug therapy.

In the literature, evidence of responsiveness is seen both as a distinct property of a quality of life measure and as an indication of a measure's validity.⁴⁴ In the absence of a gold standard for quality of life, validity can be assessed by examining the correlations between the questionnaires and measures of respiratory function or symptoms, as these are assumed to influence physical functioning and perhaps other aspects of quality of life. This kind of validity is referred to as construct validity. With respect to the longitudinal correlation with lung function and symptom scores, the AQLQ again performed best. The AQLQ showed relatively high correlation with both daytime and nighttime symptom scores. The changes in LWAQ, SIP and utilities did hardly show any correlation with changes in lung function parameters. These 3 instruments also failed to correlate with the patients' overall assessments of efficacy. A number of correlations had paradoxical directions. However, these correlations were all non-significant.

Construct validity can also be assessed by comparing the correlations among various quality of life measures. The correlations among quality of life measures

were much higher than the correlations between quality of life measures and respiratory function. For example the cross-sectional correlation between the LWAQ and the AQLQ was 0.73. This might be expected given the large degree of overlap in the content of the items which comprise the scales. However longitudinally the correlation between the LWAQ and the AQLQ was only 0.29. The change in AQLQ correlates better with the change in rating scale utilities than with the change in LWAQ.

It was not expected that the longitudinal correlation with the domains of the AQLQ would be weaker for the physical domain of the SIP than for the psychosocial domain of the SIP. This may be because the physical domain of the SIP tends to capture only very severe physical dysfunction, whereas the psychosocial domain also captures less severe psychosocial problems. Thus, there is more room for change in the latter domain. Furthermore, we did not expect the longitudinal association between the emotional domain of the AQLQ and the health knowledge construct of the LWAQ to be higher than the association between the emotional domain and the health appraisal construct. This may be because the emotional domain of the AQLQ assesses only concern, frustration and fear, whereas the health appraisal construct of the LWAQ is more broadly focused. But, it may also point to the enormous complexity of emotions and psychological assessments of functional limitations.

Construct validity evaluation showed that rating scale utilities correlated better with symptom scores than standard gamble utilities. It also showed that rating scale utilities correlated better with other quality of life measures than standard gamble utilities. The correlation between the two types of utilities was weak. Either the standard gamble measures largely different aspects of health -in which case our validity concept is not valid- or the standard gamble has indeed lower construct validity. It may be that it chiefly measures patients' risk behaviour, since patients are asked to value their health states under risk conditions, whereas with the rating scale health states are valued under certainty.

In summary, the results of this study indicate that salmeterol tends to have a more favourable effect on quality of life than salbutamol. The AQLQ and the rating scale utilities were found to be sufficiently responsive to detect the clinically important changes due to the use of salmeterol, whereas the LWAQ, the SIP and the standard gamble utilities were not. Both disease-specific instruments appear to be reliable, but the AQLQ is more valid than the LWAQ. Of the generic instruments, the SIP showed better reliability than the utilities instruments. The validity of the standard gamble was worst.

6.6 Acknowledgements

We would like to thank Michael Hyland and Elisabeth Juniper for discussing the results with us. The study has been financially supported by GLAXO B.V. The Netherlands.

6.7 References

1. Spilker B, Molinek F, Johnston K, *et al.*. Quality of Life Bibliography and Indexes. *Medical Care* 1989, 27 (Supplement), S77
2. Smith ND, Quality of life studies from the perspective of an FDA reviewing statistician. *Drug Information Journal* 1993, 27, 617-623
3. McSweeney AJ, Heaton, RK, Grant I, Cugell D, Solliday N, Timms R. Chronic obstructive pulmonary disease; socioemotional adjustment and life quality. *Chest* 1980, 77, 309-11
4. Guyatt GH, Berman LB, Townsend M, Pugsley S, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987, 42, 773-8
5. Williams SJ, Bury MR. Impairment, disability and handicap in chronic respiratory illness. *Social Science and Medicine* 1989, 29, 609-16
6. Schrier AC, Dekker F, Kaptein AA, Dijkman JH. Quality of life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990, 89, 894-9
7. Van Schayck CP, Rutten-van Mólken MPMH, Van Doorslaer EKA, Folgering H, Van Weel C. Two-year bronchodilator treatment in patients with mild airflow obstruction. Contradictory effects on lung function and quality of life. *Chest*, 1992, 102, 1384-1391
8. Malo J-L, Boulet L-P, Dewitte J-D, Cartier A, L'Archeveque J, Coté J, Bedard G, Boucher S, Champagne F, Tessier G, Contandriopoulos A-P, Juniper EF, Guyatt GH. Quality of life of subjects with occupational asthma. *Journal of Allergy and Clinical Immunology* 1993, 91, 1121-1127
9. Hyland ME. The Living with Asthma Questionnaire. *Respiratory Medicine* 1991, 85 (Supplement), 13-16
10. Hyland ME, Finnis S, Irvine SH. A scale for assessing quality of life in adult asthma sufferers. *Journal of Psychosomatic Research* 1991, 35, 99-110
11. Marks GB, Dunn SM, Woolcock AJ. A scale for the measurement of quality of life in adults with asthma. *Journal of Clinical Epidemiology* 1992, 45, 461-472
12. Creer TL, Wigal JK, Kotses H, McConnaughy K, Winder JA. A life Activities Questionnaire for Adult Asthma. *Journal of Asthma* 1992, 29, 393-399
13. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire *American Journal of Respiratory Disease* 1992, 145, 1321-1327
14. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992, 47, 76-83

15. Ullman A, Hedner J, Svedmyr N. Inhaled Salmeterol and Salbutamol in Asthmatic Patients. An evaluation of asthma symptoms and the possible development of tachyphylaxis. *American Review of Respiratory Disease* 1990, 142, 571-575
16. Britton MG, Earnshaw JS, Palmer JBD, on behalf of a European Study Group. A twelve months comparison of salmeterol with salbutamol in asthmatic patients. *European Respiratory Journal* 1992, 5, 1062-1067
17. Lundback B, Rawlinson DW, Palmer JBD, on behalf of a European study group. Twelve month comparison of salmeterol and salbutamol as dry powder formulations in asthmatic patients. *Thorax* 1993, 48, 148-153
18. Pearlman DS, Chervinsky P, LaForce G, Seltzer JM, Southern DL, Kemp JP, Dockhorn RJ, Grossman J, Liddle RF, Yancey SW, Cocchetto DM, Alexander WJ, Van As A. A comparison of salmeterol with albuterol in the treatment of mild-to-moderate asthma. *The New England Journal of Medicine* 1992, 327, 1420-1425
19. Fitzpatrick MF, Mackay T, Driver H, Douglas NJ. Salmeterol in nocturnal asthma: a double blind, placebo controlled trial of a long acting inhaled β_2 agonist. *British Medical Journal* 1990, 301, 1365-1368
20. Palmer JBD, Hyland ME. Salmeterol in clinical practice: comparator and safety studies, quality of life studies. *European Respiratory Journal* 1991, 4, 301-303
21. Juniper EF, Johnston P, Borkhoff C, Haukioja A. Effect of salmeterol on asthma quality of life. (Abstract) *American Review of Respiratory Disease* 1993, A60
22. Guyatt GH, Kirshner B, Jaeschke R. Measuring Health Status: What are the necessary measurement properties? *Journal of Clinical Epidemiology* 1992, 45, 1341-1345
23. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993, 46, 417-432
24. European Group for Health Measurement and Quality of Life Assessment: Hunt SM, Alonso J, Bucquet D, Niero M, Wiklung I, McKenna S. Cross-cultural adaption of health measures. *Health Policy* 1991, 19, 33-44.
25. Juniper EF, Guyatt GH, Ferrie PJ, Griffith E. Measuring Quality of Life in Asthma. *American Review of Respiratory Disease* 1993, 147, 832-838
26. Rowe BH, Oxman AD. Performance of an Asthma Quality of Life Questionnaire in an Outpatient Setting. *American Review of Respiratory Disease* 1993, 148, 675-681.
27. Hyland ME, Thompson P, Bellesis M, Kenyon CAP. Quality of life consequences of morbidity reduction in asthma: Health-knowledge versus Health-appraisal. (Submitted)
28. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. *Medical Care* 1981, 19, 787-805
29. Jacobs HM, Luttik A, Touw-Otten FWMM, De Melker RA. De 'Sickness Impract Profile'. Resultaten van een valideringsonderzoek van de Nederlandse versie. *Nederlands Tijdschrift voor Geneeskunde* 1990, 134, 1950-1954
30. Marks GB, Dunn SM, Woolcock AJ. An evaluation of an asthma quality of life questionnaire as a measure of change in adults with asthma. *Journal of Clinical Epidemiology* 1993, 40, 1103-1111
31. Jones PW and the Nedocromil Sodium Quality of Life Study Group. Quality of life, symptoms and pulmonary function in asthma: long-term treatment with nedocromil sodium examined in a controlled multicentre trial. *European Respiratory Journal* 1994, 7, 55-62

32. Bennett K., Torrance GW, Tugwell P. Methodologic challenges in the development of utility measures of health-related quality of life in rheumatoid arthritis. *Controlled Clinical Trials* 1991, 12 (supplement) 118-128
33. Bakker CH, Rutten-van Mólken M, Van Doorslaer E, Bennett K, Van der Linden Sj. Health Related Utility Measurement in Rheumatology: An Introduction. *Patient Education and Counseling* 1993, 20, 145-152
34. Bakker C, Rutten M, Van Doorslaer E, Bennett K, Van der Linden Sj. Feasibility of utility assessment by rating scale and standard gamble in patients with ankylosing spondylitis or fibromyalgia. *Journal of Rheumatology* 1994, 21, 269-274
35. Torrance GW. Measurement of health-state utilities for economic appraisal: A review. *Journal of Health Economics* 1986, 5, 1-30
36. Deyo RA, Diehr P, Patrick D. Reproducibility and Responsiveness of Health Status Measures. Statistics and Strategies for Evaluation. *Controlled Clinical Trials* 1991, 12, 142-158 (Supplement)
37. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987, 40, 171-178
38. Guyatt GH, Berman LB, Townsend M, Taylor DW. Should study subjects see their previous responses? *Journal of Chronic Diseases* 1985, 38, 1003-1007
39. Hyland ME. The validity of health assessments: resolving some recent differences. *Journal of Clinical Epidemiology* 1993, 46, 1019-1023.
40. Hyland ME. Quality-of-Life Assessment in Respiratory Disease. *PharmacoEconomics* 1992, 2, 43-53
41. Hughes MD, Van Essen-Zandvliet EEM, Kerstjens H. Issues in the Design of Long-term Clinical Trials in Asthma. *Drug Information Journal* 1993, 27, 793-803
42. Guyatt GH, Feeny DH, Patrick DL. Measuring Health Related Quality of Life. *Annals of Internal Medicine* 1993, 118, 622-629
43. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *Journal of Clinical Epidemiology* 1994, 47, 81-87
44. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Quality of Life Research* 1992, 1, 73-75

7

Methodological issues of patient utility measurement: Experience from two clinical trials

*Maureen P.M.H. Rutten-van Mölken¹, Carla H. Bakker²,
Eddy K.A. van Doorslaer³, Sjef van der Linden².*

1 Department of Health Economics, University of Limburg.

2 Department of Internal Medicine / Division of Rheumatology, University Hospital Maastricht.

3 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

7.1 Abstract

This paper explores various methodological issues of patient utility measurement in two randomized controlled clinical trials involving 85 fibromyalgia and 144 ankylosing spondylitis patients. In both trials one baseline and two follow-up measurements of patients' preferences for their own health state and several hypothetical states were performed using the rating scale and the standard gamble method.

It was confirmed that standard gamble scores are consistently higher than rating scale scores for both the experienced and hypothetical states. The three-month test-retest reliability for hypothetical states measured by intra-class correlation coefficients ranged from 0.24 to 0.33 for the rating scale and from 0.43 to 0.70 for the standard gamble. Although the reproducibility is not high, the group mean scores are fairly stable over time. Mean standard gamble scores tend to differ depending on the way the measurements are undertaken. Utilities elicited with chained gambles were significantly higher than utilities elicited with basic reference gambles. At the individual level some inconsistent responses occurred. However, more than 70% of these fell within the bounds of the measurement error which ranged from 0.11 to 0.13 on the standard gamble (0-1 scale) and from 8 to 10 on the rating scale (0-100 scale). The large number of negative utilities for the severe hypothetical state, which was used as an anchor point in the chained gambles, and the magnitude of these negative utilities (down to -19) calls for intensified research efforts in order to handle these responses in utility calculations.

7.2 Introduction

Decisions regarding the allocation of resources to health care interventions should ideally be based on the relative costs and benefits of the alternatives. This requires an assessment of the societal value of the outcomes (i.e., various health states) achieved by these interventions, which can be done by means of utility measurement. When utility measurement is applied in clinical trials, patients are asked to assign a single value to a health state on a scale ranging from 0 (usually death) to 1 (usually perfect health), by balancing the positive treatment effects against the negative side effects.¹ A utility can be seen as an inclusive, generic quality of life measure, which reflects the net effect of treatment. It is designed to allow a broad comparison of the effects of health care interventions across patient populations. When such comparisons are made in the context of cost-utility analyses, utilities are often used as weights to compute "quality adjusted life years" (QALYs). Years of life are multiplied by utility weights for the health status during those years, thus adjusting these life years for their quality.

Because the impact of a health care intervention on the value of an individual patient's health state is increasingly recognized as an essential component of the evaluation of that intervention's usefulness, utility measurement has become more widely used during the past decade. In general, two different approaches to utility measurement have been used in clinical trials.² In the *first approach*, patients are asked a number of questions about their functioning. Their answers are used to rate them on the various quality of life dimensions of a particular utility measurement instrument. Combining these dimensions results in descriptions of patients' overall health states, to which preference values are assigned. These values are obtained in a different population, usually the general population. The majority of utility analyses following this approach have used the health state preference values that were obtained by the original developers of the utility instruments.³ The most commonly used pre-packaged utility measurement instruments for this approach are Rosser's Disability Distress scale⁴, Kaplan's Quality of Wellbeing scale⁵, and Torrance's Health Utility Index.⁶

In contrast to the use of pre-packaged systems, the *second approach* to utility measurement is to ask patients to assign a single preference value to their overall quality of life. This self valuation by patients has not been widely undertaken in the past, but appears to be increasingly used.³ The Auranofin trial in rheumatoid arthritis patients is a well known example of this approach.^{7,8} It has yet to be established whether one of these two approaches or any of the available utility instruments is superior.

In the two studies in fibromyalgia and ankylosing spondylitis, described in this paper, we have opted for the second approach. It seems very appealing to us to incorporate patients' preferences in evaluating a therapy, since only they know the true implications of a particular health state from first hand experience, and their preferences reflect the relative desirability of different health states to those who should benefit from the services provided.⁹ Moreover, the first approach commonly requires decision rules to classify patients into the dimensions of a utility instrument using the answers they gave on other quality of life questions. Often these decision rules are not easily explained and justified.

When using pre-packaged systems, the underlying utilities may have been elicited by a number of techniques of which the rating scale, standard gamble and time-trade-off are the major ones. These same techniques can also be used in the approach we opted for, where utilities are derived directly from patients. A comprehensive description of these techniques is given by Torrance.¹⁰ In our study only the rating scale and the standard gamble were used. The first because it is a very simple technique in which a subject provides his preference values explicitly by placing health states on a scale with clearly defined endpoints (e.g. best imaginable health state and worst imaginable health state). The second because it is the only technique that is well-founded on an economic theory, the Von Neumann and Morgenstern expected utility theory.¹¹ We did not consider using the time trade off technique because on the one hand it seems more difficult

than the rating scale and on the other hand it is not founded on a particular theory. Furthermore, for practical reasons the number of techniques had to be limited. A fuller description of the rating scale and standard gamble techniques is given in the methods section.

A review by Froberg and Kane of measurement issues related to obtaining utilities, indicated there was a considerable lack of knowledge about the accuracy of utility measurement techniques.^{12,13d} This paper attempts to make a contribution to overcoming parts of this deficiency by focusing on those methodological issues of patient utility measurement that became apparent in our studies. In the methods section we will provide a brief description of the patients and the studies in which the utility measurements were incorporated, and the methodology of utility measurement that was used. The main body of this paper contains four separate sections on methodological issues. In the first section, a comparison is made between rating scale and standard gamble utilities whereas in the second section the reliability of both techniques is compared. The third section concerns the internal consistency of the standard gamble and the fourth section addresses health states valued worse than death. Some issues, such as the observed differences between rating scale and standard gamble methods¹³⁻¹⁶, inconsistent responses and the differences between basic reference gambles and chained gambles^{17,18} have been addressed before, but usually not in the context of clinical trials. Reliability data are scarcely reported in the existing literature and little has been reported empirically on the occurrence of negative utilities.

7.3 Methods

7.3.1 Patients and studies

Utility measurements were performed in the context of two randomized controlled clinical trials in rheumatic patients. These trials and their results are described in more detail elsewhere.^{19,20} Patients were recruited for both studies from outpatient clinics. In the first study, 85 females suffering from *fibromyalgia* (criteria of Wolfe *et al.*)²¹ were randomly divided into a standardized fitness training programme (n=35), a biofeedback training programme (n=31) and a control group (n=19). Their mean±sd age was 44±8 years, 82% percent was married, 27% employed and 66% had a low educational level. At the 6 months follow-up 12 patients dropped out, 6 in the fitness group, 5 in the biofeedback group and 1 in the control group, all of them for reasons unrelated to their disease.²⁰ The baseline characteristics of the drop-outs did not differ significantly from those of the patients who completed the study.

In the second study 144 patients (21% female) suffering from ankylosing spondylitis (modified New York criteria, 1984)²² were randomly assigned to receive

either 1) self-administered unsupervised individual physical exercise at home ($n=68$) or 2) weekly sessions of group physical therapy in addition to the same individual physical exercise at home ($n=76$). Their mean \pm sd age was 43 ± 10 years, and 67% was married, 72% employed and 35% had a low educational level. By the 9 months follow-up 9 patients had withdrawn (8 in the experimental and 1 in the control group), 4 because of the inability to exercise individually, 4 due to other diseases or pregnancy and 1 who had moved. Their baseline characteristics did not differ significantly from those who did not withdraw.¹⁹

Since the focus of this paper is on methodological issues, it is not our intention to address the effectiveness of the therapies (details on effectiveness can be found in Hidding *et al.*¹⁹, and Van Santen-Hoeufft *et al.*²⁰). The trial data are only used to illustrate some methodological issues.

7.3.2 Methods of utility measurement

In the fibromyalgia study patients were seen for utility measurement at an outpatient clinic at baseline and at the 3 and 6 months follow-ups whereas in ankylosing spondylitis utility measurements were scheduled at baseline, 3 and 9 months follow-ups. The measurements were done by means of the Maastricht Utility Measurement Questionnaire, a translated and slightly adapted version of the McMaster Health Utility Index.^{23,24} This instrument is administered by a trained interviewer and takes about 30 minutes to complete.

All patients were asked to provide utilities for three hypothetical reference health states as well as for their own health state. The description of each health state covers 6 dimensions: 1) activities of daily living, 2) self-care functions, 3) anxiety and depression, 4) leisure activities, 5) pain and discomfort and 6) side effects from treatment. Each patient described his or her own health state by ticking off one of the 5 functional levels within each dimension (1 being the best functional level and 5 being the worst functional level). For all health states, duration was specified as "the rest of your life". The reference states describe typical mild, moderate and severe states of patient's illness. They help patients to determine the position of their own health state on the spectrum of possibilities.²⁵ Since the reference states remain the same throughout the study, they also enable the calculation of the test-retest reliability of the utility measurement instrument at the repeated follow-ups.²⁵

As part of the Maastricht Utility Measurement Questionnaire, utilities were elicited using both the rating scale and the standard gamble method. In the first part of the baseline and follow-up interviews patients rank the reference health states and their own health state by preference on a rating scale with the endpoints of perfect health (100) and the severe reference state (0). They are asked to do this in such a way that the distances between the states represent the differences in their

preferences. The baseline interview then continues with standard gamble questions 1, 2 and 3 as shown in table 1. At follow-up interviews an additional question (question 4) is asked to assess the change which the patient has experienced relative to baseline (see table 1). Before question 4, each patient is explicitly asked whether his health state has improved, deteriorated or remained stable compared to baseline. If a patient indicates an improvement compared to baseline, question 4a is asked. When a patient indicates a deterioration, question 4b is asked. The first 80 ankylosing spondylitis patients coming to the 9 months follow-up were asked an additional fifth question in which they valued their own current health state using perfect health and death as outcomes of the gamble. At 3 and 9 months follow-up all ankylosing spondylitis patients who found that the severe reference state was worse than death were asked a sixth question to assess the magnitude of the negative utility for the severe reference state. Throughout this paper we will refer to the question numbers in table 1.

Table 1. Standard gamble questions at baseline (1,2 and 3) and at follow-up (1,2,3,4,5,6)[†]

Standard gamble question	Health state being valued	Outcomes of the gamble	
		Best (p)	Worst (1-p)
1	mild reference state	perfect health	severe reference state
2	own health state	perfect health	severe reference state
3	severe reference state	perfect health	death
4a	own health state at follow-up	perfect health	own health state at baseline
4b	own health state at baseline	perfect health	own health state at follow-up
5	own health state	perfect health	death
6	death	perfect health	severe reference state

[†] question 5 was only put to 80 ankylosing spondylitis patients at 9 months follow-up; question 6 only at 3 and 9 months follow-up to ankylosing spondylitis patients who found the severe state worse than death.

The standard gamble is sometimes seen as the gold standard for utility measurement, because it is directly based on the axioms of the Von Neumann and Morgenstern expected utility theory.¹¹ This theory consists of a number of axioms for rational decision-making under risk. One of these axioms specifies the standard gamble approach to utility measurement. In each standard gamble question a patient

is offered a choice between certain continued life in the health state being evaluated (h_i), and a gamble with chance p to gain the best outcome of the gamble (perfect health) and chance $1-p$ of attaining the worst outcome of the gamble. The health state being valued must be intermediate between the two outcomes of the gamble in terms of preference. Chance p is systematically varied until the patient is indifferent between continued life in state h_i and taking the gamble. In our studies p was varied in steps of 10% ($p/1-p$: 100/0, 90/10, 10/90, 80/20, 20/80 etc.) When the indifference point has been found, the utilities of the health states (U_{h_i}) are calculated using the expected utility equation: $U_{h_i} = pU_x + (1-p)U_z$, where p is the indifference probability, U_x is the utility of the best outcome of the gamble and U_z is the utility of the worst outcome of the gamble¹⁾.

Standard gambles with perfect health and death as potential outcomes are called *basic reference gambles*. Since by definition the utility of perfect health is 1 and the utility of death is 0, the utility of the health state being valued in a basic reference gamble is equal to $p \cdot 1 + (1-p) \cdot 0 = p$, the indifference probability. Generally, the more undesirable the health state being valued, the greater the willingness to take a risk in order to escape that health state, the lower the indifference probability p , and thus the lower the utility for that state. Thus, the standard gamble provides an implicit valuation of a health state relative to the two possible outcomes of the gamble.

The worst (or best) outcome of the gamble can be replaced by any other health state as long as it is worse (or better) than the health state being valued.²⁶ Such gambles are called *chained gambles*, because they have to be chained to a basic reference gamble that assesses the utility of that other health state. In our studies, the severe reference state was substituted for death, in order to avoid using death in a gamble that involved a chronically ill patient's own health state. Including death could upset them. Moreover, in the period covered by our studies, death was unlikely to be a relevant outcome in the rheumatic disease patient groups we studied. When the severe reference state is used as the worst outcome of the gamble (as in standard gamble question 1 and 2) the utility for the health state being valued can be calculated using the same equation as above, where U_z is the utility of the severe reference state, a utility which is measured in basic reference gamble 3. In standard gamble question 4a, U_z is the utility of the patient's own health state at baseline, which was measured during the baseline-interview. In standard gamble 4b, U_z is the utility of the patient's own health state at follow-up, which is being measured in the follow-up interview.

In general, when the severe reference state was considered worse than death, a utility of 0 was assigned to that state in order to avoid using negative utilities.

1) The indifference probability is defined as the midpoint of the two probabilities of perfect health between which the preference shifts from the gamble to the sure state. For example, if a patient prefers a gamble with probabilities 90/10 to the sure health state, but prefers the sure health state to a gamble with probabilities 80/20, the indifference probability is 0.85

When indifference is reached in standard gamble question 6, the magnitude of the negative utility is calculated as $-p/(1-p)$, where p is again the indifference probability.¹⁰ Although calculated, these negative utilities were not used in the chained gambles.

To facilitate the patients' understanding of the standard gamble questions a probability wheel was used as a visual aid.¹⁵ This is an adjustable disk with two different colored sectors that reflect the probabilities of getting the two outcomes of the gamble. The outcomes of the gamble are described on cards that have the same color as the sectors. The size of the sectors is changed according to the change in probability.

7.3.3 Statistical analysis

Results will be presented using means and associated standard errors. However, since the negative utilities of the severe reference state elicited in standard gamble question 6 included a number of extremely negative values we present the 5% trimmed mean and the 5% trimmed standard deviation for these results. This means that the upper and lower 5% of all observations were excluded when calculating the mean and standard deviation, thus removing the influence of the outlier values that caused the distribution of negative utilities to be heavily skewed to the left.²⁷ Within-patient analyses by means of paired t-tests were performed 1) to compare rating scale with standard gamble scores, 2) to compare chained gambles with the basic reference gamble, and 3) to test for differences between reference state scores over time. Pearson product moment correlations are reported as measures of association between rating scale and standard gamble scores. A logistic regression analysis was performed to test for differences between patients who did and those who did not give inconsistent responses. Intra-class correlation coefficients were calculated to examine test-retest reliability.²⁸ Reliability was further assessed in terms of the precision of an individual measurement. This precision, expressed as σ_e , is the standard deviation of the measurement error also called the standard error of measurement.²⁹ It is calculated as the square root of the mean square error (MSE) which is given by an analysis of variance.²⁸ It can also be calculated as $\sigma\sqrt{1-R}$, where σ is the standard deviation of all measurements and R is the test-retest reliability coefficient.²⁹

7.4 Results

7.4.1 Comparing rating scale and standard gamble utilities

Although the standard gamble method is sometimes seen as the gold standard, the rating scale method is far more frequently used, probably because it is less time consuming and easier to apply. In our studies Pearson product moment correlations between rating scale and standard gamble scores for various health states were found to range between 0.31 and 0.48 ($p < 0.001$). However, highly significant correlations can coexist with systematic differences between the methods. As can be seen in table 2, the mean utilities of the patients' own health states assessed via the standard gamble method were significantly higher than the utilities assessed via the rating scale method. This pattern, which was also found for the mild reference health state, is consistent with earlier findings.¹³⁻¹⁶

Table 2. Mean (se) rating scale and standard gamble values for the patient's own health state

	N	Rating scale† endpoints: perfect health - severe	Standard gamble outcomes: perfect health - severe	p-value‡
<i>Fibromyalgia</i>				
baseline	85	0.54 (0.020)	0.67 (0.028)	<0.001
3 months	76	0.57 (0.023)	0.76 (0.023)	<0.001
6 months	73	0.60 (0.022)	0.76 (0.025)	<0.001
<i>Ankylosing spondylitis</i>				
baseline	144	0.72 (0.013)	0.75 (0.018)	0.095
3 months	137	0.74 (0.012)	0.78 (0.016)	0.009
9 months	133	0.75 (0.013)	0.79 (0.015)	0.002

† Rating scale preferences were divided by 100.

‡ Paired t-test

In ankylosing spondylitis we found a difference of somewhat less than 5% between the methods, and in fibromyalgia we found a difference of more than 10%. Such differences might considerably affect the results of a cost-utility analysis and alter the conclusions drawn. Whether this happens depends on the sensitivity of the decisions to the observed range of variation. Some cost utility ratios may be very robust to the magnitude of the utility, whereas others may even change as a result of only a very small change in utility. The variability of responses among patients is somewhat greater for the standard gamble method than for the rating scale method. Several phenomena might explain the differences between rating scale and standard gamble preferences. Three of them are discussed below.

Response Spreading

The first is called "response spreading" on the rating scale. This means that patients tend to distribute the health states over the entire scale, even if the true values were bunched at one end.^{14,30} The mean baseline rating scale scores of 26.8, 54.1 and 72.6 assigned by the fibromyalgia patients to the moderate reference health state, their own health state and the mild reference health state respectively, may indicate a tendency to use the whole scale. In ankylosing spondylitis these baseline scores were 36.7, 80.0 and 76.5. Utilities are cardinal measures, reflecting not only the ranking of various health states relative to perfect health and death, but also the magnitude of the difference between these different health states.²¹ If response spreading occurs, then the rating scale gives an indication of ordinal rankings and the intensity of the preferences, but it does not provide interval-scale utilities.

Risk Attitude

A second explanation for the significant difference between mean rating scale and mean standard gamble scores may be the patients' attitudes towards risk itself. Rating scale scores are measured under certainty, and do not capture the respondent's attitude towards risk. In contrast, the standard gamble approach incorporates the respondent's risk attitude, which may be risk averse, risk neutral or risk seeking. If subjects are not risk neutral, differences can be expected between rating scale and standard gamble values. If they are risk averse their indifference probability increases, thus increasing the utility of the health state being valued. Kahneman and Tversky have shown that people generally acted as if they were risk averse when choices were framed in terms of potential gains and as risk seeking when choices were framed in terms of potential losses.³¹ According to their "prospect theory" in which an S-shaped value function is assumed, the displeasure of a loss is generally greater than the pleasure associated with an equivalent gain.³² Although our gambles are mixed gambles with both a positive and negative potential outcome, the patients in our studies might have focused more on the negative outcome of the gamble, i.e., the severe reference state. In that case the shape of the value function might have contributed to risk-averse behavior. As a result of this behavior, the standard gamble utilities would be biased upward compared to rating scale values.¹⁷

Standard gamble utilities may also be biased upward because people tend to over-weight sure outcomes relative to outcomes which are highly probable. This is called the certainty effect but it is also referred to as the Allais paradox.^{17,31} Kahneman and Tversky's "prospect theory" assumes a decision weight function which over-weights small probabilities and under-weights moderate and high probabilities. If, in our studies, moderate and high probabilities were under-weighted, this might have contributed to the relative attractiveness of the sure outcome, even when the probability of gaining perfect health was rather high. Over-weighting a small chance of ending up in the severe reference state might

have reinforced the attractiveness of the sure outcome. Moreover, the fact that patients knew from experience that they had been able to adapt to their illness before may have diminished both the severity of the health state being valued and the relative value of the therapeutic pay-off from treatment gambles.³³

Cognitive and Emotional Factors

Other explanations for the difference between rating scale and standard gamble values are all related to the previous explanations. Such explanations might include differences in cognitive processes such as recalling and taking account of past events, life goals, family circumstances and the selection of reference points against which consequences are evaluated.¹⁴

Cognitive factors play an important role in Loomes and Sugden's alternative to expected utility theory, called "regret theory".^{34,35} According to that theory, the value a person assigns to a health state does not only depend on that health state but also on how that health state compares to the health state the person might have had if he or she had made a different choice. If what is obtained is better than what might have been, feelings of rejoice may increase the utility; if what comes is worse than what might have been, regret may reduce the utility. In standard gambles where patients are explicitly asked to make a choice, feelings of regret and rejoice may be anticipated, whereas such feelings are absent in the choiceless rating scale valuation process. Subjects may shy away from the gamble choice in the standard gamble because of regret aversion (regret may occur if they "lose" the gamble and end up with the worst outcome). By means of experiments, Loomes and colleagues have shown that regret theory is able to explain why observed preference reversals may have occurred.³⁶

Finally, differences may arise in emotional reactions to past and future health states and events, such as more intense emotional reactions to bad outcomes when they are presented in gambles or more intense emotional reactions to probabilities of death if a family member or friend has died recently. The likelihood of such "recall reactions" is lower for the rating scale than for the more confrontational standard gamble method.

7.4.2 Reliability of the rating scale and the standard gamble

Test-retest reliability

The reproducibility of the rating scale and standard gamble methods has been scarcely reported in the literature. O'Connor reported a Pearson product moment correlation of 0.77 and 0.80 for the one-week test-retest reliability of the rating scale and the standard gamble,³⁷ while Torrance reported product moment correlations of 0.49 and 0.53 for the one-year test-retest reliability of these methods.¹⁵ In our studies reproducibility was assessed by calculating 3 month intra-

class correlation coefficients (ICC) for the values assigned to the reference health states. The results are given in table 3.

Table 3. Test-retest reliability: three months intra-class correlation Coefficients

Reference states	Rating Scale		Standard Gamble	
	Fibromyalgia	Ank. Spondylitis	Fibromyalgia	Ank. Spondylitis
mild	0.33	0.26	0.43	0.50
moderate [†]	0.24	0.29	-	-
severe [‡]	-	-	0.70	0.65

[†] the moderate reference health state was not valued by means of the standard gamble

[‡] the severe reference health state was not valued by means of the rating scale

Although the reproducibility of the standard gamble is somewhat higher than that of the rating scale, the ICCs of the scores assigned to the reference states are generally not very high. This may point at difficulties in valuing abstract, hypothetical health states that have never been experienced. There are good reasons why it may be difficult to envision the well-being associated with a hypothetical health state. One is the inevitable gap between imagination and the actual experience of a health state. Individuals may overestimate or underestimate their ability to accommodate or to cope with adversity.¹³

Although the reproducibility is not high, table 4 shows that -in spite of the occurrence of some slight but statistically significant changes in the preferences for the mild reference state in fibromyalgia and the severe reference state in ankylosing spondylitis- the mean scores are fairly stable over time. This stability may point at the usefulness of aggregated scores for group decision making.

Table 4. Mean (se) utilities of the reference health states

	0 months	3 months	6/9 months†	p-value‡ 0-3	p-value‡ 3-6/9
<i>Fibromyalgia</i>					
rating scale					
moderate	27.3 (1.55)	29.1 (1.68)	30.4 (1.54)	0.382	0.388
mild	72.6 (1.54)	79.7 (1.26)	76.8 (1.28)	0.000	0.035
standard gamble					
mild	0.83 (0.02)	0.85 (0.02)	0.85 (0.02)	0.366	0.851
severe	0.42 (0.04)	0.35 (0.04)	0.31 (0.04)	0.083	0.095
<i>Ankylosing Spondylitis</i>					
rating scale					
moderate	36.7 (1.14)	39.0 (1.28)	39.1 (1.37)	0.116	0.909
mild	76.5 (0.83)	77.3 (0.70)	77.5 (0.90)	0.346	0.888
standard gamble					
mild	0.84 (0.02)	0.85 (0.01)	0.86 (0.01)	0.777	0.453
severe	0.33 (0.03)	0.31 (0.03)	0.39 (0.03)	0.449	0.014

† The second follow-up measurement was scheduled at 6 months in fibromyalgia and at 9 months in ankylosing spondylitis.

‡ paired t-test

The observed slight changes in reported preferences may result from real changes in the patients' health states. However, in fibromyalgia no significant changes in patients' health states were found on a variety of clinical and quality of life outcome measures.²⁰ Moreover, there is some evidence to support the hypothesis that patients' valuations of states of health are not influenced by their actual health state.³⁸ For example, ankylosing spondylitis patients' health improved as measured on a number of outcomes, including a global assessment scale, but their valuations of the reference states remained relatively stable.

A patient's true preference may change over three months and the preference at one time may not be representative of the patient's long-term preference.³⁹ This hypothesis is supported by the higher reliabilities that were found when the first 15 fibromyalgia patients from the control (no intervention) group to report for the 6 months follow-up were asked to return for a 4-week test-retest reliability assessment.⁴⁰ In this assessment, rating scale ICCs of 0.56 and 0.67 were found for the patient's own health state and the mild reference state respectively. The ICC of the standard gamble utilities was 0.66 for the patient's own health state, 0.74 for the mild and 0.94 for the severe reference state. The generally higher ICCs for the severe reference state are partly due to the fact that negative utilities for the severe reference state were recoded to zero.

Standard error of measurement

Another way of looking at the reliability of our utility measurements is by looking at the measurement error.^{15,28,29} Each single patient's preference measurement contains some measurement error, which causes part of the variance among the scores. The standard error of measurement, which is the standard deviation of the measurement error, was calculated to be 0.13 for the standard gamble (scale from 0-1) in fibromyalgia and 0.11 for the standard gamble in ankylosing spondylitis. For the rating scale method (scale from 0-100), the standard deviation of the measurement error was 10 in fibromyalgia and 8 in ankylosing spondylitis. These figures suggest that both methods contain considerable measurement error, implying relative instability of an individual patient's preferences. This supports the notion that utilities may be less useful for individual decision-making than for group decision-making, as was indicated by the relative stability of preferences over time shown in table 4.

7.4.3 Internal consistency of the standard gamble

Basic reference gambles versus chained gambles

According to the axioms of expected utility theory, the outcomes of the gamble should not influence a patient's utility for a particular health state. The patients are supposed to adjust their indifference probability to allow for alterations in the gamble outcomes. At 9 months follow-up 80 consecutive ankylosing spondylitis patients were asked to value their own health state both in comparison to perfect health and death and in comparison to perfect health and the severe reference state. The latter was the first of a chained pair of gambles. According to expected utility theory, there should be no difference in utilities elicited by a basic reference gamble or a chained gamble. However, the mean utility value of 0.83 in case of the basic reference gamble was statistically significantly lower than the mean utility of 0.87 when the severe reference state was used as a gamble outcome in a chained pair of gambles (paired t-test; $p=0.018$). This is in accordance with earlier findings.^{18,41} A majority of the patients (75%) assigned lower utilities in the basic reference gamble than in the chained gambles. Table 5 gives an example of this phenomenon for one patient.

Table 5. Difference between a chained and a basis reference gamble

SG question	Health state being valued	Gamble outcomes [†]	p [‡]	U _{in} [§]
2 (chained)	own	perfect health (1) severe ref. state (0.55)	0.75	0.89
5 (basic)	own	perfect health (1) death (0)	0.85	0.85

† Utilities for outcomes in parentheses

‡ Indifference probability for gamble

§ Utility for a patient's own health state

The fact that the chained gambles resulted in higher utilities than the basic reference gamble does not imply that patients took a smaller risk when death was replaced by the severe reference state. On the contrary, generally, patients took a greater risk in the gamble where the severe reference state was the worst outcome than in the gamble where death was the worst outcome. The eventually higher utilities in the chained gamble resulted from a combination of relatively small differences between the indifference points in the chained and basic reference gamble and the relatively high utilities assigned to the severe reference state (see example in table 5). This finding is in contrast to Llewellyn-Thomas *et al.* who found that raters were prepared to take a greater risk in gambles when death was the worst outcome.¹⁸ This difference may be explained by the fact that Llewellyn-Thomas *et al.* elicited utilities from cancer patients, to whom death usually is a real risk, whereas death to fibromyalgia or ankylosing spondylitis patients is not or only a remote issue.

Since the standard gamble method seems to be susceptible to the characteristics of the worst outcome of the gamble, Hellinger and Llewellyn-Thomas conclude that this method is internally inconsistent.^{17,18} However, when a change of focus or a change of reference point occurs as a result of a change in gamble outcomes, the preference shifts are not necessarily illogical.

Inconsistent Responses

Respondents are expected to provide preferences that are consistent with the natural underlying order of our health state descriptions. In other words, dominance violations should not occur. For the standard gamble as applied in our studies, dominance implies that when a patient's own health state is compared to the mild reference state and all 6 dimensions indicate a better (worse) or equal functional level, the utility the patient assigns to his own health state should be higher (lower) or equal to the utility of the mild reference state. However, 7 fibromyalgia patients—each only once—did not provide preferences in accordance with this expectation. In ankylosing spondylitis dominance was violated by 17 patients on the standard

gamble. On the total number of questions this number of dominance violations is rather low. Moreover, in 17 of the 24 (71%) inconsistent answers, the difference between the utilities of the two compared health states was smaller than the standard deviation of the measurement error reported in the previous section. Thus, most of the inconsistent responses fell within the bounds of the measurement error of 0.13 in fibromyalgia and 0.11 in ankylosing spondylitis.

In our studies every patient preferred his or her baseline health state to the severe reference state. Therefore, the valuation of a patient's follow-up health state in a gamble with perfect health and the severe reference state should result in a higher or equal indifference probability than the valuation of the same health state in a gamble with perfect health and the patient's baseline health state. When every follow-up measurement was checked, we found that 6 (9%) fibromyalgia patients and 6 (4%) ankylosing spondylitis patients violated this rule once. It would also be expected that, if death is regarded as worse (better) than or equal to the severe reference state, the indifference probability of the patient's own health state when compared to perfect health and death should be higher (lower) than or equal to its indifference probability when compared to perfect health and the severe reference state. This was checked for the 80 ankylosing spondylitis patients who were asked the additional fifth standard gamble question. Twelve of them (15%) gave an inconsistent response. Again the majority (18) of these 24 inconsistent responses fell within the bounds of the measurement error.

Overall, about 21% of all patients gave an inconsistent response. By means of a logistic regression, we enquired whether the patients who gave inconsistent responses were somehow different from the patients who did not. In ankylosing spondylitis we found that, when controlling for the influence of all other variables, males were more likely to give an inconsistent response than females. Otherwise, no differences with respect to age, duration of illness, education and marital status was found. We have no explanation for this gender related difference. Perhaps it is due to chance.

The inconsistencies described above are perhaps due to the fact that the health state descriptions cover too many dimensions for some people to include every dimension in their overall valuation. Concentration on just one or two dimensions that are considered important can lead to inconsistent responses. Some patients assigned higher preferences to their own follow-up health state when their baseline health state was used as the worst outcome of the gamble than when the severe reference state was used. This may be more an indication of appreciation for even a small improvement than of a change in preference.

Extreme Risk Averse Behavior: assigning a utility of 0.95 to each health state

Eight out of 85 (9%) fibromyalgia patients reached the same indifference probability of 0.95 for all health states being valued in the first three standard gamble questions on at least one of the measurement times. Eleven out of 144 (8%) ankylosing spondylitis patients assigned a value of 0.95 to all three health states at least once. One fibromyalgia patient and one ankylosing spondylitis patient did this consistently at baseline and each follow-up measurement. When controlling for other patient characteristics, fibromyalgia patients in which this phenomenon was found tended to be somewhat older than the patients in which this was not found. However this difference was not statistically significant (logistic regression; Wald statistic = 2.975; $p=0.08$).

Apparently these patients were never willing to take a larger than 10% risk of getting the worst outcome, irrespective of the severity of illness in the health state being valued. This behavior can be explained by a general aversion to gambling.¹³ Such a reluctance to comply with the standard gamble questions reflects either reluctance to face the reality of the decision problem, reluctance to bear decision-making responsibility or inability to grasp hypothetical, unrepresentative experiments presented in a necessarily simple and abstract way. It may have been too difficult for these patients to imagine well-being associated with a hypothetical health state or they may have underestimated their ability to cope with the severe reference state.

The difference between the individual and the group perspective also sheds some light on this phenomenon.⁴² Risk might seem higher from an individual perspective than from an aggregate, group perspective. A single patient will either become perfectly healthy or severely ill (or die in case of standard gamble question 3), and will never receive the average burden or average benefit. Therefore the expected utility may seem very abstract to a single patient in the 1-game setting, but not in the 100 game setting.⁴³

Finally, the probability steps of 10%, from 100/0 to 90/10 etc., may have been too large. Changes of 5% (100/0, 95/5, 90/10 etc.) or even 1% (100/0, 99/1, 98/2 etc.) could have produced a difference in utilities between the different health states. However, findings by Kahneman and Tversky suggest that probabilities of less than 0.1 and greater than 0.9 are difficult for people to handle.³¹

7.4.4 A health state worse than death

In standard gamble question 3 the severe reference health state is valued against perfect health and death. Some 41 of the 85 fibromyalgia patients (48%) and 78 of 144 (54%) ankylosing spondylitis patients indicated at least once that the severe reference state was worse than death. This means that they were prepared to accept a 100% risk of dying to avoid this state. Eighteen fibromyalgia patients and 20 ankylosing spondylitis patients preferred death to the severe reference state at all

3 measurements. When asked explicitly, all these patients confirmed that they would rather die than live in the severe reference state.

As mentioned before, when patients indicated that the severe reference state was worse than death, a utility of zero was assigned to this state. To actually measure the magnitude of a negative utility, Torrance has suggested a slight modification of the usual standard gamble question.¹⁰ In this modified question, patients are offered a choice between a sure death or a gamble with chance p of perfect health and chance $1-p$ of living irreversibly in the severe reference state. This question is presented by asking patients to imagine that they suffer a rapidly progressing terminal disease which -if left unattended- will lead to death. However, if treated there is a chance of gaining perfect health or of becoming like the severe reference health state for the remainder of their life. The utility of the severe reference state is calculated, from the indifference probability, as $-p/(1-p)$.

At the 3 and 9 months follow-ups all the ankylosing spondylitis patients who indicated that the severe reference state seemed worse than death (51 at 3 months and 36 at 9 months) were asked this standard gamble question. At 3 months follow-up, the 5% trimmed mean utility of the severe reference state was -0.16 with a 5% trimmed standard deviation of 0.19; at 9 months follow-up the 5% trimmed mean utility of the severe reference state was -0.18 with a 5% trimmed standard deviation of 0.34. The mean and standard deviation were trimmed 5% because the distribution of the answers to standard gamble question 3 is heavily skewed to the left. The standard gamble to assess the magnitude of a negative utility can result in utilities smaller than -1 , thus causing the upper end of the utility scale (from 0 to 1) to be shorter than the lower end of the scale (from 0 to -19 or less, depending on the size of the probability increments in the measurement instrument). At 3 months follow-up 1 patient assigned a utility of -3.00 and at 9 months follow-up 3 patients assigned a utility lower than -1 (-1.86 , -3.00 and -19.00) to the severe reference state.

The finding that for so many patients death was not the worst imaginable outcome has led some authors to conclude that death is indeed not the logical zero point for a utility scale.⁴⁴ Although we certainly recognise the existence of health states valued worse than death, the allowance of negative utilities is problematic. Furthermore, the descriptive validity of negative utilities may be questioned because probably few people would act in accordance with their statement that they would be better off dead. The occurrence of utilities below -1 is problematic too. Based on the large number of negative utilities for the severe reference state used as the anchor point in our gambles and the observed magnitude of these utilities (some of which were much lower than -1), we conclude that, for the present, it is more convenient to use death as the anchor point in the gambles. Moreover, this increases the comparability of utilities measured in different patient groups and in different studies.

7.5 Conclusions and discussion

Table 6 provides an overview of the most important findings, possible explanations and preliminary implications reported in the previous sections.

In the introduction of this article two different approaches to utility measurement in clinical trial settings were distinguished. One can either use prepackaged systems (e.g. the Quality of Wellbeing scale, Rosser's Disability Distress scale or the recently developed EuroQol⁴⁵) or directly elicit preferences from patients. When using prepackaged systems one needs to acquaint oneself with the method on which the underlying utilities were based. As has been reported before and reconfirmed in our studies, standard gamble preference scores for a particular health state are significantly higher than rating scale preference scores for that same health state. Hence, since the Quality of Wellbeing scale, the EuroQol and Rosser's Disability Distress scale are based on rating scales, these prepackaged instruments would be expected to produce lower scores than the McMaster Health Utility Index which is based on the standard gamble. When measuring preferences directly one again has the choice between preference rankings or choice-based methods. Moreover, if one decides to use the standard gamble one can also choose between various ways of taking the measurements: between basic reference gambles or chained gambles and between absolute or relative change questions. It was found in our studies that these different versions of the same method resulted in statistically significant utility differences. Chained gambles resulted in higher utilities than the basic reference gamble and absolute changes in utilities were smaller than relative changes measured directly in comparison to the baseline health state.

Utilities are proposed as a decision aid in two different contexts: 1) where choices have to be made between alternative therapies for the same individual,²⁶ and 2) where choices have to be made between alternative ways of allocating limited resources among different health care activities serving the same or different patient groups.⁴⁶ There is cause for misgivings regarding the use of utilities for clinical decision analysis in the first context. For example, a number of inconsistencies were found in the responses of single patients to different standard gamble questions. It is likely that these reflect the underlying measurement error in taking a single preference measure. Our study indicates that the standard deviation of the measurement error ranged from 0.11 to 0.13 for the standard gamble and from 8 to 10 for the rating scale. Overall, more than 70% of the inconsistent responses fell within the bounds of the measurement error. The relatively poor stability of measurements from an individual patient limits the use of utilities based on a single measurement only for individual decision-making. This increases the need to undertake the measurements repeatedly to average out the measurement noise within the individual.

Table 6. Summary table

Most important findings	Possible explanations	Implications
1 Comparing RS and SG utilities†		
- SG utilities significantly higher than RS utilities	- response spreading on the RS - risk attitude incorporated in the SG - cognitive and emotional factors (eg. regret theory)	- prepackaged utility measurement instruments based on the SG are bound to produce higher utilities - costs per QALY will be sensitive to the method of utility measurement used
2 Reliability of RS and SG		
- the ICCs of the reference health states were higher for the SG than for the RS, but low for both methods - mean scores of both methods were rather stable - the standard error of measurement was about 0.12 for the SG (0-1 scale) and about 9 for the RS (0-100 scale)	- instability of intra-individual valuations due to difficulties in valuing hypothetical states, but stability of mean values at the group level - substantial measurement error inherent to utility measurement - precision decreases as choice is introduced	- limited use of utilities for individual decision making in clinical decision analysis - more confidence in the use of utilities for program evaluation - need to undertake repeated measurements
3 Internal consistency of the standard gamble		
- chained gambles resulted in higher utilities than basic reference gambles - about 21% of all patients gave at least 1 inconsistent response - about 70% of the inconsistent responses fell within the bounds of the standard error of measurement - almost 10% of all patients assigned a utility of 0.95 to three very different health states; they were not willing to take any risk	- outcomes of the gamble influence an individual patient's utility for a particular health state - inconsistent responses are to a large extent due to measurement error - general aversion to gambling - inability to understand hypothetical, abstract questions - risk may seem higher from an individual than from a group perspective - probability increments of 10% are too large - difficulty of SG method - health state descriptions cover too many dimensions	- costs per QALY depend on the outcomes used in the SG - repeated measurements may reduce inconsistency - SG not suited for everyone
4 A health state worse than death		
- about 50% of the patients valued the severe reference state used in the chained gamble as worse than death on at least 1 occasion - negative utilities smaller than -1 occurred	- recognition of the existence of health states worse than death - low validity of the response	- an appropriate way to incorporate negative utilities is needed - searching for alternative ways to calculate negative utilities

† RS = Rating Scale; SG = Standard Gamble; ICC = Intra-class Correlation Coefficient

Even though the 3 months test-retest reliability was not very high, the relative stability of the mean utilities over time on the group level gives some confidence in the use of utilities for program evaluation in cost-utility analysis. However, since generally only a single estimate of utility is used in cost-utility analysis, this analysis may be very sensitive to the method used to elicit utilities. Recently Hornberger has shown that the effect of different methods on the final cost-utility may be considerable.⁴⁷ It is as yet premature to suggest a preferred method of utility measurement. Neither the rating scale nor the standard gamble method seems superior. As to the rating scale, there remain fundamental doubts about the interval properties of the scale.⁴⁸ Moreover, the repeatability of this method was found to be lower than that of the standard gamble. As to the standard gamble, it would be interesting to determine to what extent the observed difference between basic reference gambles and chained gambles influences the final results of cost-utility analysis.

Many of the inconsistencies and responses that seem to violate expected utility theory which we found may be associated with the description of the severe reference state used as the worst outcome in the first and as the certain outcome in the second of a chained pair of standard gambles. The health state of a severely ill person may be so hard to imagine that patients put all their effort into understanding it, instead of paying attention to what is really being asked. This possible explanation is supported by the fact that a number of patients changed their opinions as to whether the severe state was better or worse than death. Overall about 50% of the patients indicated at least once that the severe reference state was worse than death. When attempting to measure how much worse than death, a small number of highly negative utilities occurred that greatly influenced the mean utility. These findings and the difficulties of handling extreme negative utilities may argue in favor of using death as the worst outcome of the gamble. Death may be a more imaginable zero point to anchor the scale. On the other hand, the existence of health states worse than death should be recognised. Finding ways to handle negative utilities in utility and QALY calculations is one of the most challenging problems facing the researchers involved.

As preference scores elicited by the standard gamble seem to be susceptible to the way questions are presented and the endpoints used, doubts may be raised about the validity of expected utility theory. However, since there is no evidence that less formal procedures to guide individual therapy decisions and resource allocation decisions are any less susceptible to the effects of different methods of presentation and various biases, it is not appropriate to reject utilities as useful outcome measures. Moreover, as Torrance and Feeny point out, expected utility theory may be regarded as normative as opposed to behavioral.²⁵ This theory describes how an individual should behave if he or she wished to act rationally in order to maximize the expected utility. It does not describe how an individual actually makes a decision under uncertainty. Perhaps patients would have made more rational responses if they had been better informed about the meaning of their responses and

the consequences of their choices. Furthermore, many of the observed inconsistencies or preferences that do not seem to fit expected utility theory are not necessarily illogical. Some of the alternatives to expected utility theory, such as prospect theory and regret theory, may help to explain several of the seemingly inconsistent answers from a more behavioral perspective. The challenge is to explore the potential contribution of such theories to utility measurements in health care decision settings.

7.6 Acknowledgements.

We would like to thank the anonymous reviewer, George Torrance, Han Bleichrodt, Grant Rhodes, Silvia Evers, Margreet Jansen, Linda Heurman and Mariëlle Goossens for their valuable comments on an earlier draft of this paper.

This study was supported by grants from the Netherlands Health Research Promotion Program and the National Rheumatic Disease Foundation of the Netherlands.

7.7 References

1. Feeny D, Labelle R, Torrance GW. Integrating Economic Evaluations and Quality of Life Assessments. In: Spilker B (ed.). *Quality of Life Assessment in Clinical Trials*. Ltd. New York, Raven Press, 1990.
2. Bell MJ, Bombardier C, Tugwell P. Measurement of functional status, quality of life and utility in rheumatoid arthritis. *Arthritis and Rheumatism* 1990;33:591.
3. Gerard K. Cost-utility in practice: A policy maker's guide to the state of the art. *Health Policy* 1992;21:249.
4. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiology* 1978;7:347.
5. Kaplan RM, Bush JW. Health status: types of validity and the index of well-being. *Health Serv Res* 1976;478.
6. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res* 1982;30:1042.
7. Bombardier C, Ware J, Russell IJ, Larson M, Chalmers A, Read JL, and the Auranofin Cooperating Group. Auranofin Therapy and Quality of Life in Patients with Rheumatoid Arthritis. Results of a Multicenter Trial. *Am J Med* 1986;81:565.
8. Thompson MS, Read JL, Hutchings HC, Paterson M, Harris ED. The Cost Effectiveness of Auranofin: Results of a Randomized Clinical Trial. *J Rheumatol* 1988;15:35.
9. Mehrez A, Gafni A. Quality-adjusted Life Years, Utility Theory, and Health-years Equivalents. *Med Decis Making* 1989;9:142.
10. Torrance GW. Measurement of health state utilities for economic appraisal. A Review. *J Health Econ* 1986;5:1.

11. Neumann Von J, Morgenstern O. Theory of games and economic behavior. Princeton University Press, 1944 (1st ed.), 1947 (2nd ed.).
- 12a. Froberg DG, Kane RL. Methodology for measuring health-state preferences-I: Measurement Strategies. *Journal of Clinical Epidemiology* 1989;42:345.
- 12b. Froberg DG, Kane RL. Methodology for measuring health-state preferences-II: Scaling Methods. *Journal of Clinical Epidemiology* 1989;42:459.
- 12c. Froberg DG, Kane RL. Methodology for measuring health-state preferences-III: Population and Context Effects. *Journal of Clinical Epidemiology* 1989;42:585.
- 12d. Froberg DG, Kane RL. Methodology for measuring health-state preferences-IV: Progress and a Research Agenda. *Journal of Clinical Epidemiology* 1989;42:675.
13. Mulley AG. Assessing Patients' Utilities. Can the Ends Justify the Means? *Medical Care* 1989;27:S269.
14. Read JL, Quinn RJ, Berwick DM, *et al.* Preferences for health outcomes: comparisons of assessment methods. *Med Decis Making* 1984;4:315.
15. Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio Econ Plan Sci* 1976;10:129.
16. Wolfson AD, Sinclair AJ, Bombardier C, McGreer. Preference Measurement for Functional status in Stroke Patients: Interrater and Intertechnique Comparisons. In: Kane RL and Kane RA (eds.). *Values and Long-Term Care*. Lexington Books, D.C. Heath and Company, Lexington, 1982.
17. Hellinger FJ. Expected Utility Theory and Risky Choices with Health Outcomes. *Medical Care* 1989;27:273.
18. Llewellyn-Thomas HA, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The Measurement of Patients' Values in Medicine. *Med Decis Making* 1982;2:449.
19. Hidding A, van der Linden Sj, Boers M, Gielen X, Kester A, de Witte L, Dijkmans B, Moolenburgh D. Is group physical therapy superior to individual therapy in Ankylosing Spondylitis. *Arthritis Care Res* 1993;6:117-125.
20. Van Santen-Hoeufft M, Bolwijn P, Van der Linden Sj. Effectiveness of fitness and biofeedback training in fibromyalgia. (submitted).
21. Wolfe F, Smythe HA, Yunus MB, *et al.* The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the multicenter criteria committee. *Arthritis Rheum* 1990;31:1135.
22. Linden van der SJ, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361.
23. Bennett K, Torrance G. Methodologic challenges in the development of utility measures of health-related quality of life in Rheumatoid Arthritis. *Controlled Clinical Trials* 1991;12:118S.
24. Bakker CH, Ruten-van Mólken M, Van Doorslaer E, Bennett K, Van der Linden Sj. Health Related Utility Measurement in Rheumatology: An Introduction. *Patient Education and Counseling* 1993;20:145.
25. Torrance GW, Feeny D, Utilities and Quality-Adjusted Life Years. *Int J Techn Ass in Health Care* 1989;5:559.
26. Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. Philadelphia: W.B. Saunders Company, 1980.
27. Norusis/SPSS INC. *Advanced Statistics SPSS/PC+*. Chicago, 1988.

28. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Statistics and Strategies for Evaluation. Controlled Clinical Trials* 1991;12:142S-158S, 1991.
29. Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to their Development and Use.* Oxford University Press, 1989;83-89.
30. Kaplan RM, Bush JW, Berry CC. Health status index: category ratings versus magnitude estimation for measuring levels of well being. *Med Care* 1979;17:501.
31. Kahneman D, Tversky A. Prospect theory. An analysis of decision under risk. *Econometrica* 1979;47:263.
32. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;211:453.
33. O'Brien BJ, Elswood J, Calin A. Willingness to accept risk in the treatment of rheumatic disease. *J Epid Comm Health* 1990;44:249.
34. Loomes G, Sugden R. Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 1982;92:805.
35. Loomes G, Sugden R. Some implications of the more general form of regret theory. *Journal of Economic Theory* 1987;41:270.
36. Loomes G, Starmer C, Sugden. Observing violations of transitivity by experimental methods. *Econometrica* 1991;59:425-439
37. O'Connor AM, Boyd NE, Till JE. Methodological Problems in Assessing Preferences for Alternative Therapies in Oncology: The Influence of Preference Elicitation Technique, Position Order and Test-retest Error on the Preferences for Alternative Cancer Drug Therapies. *Nursing Research: Science for Quality Care; Proc 10th National Nursing Research Conference.* Toronto: University of Toronto, 1985:49.
38. Llewellyn-Thomas HA, Sutherland HJ, Tritchler DL, Lockwood GA, Till JE, Ciampi A, Scott JF, Lickley LA, Fish EB. Benign and malignant breast disease. The relationship between women's health status and health values. *Med Decis Making* 1991;11:180.
39. Christensen-Szalanski JJJ. Discount Functions and the Measurement of Patients' Values. *Women's Decisions During Childbirth.* *Med Decis Making* 1984;4:46.
40. Bakker C, Rutten-van Mólken M, Van Doorslaer E, Bennett K, Van der Linden Sj. Feasibility of utility assessment by rating scale and standard gamble in ankylosing spondylitis or fibromyalgia. *Journal of Rheumatology* 1994 (in press)
41. Bleichrodt H. Testing the validity of expected utility theory in health state valuation: Some experimental results. *Institute for Medical Technology Assessment paper no 93.23, Rotterdam, Erasmus University of Rotterdam, 1993.*
42. Redelmeier DA, Tversky A. Discrepancy between medical decisions for individual patients and for groups. *New Eng J Med* 1990;322:1162.
43. Asch DA. Choices for individual patients vs groups. Letter to the editor. *N Eng J Med* 1990;323:922.
44. Haig THB, Scott DA, Wickett LI. The Rational Zero Point for an Illness Index with Ratio Properties. *Med Care* 1986;24:113.
45. The EuroQol Group. A new facility for the measurement of health related quality of life. *Health Policy* 1990;16:199
46. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? *Can Med Assoc J* 1992;146:473.

47. Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *J Clin Epidemiol* 1992;45:505.
48. Nord E. The validity of a visual analogue scale in determining social utility weights for health states. *Int J Health Planning and Management* 1991;6:234.

8

Statistical analysis of cost outcomes in a randomized controlled clinical trial

*Maureen P.M.H. Rutten-van Mölken¹, Eddy K.A. van Doorslaer²,
René C.J.A. van Vliet³.*

1 Department of Health Economics, University of Limburg.

2 Institute for Medical Technology Assessment, Erasmus University Rotterdam.

3 Department of Health Care Policy and Management, Erasmus University Rotterdam.

8.1 Abstract

This paper suggests an approach to deal with an estimation problem which is often encountered in analyzing the longitudinal cost data gathered in a clinical trial. The source of that estimation problem is twofold: 1) a considerable number of missing data due to treatment-related withdrawal of severely affected patients with high health care costs in only one the treatment groups and 2) a heavily skewed cost distribution due to rare high-cost events. The approach is illustrated using data from a trial comparing 3 different drug regimes.

In order to calculate costs per patient-year in case of selectively missing data we extrapolated the costs of patients with incomplete follow-up. Due to the skewness and the associated large variance in costs per patient-year, these costs cannot be analyzed using common parametric statistical methods relying on underlying normal distributions. A logarithmic transformation was performed to approximate a normal distribution, reduce the impact of extreme values and create similar size variances in the treatment groups. An ordinary least squares regression analysis of transformed data then standardized for differences in patient characteristics between the groups. For the retransformation, the so-called smearing estimate was used. This 'transformation-standardization-retransformation' approach enabled us to provide more consistent and efficient estimates of cost differences that were shown to be statistically significant and judged to be important.

8.2 Introduction

During the last two decades, the growing interest in economic evaluation studies in health care resulted in over 1800 articles, mainly published in medical journals.^{1,2} Most of these articles describe cost-effectiveness studies of various health care interventions conducted since 1985. During this period the number of published clinical trials exceeded by far the number of economic evaluation studies. Although less than 1% of the trials included an economic analysis, there is a growing tendency towards performing economic evaluations as part of clinical trials,³ because a clinical trial is the optimal study design to elicit unbiased data on the efficacy of health care interventions. Several authors have published guidelines for performing economic analysis in this area.⁴⁻⁸ The methodological issues that arise from the closer integration of economic and clinical research were recently updated and the importance of standardizing the methodology has been emphasized repeatedly.^{9,10}

One of the major differences between clinical and economic evaluations concerns the statistical analysis of data. Although clinical data are commonly subjected to thorough statistical analyses in order to test for the significance of the observed differences between therapies, the economic analysis is often restricted to

a straightforward calculation of the difference in costs. This difference is then presented as a point estimate. There are several reasons why statistical analysis of economic data is often impossible. One is that economic data are often collected retrospectively, on a non-patient-specific basis, so that the variance in the cost distribution cannot be calculated. Another reason is that the sample size is sometimes too small, so that the study lacks statistical power to test differences in costs, even though the power is sufficient to detect differences in effects. This may be due to the variance in cost data often being much larger than the variance in efficacy data.

In clinical analyses, uncertainty is taken into account by reporting mean differences along with accompanying confidence intervals. The traditional way of dealing with uncertainties in economic analyses is by means of sensitivity analyses. In such analyses, critical assumptions in the cost analyses are varied in order to test the sensitivity of the results and conclusions to such changes. Although sensitivity analyses are proposed in all guidelines, recent reviews of economic evaluation studies showed that the number of studies actually performing sensitivity analysis is rather limited.^{3,11,12}

In the case of the prospective, continuous collection of patient-specific cost data (i.e., volumes of medical resource use multiplied by a fixed cost for each of the resources) alongside a trial, the variation in cost data can be observed and statistical testing of economic hypotheses can be performed. Recently, O'Brien *et al.* have proposed a method for the formal statistical testing of economic hypotheses and for calculating confidence intervals for cost-effectiveness ratios.¹³ The latter is particularly important since the economic importance of a change in costs can only be judged in combination with the clinical importance of changes in effects.¹⁴ Therefore, it is valuable to know the difference between treatments in terms of a change in the ratio of costs to effects. Nevertheless, it will always be relevant for regulators and public health officials to know the size of the change in costs (and effects) independently, which is difficult to trace from ratio's.¹⁴

The aim of this paper is to discuss, in the context of randomized controlled trials, the analysis of longitudinal cost data which are skewed and contain missing observations as a result of selective withdrawal. Highly skewed cost distributions are commonly found in economic evaluations. The skewness results from a relatively small proportion of the patients causing a relatively high proportion of the total health care costs. Selective patient withdrawal is likely to occur when one treatment is considerably more effective than the other and/or when adverse events occur more frequently in one treatment group than in the other. The purpose of our analysis is to obtain an estimate with maximum precision of the difference in costs between treatments. Since our focus is on the methods of analysis, it is not our intention to discuss other important methodological issues associated with calculating costs, such as which types of costs to incorporate, how to collect these costs (from patients, hospital bills, third party payers), how to value these costs (by means of charges or opportunity costs) or the choice of discount rate.

To illustrate the proposed approach, real cost data are used. In the next section we briefly describe the clinical trial in which these data were collected. We do not aim at summarizing the results of this trial, except to the extent that they illustrate the particular problems and the approach suggested as a solution. In section 4 we discuss the problems encountered in the trial which made the cost analysis difficult. In section 5 we compare alternative ways of dealing with selective withdrawal when calculating costs. We suggest a specific approach for dealing with the statistical analysis of skewed data in section 6 and provide a discussion in the final section.

8.3 The asthma/COPD trial

The data we use as an illustration were obtained in a randomized controlled trial, with the cooperation of 274 patients aged 18-60 years who were suffering from asthma or COPD. The patients were enlisted between October 1987 and January 1989. The trial compared three different inhaled drug regimes: a) β_2 -agonist plus corticosteroid (n=91), b) β_2 -agonist plus anticholinergic (n=92), c) β_2 -agonist plus placebo (n=91). We will refer to these treatments as A, B and C. More details on the study and the results can be found elsewhere,^{15,16} but a very brief summary is given here. Baseline characteristics of the patients in the three groups were comparable. Treatment A was found to be more effective than treatments B and C. Treatment A led to significant and clinically very important improvements in the primary endpoints: lung function, hyper-responsiveness and symptoms. Because interim analysis has signalled the superiority of treatment A, the trial was terminated earlier than originally planned. No difference was found between treatments B and C. Very few side effects were reported and those reported were minor.

Annual drug acquisition costs of treatments A, B and C were 1129, 588 and 331 Dutch Guilders per patient respectively. Our aim was to investigate whether the additional costs of A and B compared to C were compensated for by reduced utilization of other health care services. The costs of the following services were included: 1) additional drug use, 2) additional non-scheduled outpatient contacts, 3) additional non-scheduled tests, 4) hospitalizations. We included only drug use, outpatient contacts, tests and hospitalizations which were due to asthma or COPD. Whether these costs were asthma or COPD related was judged by patients and physicians. We excluded the costs of the regular outpatient visits and lung function tests that were scheduled every three months for all treatment groups because of the trial. Data on health care resource use were collected every 3 months alongside the trial by means of: 1) semi-structured interviews held with the patients every three months during outpatient visits, 2) diaries kept by the patients, and 3) patients' records. For more details see Rutten-van M lken.¹⁷ All costs were based on

1989 prices and expressed in Dutch Guilders (NLG.). (The 1989 exchange rate was 1 US Dollar to 2.12. NLG.) The costs of drugs include pharmacy delivery costs. Estimates of the costs of additional non-scheduled outpatient contacts and lung function tests were based on real resource cost calculations which in turn were based on detailed cost accounting studies. Estimates of the costs of hospitalizations and a few, rarely performed tests were based on charges.

8.4 Problems encountered in the asthma/COPD trial

8.4.1 Problem 1: Selective treatment-related withdrawal

Two different reasons for incomplete follow-up can be distinguished. First, incomplete follow-up results from the earlier termination of the trial and second it results from premature withdrawal of patients from the trial. Whereas the first does not affect the analysis, the second does so considerably. Because the trial was terminated earlier than originally planned, not all patients reached the full 30 months follow-up. However, they all reached a follow-up of at least 21 months. This early termination is unlikely to affect the analysis, since the patients with incomplete follow-up due to this reason did not differ from the patients with complete follow-up. However, a considerable number of observations were missing as a consequence of premature withdrawal of patients unrelated to the termination of the trial. Figure 1 shows the cumulative number of withdrawals over the study period, starting immediately after randomization. Withdrawal rates were significantly higher in the groups receiving treatment B (49%) and C (48%) than in the group receiving treatment A (13%) (χ^2 test; $p < 0.0001$).¹⁶ Four categories of withdrawals can be distinguished. The first and largest category contains 70 of the 101 withdrawals (5 in A; 35 in B; 30 in C) all of whom dropped out because of an increase in pulmonary problems such as shortness of breath, wheezing, asthma attacks, respiratory infections etc. The second category contains 20 patients (1 in A; 6 in B and 13 in C) who were no longer willing to participate. The third category includes 3 patients who dropped out because of adverse events (2 in A because of hoarseness and tremor and 1 in C because of a sore throat after inhalation). Eight patients (4 in A; 4 in B) who withdrew due to other reasons unrelated to the disease or the treatment were assigned to the fourth category.

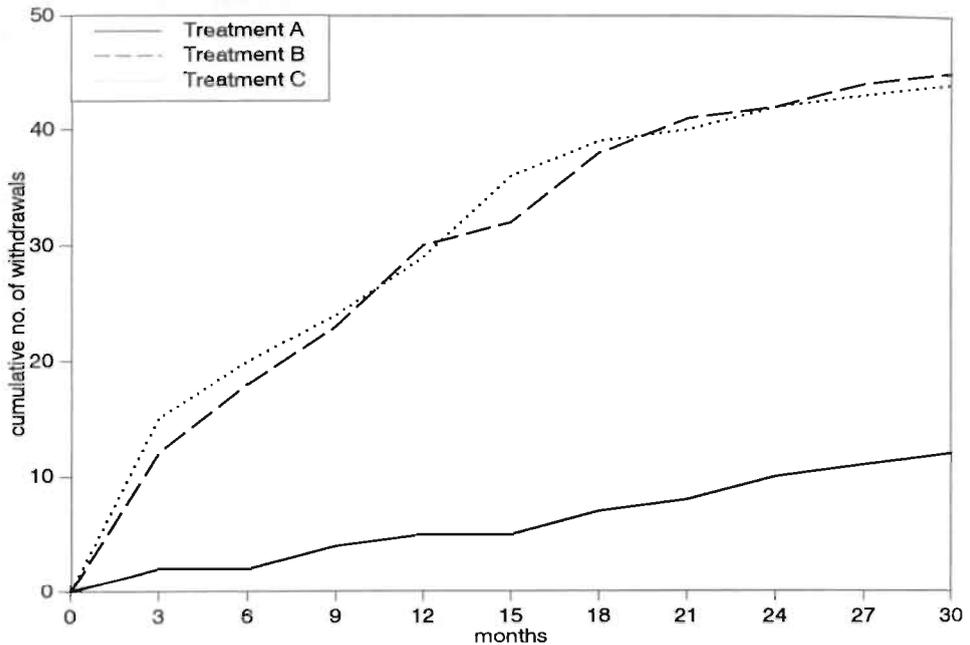


Figure 1. Cumulative number of withdrawals by treatment group

Thus, the large difference in withdrawal rates was clearly due to a lack of therapeutic effect of treatments B and C. As such, withdrawal rate turned out to be a very important outcome measure demonstrating the superiority of treatment A. Baseline characteristics were compared between withdrawals and non-withdrawals. Patients with a poorer lung function and more hyper-responsiveness at baseline -which could apparently be improved by neither treatment B nor treatment C- were more likely to drop out.¹⁶ Thus, the more severely affected patients, who are more likely to have higher health care costs, systematically dropped out of the trial. This assessment is supported by table 1 which shows that, in all treatment groups, the mean health care costs per quarter, if hospitalization costs were excluded, were more than twice as high for those who withdrew than for those who did not. They were almost six times as high if hospitalization costs were included. When compared to the non-withdrawals, the frequency distribution of costs among withdrawals shows smaller percentages of patients in the low cost categories and larger percentages of patients in the high cost categories.

Table 1. Distribution of costs per 3 months[†] among withdrawals and non-withdrawals in Dutch Guilders

	Including hospitalizations		Excluding hospitalizations	
	Withdr.	Non-withdr.	Withdr.	Non-withdr.
mean (SEM)	469.66 (155.90)	80.02 (12.24)	127.79 (14.38)	52.51 (4.04)
min - max	0 - 8897	0 - 1478	0 - 694	0 - 320
0-75 NLG	38%	74%	40%	76%
75-150 NLG	23%	17%	28%	19%
150-225 NLG	18%	3%	20%	3%
225-300 NLG	4%	1%	5%	1%
300-450 NLG	5%	2%	4%	1%
450-1000 NLG	5%	2%	3%	0%
> 1000 NLG	7%	1%	0%	0%

† Means are calculated by dividing each patient's total costs by the number of quarters they were in the study.

After their withdrawal patients were no longer followed, thus making a pure intention-to-treat analysis impossible. As can be seen from figure 1, 2 patients receiving treatment A, 12 patients receiving treatment B and 15 patients receiving treatment C dropped out even before their first follow-up visit. These 29 patients could not be included in the analysis, because no information on their costs was available. At baseline there were no dissimilarities in treatment groups with regard to patient's baseline-characteristics. However, as a result of these 29 patients dropping out prior to the first follow up visit, the populations represented by the various treatment groups may not be fully comparable. The patient group composition became slightly more favourable for treatment group A than for treatment groups B and C. Since the patients receiving treatment A had slightly better scores on a number of prognostic factors (e.g. less hyperresponsive, better lung function), correction for confounding was called for. For, the joint effect of these small dissimilarities between treatment groups, which were not significant when taken separately, may be considerable.

8.4.2 Problem 2: Skewed cost distributions

During the study there were 15 hospitalizations. The number of hospitalizations was not equally distributed over the treatment groups. In treatment group A only one patient was hospitalized for 6 days with associated costs of 7,096 NLG. In treatment group B 8 hospitalizations occurred with lengths of stay varying from 4 to 13 days and total associated costs of 43,534 NLG. In treatment group C the total costs of 6 hospitalizations with lengths of stay varying from 6 to 18 days amounted to 42,980 NLG. Table 2 shows that this relatively small number of hospital admissions has disproportionately affected the mean costs and the variance around the mean.

Table 2. Mean[†] (standard deviation) health care costs per patient per 3 months

Treatment group	including hospitalizations	excluding hospitalizations
A (n=89)	64 (109)	54 (59)
B (n=80)	236 (657)	85 (87)
C (n=76)	304 (1147)	88 (107)
Overall (n=245)	195 (748)	75 (87)

† Means are calculated by dividing each patient's total costs by the number of quarters they were in the study.

The hospitalizations cause the distribution of costs to be very skewed to the right. But, as table 1 already showed, the distribution is skewed even if the hospitalization costs are excluded. This skewness is caused by a relatively small number of patients with a relatively high number of additional outpatient contacts or expensive drugs. Due to this skewness, parametric tests that assume the observations to be random samples from normal distributions with the same variance are inappropriate. When applied nonetheless, these rare high-cost events reduce the power of these tests to detect significant cost differences between patient groups. We decided not to ignore the high-cost events but to include them in the overall statistical test of cost differences, while looking for ways to adequately account for their impact.

8.5 Selective withdrawal and cost estimation

In order to obtain an estimate of the cost differences between the treatments, an appropriate method for dealing with selective withdrawal had to be found. Three types of cost calculations are commonly undertaken to tackle this problem in cost analysis: 1) calculations based on complete data only, 2) cumulative cost approach and 3) patient-year approach. It will be shown that the latter is most appropriate. Other methods have been proposed for dealing with missing observations in analyses of clinical outcome data.^{18,19} These methods suggest the replacement of missing values by for example the most recent previous value or any other value as long as it is higher or lower than the observed values in the non-withdrawals. In case of the latter approach, the magnitude of the assigned costs does not matter, since the costs would be analyzed using statistical methods based on ranks. This may be sufficient when the only aim is to test whether a difference is statistically significant but it is not useful if the objective is to estimate the magnitude of this difference. Furthermore, such replacement approaches are unsuitable to analyze costs because they ignore the information available on costs and especially the variation in costs before withdrawal. Replacing missing values by the most recent previous value would imply, for some patients, carrying forward one measurement to a large number of follow-up visits. This would greatly affect the variance of the cost distribution. Finally, these replacement techniques may lead to wrong conclusions when the number of treatment related withdrawals differs between the treatment groups.¹⁹

8.5.1 Complete data only

First, we could simply calculate the mean costs per patient per year in each of the treatment groups using only the data of patients with complete follow-up. In clinical analysis, this approach is frequently used.¹⁹ However, since withdrawal was treatment-related, treatment comparisons based on complete data would certainly be biased. We would end up with very few patients who, moreover, were in relatively good shape at the start of the trial and remained so during the trial. Since the costs of the patients who withdrew were significantly higher than those of the patients who did not withdraw, it is obvious that the mean costs if we take only complete data sets will be lower than the mean for all participants, especially in treatment groups B and C. Thus, the difference in mean costs between treatment A on the one hand and treatments B and C on the other hand would be underestimated.

8.5.2 Cumulative cost approach

A second type of calculation is based on the assumption that the data missing within each treatment group are randomly distributed. In other words, it is assumed that every patient within any one treatment group has the same probability of dropping out. In that case the missing cost data during a certain period in a certain treatment group can be replaced by the mean of the observed costs in that same period within the same group. This approach is sometimes called the "cumulative cost approach" or "*sum-limit approach*"²⁰ and its assumptions are similar to an approach proposed to analyze longitudinal quality of life data.²¹ When applying the cumulative cost approach, we first calculated the average costs per patient for each 3-month period. Since the number of patients in the study declined, these 3-month averages were based on decreasing numbers of patients. The averages were then added up to give the average costs per patient during the entire study. For example suppose that the mean costs of 80 patients receiving treatment B are 150 NLG per patient during the first quarter. In the second quarter 6 patients drop out and the mean quarterly costs based on the remaining 74 patients are 110 NLG per patient. Thus, the cumulative cost approach suggests that the mean costs per 6 months are $150+110=260$ NLG for *each* of the 80 patients. However, it seems likely that the costs of patients who withdrew -had they continued in the trial- would have been higher than the average of those who did not withdraw in the same group. Since withdrawal rates were much higher in treatment groups B and C, the cumulative cost approach may also underestimate the difference in costs between treatment group A and the other 2 groups.

8.5.3 Patient-year approach

A third type of calculation we labelled the "*patient-year approach*". The underlying assumption of this approach is that the (non-observed) costs of patients after withdrawal equal the mean (observed) costs they had before they dropped out. In other words, the mean of the observed costs per quarter before withdrawal were linearly extrapolated to a one-year period. For example, suppose a patient's costs during the first three quarters of the study were 300, 200 and 400 NLG respectively. Thus, the mean costs per quarter for this patient are 300 NLG. This patient drops out during the fourth quarter. Then the estimated mean costs for this patient are $12/3 \times 300 = 1200$ NLG per year. When a patient drops out after more than 1 year of follow-up his or her annual mean costs are calculated in the same way. The essential aspect of this approach is that, for early withdrawals, it carries forward the mean of all available measurements. The patient-year approach may give a more realistic estimate than the first two approaches of the mean annual costs had patients continued their initial drug treatment instead of dropping out, even if their clinical condition had deteriorated. It is possible that the actual rise in

the costs of the patients who withdrew, had they stayed in the trial, would have been progressive. In that case, the patient-year approach is still underestimating costs, but not as seriously as in the other two approaches.

8.5.4 Comparing the approaches

Table 3 provides a comparison of the estimated mean costs per patient per year and the associated standard errors, calculated using the three approaches described above. According to the first approach the savings in health care utilization due to treatment A are estimated to be 38% of the costs of treatment C (the placebo), whereas using the second and third approach the savings are estimated to be 64% and 79%, respectively, of the costs of treatment C.

Table 3. Estimated mean (standard error) costs per patient per year

Approach	Treatment A	Treatment B	Treatment C
Complete follow-up only	193 (28)	444 (141)	312 (107)
Cumulative cost approach	283 (not applic.)	641 (not applic.)	795 (not applic.)
Patient-year approach	254 (46)	945 (294)	1216 (526)

No standard errors can be given for the estimates based on the cumulative cost approach, because this approach gives only cumulative group averages, not directly based on individual observations. For this reason, statistical analysis of cost differences based on the cumulative cost approach is not possible. Analyzing only complete data reduces the standard errors because the higher costs of withdrawals are not included. Because the approach using only complete data seriously underestimates both the mean costs and the associated variation, statistical analysis of costs differences calculated using this method does not seem appropriate. In general, the patient-year approach does not reduce the standard errors because it calculates each patient's mean costs per quarter based on all available observations and then extrapolates that mean and the associated standard error to a one-year period. This may even increase the variation in costs for patients withdrawing before one year, whereas it may decrease the variation in costs for patients with examination periods exceeding a year. We will concentrate, in the next section, on the appropriateness of statistically analyzing cost differences estimated using the patient-year approach, focusing on a particular method of adjusting for the overestimation due to rare high-cost events.

8.6 Skewness and statistical testing

The calculation of costs in itself may be problematic to some extent, but the statistical analysis is even more troubled. As a result of the large variation in costs per patient-year, the sample means in the treatment groups as given in table 3 are relatively unreliable estimates of the population means. If differences in untransformed costs based on the patient-year approach were to be analyzed by means of frequently used parametric analyses such as t-tests or analyses of variance, the 95% confidence interval would be very wide but meaningless because these analyses are based on the assumption of an underlying normal distribution. Most of the parametric methods require that the sample has been taken from a normal distribution with equal variances. However, the distribution underlying our sample is skewed to the right and the standard errors associated with costs per patient-year are not only very large but appear to differ between the treatment groups. A proper four-step approach to analyze this type of data is described below.

Step 1. Transformation

One of the suggested econometric solutions to the skewness problem is data transformation to attempt to symmetrize the distribution. One possibility is a simple log-transformation, which is a special case of the more general Box-Cox transformation.²² This approach was also adopted by the researchers of the Rand Health Insurance Study who were confronted with skewness in medical cost data.²³ If we denote the annualized untransformed costs per patient by Y_i and the transformed cost by Z_i , then Y_i and Z_i are related by:

$$(1) \quad Z_i = \log(Y_i + 10)$$

The results of this transformation are twofold. First, the skewness is reduced and a normal distribution is approximated.^{22,23} Second, similar size variances in the respective groups are obtained.²⁴ Hartley's test for equality of more than two population variances²⁵ indicated that the standard deviations of the log-transformed costs (which were 1.423, 1.574 and 1.570 in treatment groups A, B and C respectively) did not differ by treatment group (test-statistic=1.22; critical value at the 0.05 level=1.67).

Twenty-one patients (10 in A, 7 in B and 4 in C) had no health care costs other than the costs of the study drugs. In general, this would call for a Tobit, or for a two-part model in which first the probability of having non-zero costs is estimated and secondly the magnitude of the costs given positive costs. However, because the number of patients with zero costs was rather small (<10%) and in order to limit the number of parameters to be estimated and to keep the retransformation procedure tractable, we adopted a simpler approach by adding a

positive constant to the costs in order to avoid taking the logarithm of 0. We added 10 in order to shift the top of the logarithmic distribution further from 0, thereby better approximating a normal distribution of the residuals in the linear regression analysis that was performed on the transformed scale. The shift also increases the efficiency (i.e. reduces the variance) of the regression estimates of the parameters²⁶. We added 10 because a Kolmogorov-Smirnov test indicated that this value (as compared to values of 1 and 5) gave the best approximation to a normal distribution of the residuals.

Step 2. Regression analysis, correction for confounding and effect modification

After this transformation we performed an ordinary least squares regression analysis to test for the partial effect of treatment on the transformed costs per patient-year. The patients' baseline characteristics were also included in this regression analysis to correct for possible confounding. Correction for confounders, even if they would have been evenly distributed across the treatment groups, reduces the variance of the parameter estimates.²⁷ The regression equation is given by:

$$(2) \quad Z_i = \alpha + X_i\beta + \epsilon_i$$

where α is the value for the intercept, the values of X_i are the vectors of explanatory variables, β is a vector of regression coefficients and the values of ϵ_i are the residual errors with a mean of 0 and a variance of σ^2 .

Results of the least squares regression analysis of the transformed costs per patient-year are shown in the first column of table 4. As is indicated by the adjusted R^2 , the proportion of the variation in log-costs explained by the regression model is 12%. Other things being equal, patients receiving treatment A and patients with a better lung function had significantly lower health care costs. Patients who had used inhaled corticosteroids previously had significantly higher costs. There was no significant difference in costs between treatments B and C. The 95% confidence intervals around all other explanatory variables contained zero, thus indicating no effect on the log-transformed costs.

Table 4. Ordinary least squares regression analysis of mean costs per patient-year[†]

explanatory variables [‡]	without interaction		with interaction	
	$\hat{\beta}$ (se)	p-value	$\hat{\beta}$ (se)	p-value
dummyA	-0.671 (0.231)	0.004	-0.302 (0.477)	0.527
dummyB	0.109 (0.236)	0.643	0.747 (0.525)	0.156
sex	0.027 (0.212)	0.900	-0.118 (0.206)	0.567
smoking	-0.201 (0.215)	0.350	-0.269 (0.378)	0.477
allergy	0.062 (0.224)	0.782	0.057 (0.213)	0.789
prior use of corticosteroids	0.641 (0.192)	0.001	0.605 (0.183)	0.001
age	-0.003 (0.009)	0.739	-0.007 (0.008)	0.414
lung function	-0.021 (0.007)	0.002	-0.024 (0.006)	0.000
hyper-responsiveness	0.056 (0.047)	0.242	0.156 (0.076)	0.041
reversibility	-0.005 (0.013)	0.674	0.002 (0.020)	0.939
dummyA x smoking			1.041 (0.484)	0.033
dummyB x smoking			-0.846 (0.512)	0.099
dummyA x reversibility			-0.018 (0.028)	0.523
dummyB x reversibility			-0.016 (0.028)	0.562
dummyA x hyperresp.			-0.302 (0.104)	0.004
dummyB x hyperresp.			-0.070 (0.104)	0.499
intercept	6.545 (0.741)	0.000	6.706 (0.789)	0.000
adjusted R ²	0.120		0.211	
mean squared error	2.146		1.922	
F-value	4.313	0.000	5.090	0.000

† dependent variable: $\log(\text{costs per patient-year} + 10)$

‡ explanatory variables at baseline:

- dummy A: 1=treatment A, 0=treatment B or C
- dummy B: 1=treatment B, 0=treatment A or C
- sex: 1=male, 0=female
- smoking: 1=smoker, 0 non-smoker
- allergy: 1=allergic, 0=non-allergic
- prior use of steroids: 1=prior use, 0=no prior use

- lung function: continuous variable expressed in Forced Expiratory Volume in 1 second as % of the value predicted for age, sex, weight and height (FEV_1 % pred.). The higher the % the better the lung function

- hyper-responsiveness: continuous variable expressed in $-\log_2 PC_{20}$, which is the value of the 2log of the provocative concentration of histamine causing a 20% decrease in FEV_1 multiplied by -1.

The higher the $-\log_2 PC_{20}$, the more hyperresponsive a patients is.

- reversibility: the change in FEV_1 % pred. after drug inhalation

- the 6 explanatory variables containing an * are interaction terms

A regression analysis that also included a dummy variable distinguishing between non-withdrawals with complete follow-up and non-withdrawals with incomplete follow-up due to earlier termination of the trial did not change the overall results.

When analyzing the clinical data of this trial, Kerstjens *et al.* found that when A was compared to C the improvement in lung function was more marked in patients who were younger, allergic, more hyper-responsive and non-smoking.¹⁶ Therefore we enquired whether the cost difference between the treatments also depended on the type of patient. When performing subgroup analysis of log-cost differences we found a significant interaction between treatment and smoking, treatment and hyper-responsiveness, and treatment and reversibility. When A was compared to C, the reduction in costs due to A was more marked in patients who did not smoke, who were more hyperresponsive and whose airway obstruction was more reversible.

In order to avoid a large increase in the number of explanatory variables in relation to the number of observations we only included the significant interaction terms into the regression analysis. The results of including these interactions as explanatory variables in the regression analysis are shown in the second column of table 4. It can be seen that the cost reducing effect of treatment A is mainly obtained through the significant interaction with smoking and hyper-responsiveness. Besides patients with a better lung function and patients who did not use inhaled corticosteroids prior to the start of the trial, also patients who are less hyper-responsive have lower health care costs than patients who are more hyperresponsive. The largest reduction in costs due to treatment A was found in patients with greater hyper-responsiveness and non-smoking patients. By including the interaction terms, the goodness of fit of the regression model is improved. This improvement is illustrated by a nearly twofold increase of the adjusted R^2 and a decrease of the mean squared residual from 2.15 to 1.92.

Step 3. Calculating standardized expected costs

The expected value of Z_i , denoted as \hat{Z}_i , can be estimated as:

$$(3) \quad \hat{Z}_i = \hat{\alpha} + X_i \hat{\beta}$$

where $\hat{\alpha}$ is the estimated value for the intercept, the values of X_i are the vectors of explanatory variables and $\hat{\beta}$ is a vector of estimated regression coefficients.

To evaluate the impact of the estimated coefficients on treatment costs, we estimated (for each of the 245 patients) three versions of expected costs: first, on the assumption that everybody had in fact received treatment A, secondly, assuming that everybody had received treatment B and finally assuming that all patients had received treatment C. This procedure standardizes for any differences in the X-characteristics between the three groups.

Step 4. Retransformation

Our aim is not to predict the expected costs by treatment group on the transformed scale but on the untransformed scale (i.e. Dutch Guilders). It may seem easy to retransform the estimated transformed costs into Guilders, by simply taking the exponential function of the estimated mean of the transformed costs. However, this is inappropriate because in that case the estimated mean retransformed costs will be neither unbiased nor consistent unless the transformation was linear.²³ The appropriate equation to calculate the expected value of Y_i for an individual with characteristics X_i is:²³

$$(4) \quad E(Y_i|X_i) = \phi \cdot \exp(\alpha + X_i\beta) - 10$$

where ϕ is a retransformation factor that can be estimated by:

$$(5) \quad \hat{\phi} = \frac{1}{N} \sum_{i=1}^n \exp(\hat{\epsilon}_i)$$

where $\hat{\epsilon}_i$ is defined as $Z_i - \hat{\alpha} - X_i\hat{\beta}$. This nonparametric retransformation factor is called a smearing estimator.²⁸ When retransforming estimated log-costs the use of this smearing estimator, $\hat{\phi}$, has been found to outperform the use of a parametric retransformation factor.²⁸ Based on our data, the smearing estimator was 2.860 in case of the regression without interaction terms and 2.513 in case of the regression with interaction terms.

Calculating the corresponding variance of the resulting average expected cost figures for the treatment groups becomes slightly more complicated. The variances can be approximated by means of the so-called delta method.²⁹

$$(6) \quad \text{var}(f(X\beta)) \approx \left[\frac{\partial f}{\partial \beta} \right]' \text{covar}(\beta) \cdot \left[\frac{\partial f}{\partial \beta} \right]$$

which entails in this case:

$$(7) \quad \text{var}(\bar{Y}) = [\bar{Y}]^2 \cdot [\bar{x}]' \cdot \text{covar}(\hat{\beta}) \cdot \bar{x}$$

with $\text{covar}(\hat{\beta})$ being the variance-covariance matrix of the regression coefficients.

Results of the retransformation process

The estimates of the cost differences between the treatment groups that result from the proposed retransformation process are given in the first two rows of table 5. Based on the regression analysis without interaction terms, the savings in retransformed costs due to treatment A are estimated to be 370 NLG (95% CI 97; 643) or 49% of the costs of treatment C. Although it increases the precision of the estimates, taking into account the interactions between treatment and patient-characteristics turns out to have very little influence on the estimated savings. When accounting for interactions, the estimated cost difference between treatment A and C are 366 NLG (95% CI 123 ; 610). Since treatment A was 798 NLG more expensive than treatment C (See the section 'The asthma/COPD trial'), it can be concluded that roughly half of the additional costs of treatment A are compensated for by a reduction in the use of other health care services. Whether or not this is worthwhile depends on the effects of the pharmacotherapy on clinical parameters. The large and important improvements in lung function, hyperresponsiveness and symptoms¹⁶, seem to easily justify this small, but net increase in overall health care costs due to treatment A. The difference between A and C in untransformed costs was 962 NLG (95% CI -81 ; 2004). Thus, the retransformed costs show a smaller difference between the treatment groups than do the untransformed costs per patient-year. The estimated difference in costs between treatments A and C is smaller, but the confidence intervals show that this estimation is much more precise. The regression analysis has shown that this difference in retransformed costs is significant, whereas the difference in untransformed costs was too unreliable as a consequence of the variability in these costs. Furthermore, given the clinical improvements, the difference in retransformed cost is judged to be economically relevant.

Table 5. Comparing the estimated mean (standard error) differences in retransformed and untransformed costs (NLG) per patient-year

	Mean	95% CI
<i>Retransformed costs without interaction</i>		
A - B	457.81 (149.55)	164.70 ; 750.92
A - C	370.38 (139.29)	97.37 ; 643.39
B - C	87.43 (86.51)	-278.13 ; 452.99
<i>Retransformed costs with interaction</i>		
A - B	487.11 (138.80)	215.05 ; 759.16
A - C	366.43 (124.10)	123.20 ; 609.67
B - C	120.67 (171.82)	-216.09 ; 457.44
<i>Untransformed costs</i>		
A - B	690.27 (299.13)	103.98 ; 1276.56
A - C	961.71 (531.98)	-80.98 ; 2004.40
B - C	271.44 (606.78)	-917.85 ; 1460.73

8.7 Discussion

In this paper an approach is suggested to analyze, in the context of economic appraisal alongside a clinical trial, highly skewed cost distributions containing missing data due to subjects dropping out selectively. Frequently, premature withdrawal is associated with a deteriorating clinical condition due to a lack of therapeutic effect, which in turn is associated with higher health care utilization. Since withdrawal was clearly not random across interventions, withdrawal rates may be seen as one of the main outcomes of the treatments. Recognition of this fact, does not imply that the observations prior to withdrawal cannot be included in the analysis. Inevitably, excluding all the cost observations of patients who drop out or assuming that the missing data are randomly distributed, as in the cumulative cost approach, will bias the estimation of costs in the treatment groups and cost differences between the treatment groups. Serious underestimation is avoided when using the patient-year approach. This approach implicitly assumes that the increased use of health care services observed prior to withdrawal will sustain if patients keep on taking the same initial drugs even though their condition deteriorates. Undoubtedly, a pure intention to treat analysis would have been the first best option. However, given the impossibility of such an analysis when patients are lost to follow-up, the patient-year approach is a second best alternative.

The predetermined number of patients needed in this trial was based on the expected size of the difference in lung function parameters, which were the primary outcome measures. There had been no prior discussion of the minimum difference in costs that would be judged important enough to be detected. Moreover, it would have been hard to give any indication of the cost differences that were to be expected, since this study is one of the first pharmaco-economic studies in asthma and COPD. In our study we found what is quite commonly observed in economic evaluations: only a few hospitalizations or other high cost events which cause the variance in cost data to become very large. As a result of such large variance, the sample size which may well be large enough to detect differences in clinical outcomes, may not be sufficient to demonstrate an economic difference at customary levels of statistical significance. Our trial lacked the power to demonstrate that the observed untransformed cost differences were significant. To demonstrate significance at the 5 percent level with a power of 80% in conventional parametric analysis, the number of patients in our trial would have had to be at least five times higher. As Drummond and O'Brien point out, *"a larger trial would have required continuation of random allocation beyond the point at which clinicians would have considered one treatment to be superior to the other"*.¹⁴ Given the clinical superiority of treatment A, resulting in an earlier termination of the trial than originally planned, this would undoubtedly have been judged unethical. Under such circumstances it may be appropriate to accept lower precision in the estimate of cost-differences.¹⁴ Moreover, since skewness seems to be inextricably bound up with the distribution of health care cost, an approach as the one suggested here will

also be needed in the future. The problem of skewness is not solved by excluding the high cost events from the analysis, especially not when the number and size of high cost events are not equally distributed across treatment groups.

In the approach we have presented in this paper we propose to keep as close as possible to the observed cost data, but at the same time accepting its various limitations and gaps. We suggest that high cost events should not be ignored, but the analysis rather adjusted so that the impact of these events is reduced. This can be done by performing a logarithmic transformation of the cost data followed by standardization and a correct retransformation by using a smearing estimator. Such an approach can only be applied to economic data if they are observable on a patient-specific basis. For example, in our trial the costs of the treatment drugs could only be presented as point estimates (see section 2), because the actual drug use as compared to the prescribed dose was not measured. When patient-specific data are not available, sensitivity analysis remains a well-established, appropriate tool for assessing uncertainty. But as more economic evaluations are undertaken alongside prospective clinical trials, the opportunities for systematically measuring economic data on an individual basis grow. We have argued in this paper that such longitudinal economic data should be subjected as far as possible to the same standards of significance testing as is considered acceptable for the analysis of clinical data. When we get a feeling for the magnitude of cost differences that are judged economically important and can be shown to be statistically significant, we may be able determine the appropriate sample size for future economic evaluations. However we have also recognized that this sample size is always likely to be larger than the sample size needed for the clinical evaluation. Thus, for ethical reason, second best analyses will remain necessary in the future.

8.8 Acknowledgements

The authors are grateful to the Dutch Chronic Non-Specific Lung Disease Study Group for providing data. The study was supported by the Dutch Asthma Foundation and Glaxo.

8.9 References

1. Backhouse, M.E., Backhouse, R.J., Edey, S.A. Economic Evaluation Bibliography. *Health Economics*, 1992; **1 (Supplement)**: 1-236
2. Elixhauser, A., Luce, B.R., Taylor W.R., Reblando J. Health Care CBA/CEA: An update on the growth and composition of the literature. *Medical Care*, 1993; **31 (Supplement)**: 1-11
3. Adams, M.E., McCall, N.R., Gray, D.T., Orza, M.J., Chalmers, T.C. Economic Analysis in Randomized Control Trials. *Medical Care*, 1992; **30**: 231-243
4. Weinstein, M.C., Economic Assessment of Medical Practices and Technologies. *Medical Decision Making*, 1981; **1**: 309-330
5. Drummond, M.F., Stoddart, G.L., Torrance, G.W. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press, Oxford, 1987
6. Guyatt, G., Drummond, M., Feeny D. *et al.* Guidelines for the clinical and economic evaluation of health care technologies. *Social Science and Medicine*, 1986; **22**: 393-408
7. Drummond, M.F., Stoddart, G.L. Economic analysis and clinical trials. *Controlled Clinical Trials*, 1984; **5**: 115-128
8. Luce, B.R., Elixhauser, A. Estimating costs in the economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care*, 1990; **6**: 57-75
9. Drummond, M.F., Davies, L. Economic Analysis Alongside Clinical Trials. Revisiting the Methodological Issues. *International Journal of Technology Assessment in Health Care*, 1991; **7**: 561-573
10. Drummond, M., Brandt, A., Luce, B., Rovira, J., Standardizing methodologies for economic evaluation in health care. Practice, problems and potential. *International Journal of Technology Assessment in Health Care*, 1993; **9**: 26-36
11. Gerard, K. Cost-utility in practice: A policy maker's guide to the state of the art. *Health Policy*, 1992; **21**: 249-279
12. Rutten-van Mülken, M.P.M.H., Van Doorslaer, E.K.A., Rutten, F.F.H. Economic Appraisal of asthma and COPD Care: a literature review 1980-1991. *Social Science and Medicine*, 1992; **35**: 161-175
13. O'Brien, B.J., Drummond, M.F., Labelle, R.J. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. McMaster University Centre for Health Economics and Policy Analysis Working Paper #92-23, November 1992
14. Drummond, M., O'Brien, B. Clinical importance, statistical significance and the assessment of economic and quality of life outcomes. *Health Economics*, 1993; **2**: 205-212
15. Brand, P.L.P., Kerstjens, H.A.M., Postma, D.S. *et al.* Long-term multicentre trial in chronic non-specific lung disease. Methodology and baseline assessment in adult patients. *European Respiratory Journal*, 1992; **5**: 21-31
16. Kerstjens, H.A.M., Brand, P.L.P., Hughes, M.D. *et al.* A comparison of bronchodilator therapy with or without inhaled corticosteroid therapy in obstructive airways disease. *The New England Journal of Medicine*, 1992; **327**: 1413-1419
17. Rutten-van Mülken, M.P.M.H., Van Doorslaer, E.K.A., Jansen, M.C.C., Kerstjens, H.A.M., Rutten, F.F.H. Cost-effectiveness of inhaled corticosteroid and bronchodilator therapy in asthma and COPD. *American Journal of Respiratory and Critical Care Medicine*. (forthcoming)
18. Gillings, D., Koch, G., The application of the principle of intention-to-treat to the analysis of clinical trials. *Drug Information Journal*, 1991; **25**: 411-424

19. Gould, A.L. A New Approach to the Analysis of Clinical Drug Trials with Withdrawals. *Biometrics*, 1980; **36**: 721-727
20. Van Hout, B., Bonsel, B., Habbema D., Van der Maas, P., De Charro, F., Heart transplantation in the Netherlands. Costs, effects and scenarios. *Journal of Health Economics*, 1993; **12**: 73-93
21. Zwinderman, A.H. Statistical analysis of longitudinal quality of life data with missing measurements. *Quality of Life Research*, 1992; **1**: 219-224
22. Maddala, G.S. *Econometrics*. page 314-317, New York: McGraw Hill, 1979
23. Duan, N., Manning, W.G., Morris, C.N., Newhouse, J.P. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Economic and Business Statistics*, 1983a; **1**: 115-126
24. Woolson, R.F. *Statistical methods for the analysis of biomedical data*. New York: John Wiley & Sons, 1987
25. Pearson, E.S., Hartley, H.O. *Biometrika tables for statisticians*. London: Cambridge University Press, 1966
26. Wonnacott, Th.,H., Wonnacott, R.J. *Regression: a second course in statistics*. New York: John Wiley & Sons, 1981
27. Newhouse, J.P., Manning, W.G. *et al.* The findings of the Rand Health Insurance experiment. A response to Welch *et al.* *Medical Care*, 1987; **25**: 157-179
28. Duan, N. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*, 1983b; **78**: 605-610
29. Bishop, Y.M.M., Fienberg, S.E., Holland, P.W. *Discrete multivariate analysis: Theory and Practice*. London: MIT Press, 1975

9

General discussion and areas for future research

This last chapter presents a general discussion of the main conclusions that were drawn from the collection of studies in this thesis, avoiding as far as possible any repetition of the detailed conclusions presented at the end of each chapter. Based on these conclusions and on the gaps in our knowledge that became apparent in the previous chapters, a number of recommendations for future research are made.

9.1 Specific pharmaco-economic evaluations

Like health care resources, funds for economic evaluations are limited. An economic evaluation can be expensive, so it is not possible or necessary to conduct economic evaluations of all technologies, or even all new technologies. The selection of technologies which have the highest priority for evaluation should be based on explicit criteria such as those recently been proposed and applied by *several research groups or organizations*.¹⁻⁴ In general, it is relevant to perform economic evaluations when technologies are extremely costly per case or when they are likely to be used by a considerable proportion of the population, so that they are very costly in aggregate.⁵ Pharmacotherapies for asthma and COPD meet this second criterion, but our literature review in chapter 2 showed that by the end of 1991 there was virtually no information available on the cost-effectiveness of anti-asthmatic drugs. Fortunately, the need to assess the cost-effectiveness of the competing pharmaceutical treatment plans was and is increasingly recognised by those involved in the field of asthma and COPD.^{6,7} The Dutch Chronic Non-specific Lung Disease study-group was one of the first to adopt the idea of linking economic evaluations to clinical trials of pharmacotherapy for asthma and COPD. These economic evaluations, which were described in chapter 3 and 4, have led to very straightforward conclusions regarding combined maintenance therapy with an inhaled corticosteroid plus a bronchodilator. It was found that the additional health care costs of adding an inhaled corticosteroid to maintenance β_2 -agonist therapy are to a large extent, though not fully, compensated by savings in other health care costs. The small but net increase in direct costs of health care that results from adding an inhaled corticosteroid turns into net savings to society when the economic benefits of increased productivity due to inhaled corticosteroids are taken into account. Given the clinical improvements of combined inhaled corticosteroid plus β_2 -agonist therapy over β_2 -agonist monotherapy in terms of lung function, hyper-responsiveness and symptom-free days, a small increase in health care costs seems justifiable. These conclusions were reached for both adults and children. For adults, it was also shown that the addition of an inhaled anticholinergic to an inhaled β_2 -agonist did not result in additional effects or savings.

If the effects of inhaled corticosteroids were to continue after treatment is stopped in patients who have reached a clinically stable phase of the disease, the cost-effectiveness of inhaled corticosteroids would be even better. However a number of studies, including a follow-up of the children participating in the SGO children study described in chapter 3, indicate that cessation of treatment with inhaled corticosteroids causes problems.^{8,9} The improvements achieved after long-term treatment with inhaled corticosteroids on symptoms, airway calibre and airway responsiveness do not seem to persist after cessation of treatment. Thus, it can be concluded that inhaled corticosteroids do not cure the disease. The best policy for physicians is to search for the minimal daily dose of inhaled corticosteroids that is effective enough to accurately control the disease.⁸

By emphasizing the need to use inhaled corticosteroids, current treatment guidelines contribute to increasing the efficiency of respiratory therapy. At the moment, physicians are advised to treat moderate asthma with a combination of maintenance inhaled corticosteroid therapy and 'on demand' inhaled short-acting β_2 -agonists therapy. Whether this combination is more cost-effective than the use of both inhaled corticosteroids and β_2 -agonists for maintenance treatment -which was found to be the most favourable therapy in our studies- needs to be established in future research.

International consensus on the treatment of childhood asthma is that cromoglycate is the drug of first choice for maintenance treatment in children with moderately severe asthma. Inhaled corticosteroids are not recommended as first choice for fear of the potentially serious side-effects of long-term maintenance treatment with these drugs. However, evidence on the cost-effectiveness of cromoglycate versus inhaled corticosteroids is lacking. Long-term studies comparing these two drugs in children with asthma of various levels of severity are needed.

Given the evidence of their efficacy and cost-effectiveness, inhaled corticosteroids are likely to be the fastest-growing group of anti-asthmatic drugs. A compound growth rate of 8.9% per year has been estimated.¹⁰ New inhaled corticosteroids such as fluticasone have recently been developed. They have the potential to offer a level of efficacy similar to the existing drugs, using only half the dose. If this improves the safety/efficacy profile the use of inhaled corticosteroids may be extended even more, especially in childhood asthma. Again, these potential additional effects must be weighed against the increase in direct costs of treatment.

In Europe and the US, β -agonists can be expected to remain by far the most commonly used and very effective of the bronchodilator drugs. They can be life-saving when used to treat asthma attacks. However, the uncertainties surrounding the possible causal link between the regular use of certain representatives of this category of drugs and adverse effects on the clinical course of asthma and increased risk of near-fatal or fatal asthma need to be resolved.

Developments in the field of β -agonists have concentrated on the introduction of new long-acting bronchodilators such as salmeterol and formoterol, the first of which was recently approved in the UK for paediatric use also. These drugs, with a duration of action of at least 12 hours, have potential advantages over short-acting bronchodilator drugs, especially with regard to compliance and night rest. Evidence of their bronchodilator effects, even after long-term use, has been provided and numerous randomized controlled studies, including the one described in chapter 6, have shown that there is no deterioration in lung function and no increase in exacerbations with regular salmeterol therapy, whether or not patients were receiving corticosteroids. Our study, described in chapter 6, suggested that salmeterol also tends to improve quality of life when compared to salbutamol. Nevertheless, doubts about the place of maintenance treatment with long-acting β -agonists in asthma management continue to be expressed in the literature. Recently, the results of a British randomized controlled nationwide surveillance study of salmeterol, by Castle *et al.*,¹¹ fuelled these concerns.¹²⁻¹⁴ Although they found that the withdrawal rate due to asthma was lower in the group given salmeterol, it was also found that the death rate from asthma in the group given salmeterol was three times that in the group given salbutamol. However the trial *had insufficient statistical power to detect even a fourfold increase in deaths.*

Various treatment guidelines for adult asthmatics emphasize the appropriate use of these long-acting bronchodilators. They should only be used in combination with inhaled steroids. Whether it is better to combine an inhaled corticosteroid with a long-acting β -agonist, rather than with a short-acting β -agonist on demand, and which patient groups benefit most from these two therapies, still has to be established in future long-term research. If the long-acting β -agonist turns out to be better, it might be useful to develop a standard combination of a β -agonist and a corticosteroid together in one inhaler.¹⁵ This could greatly improve on the compliance rate for inhaled corticosteroids, which is generally low because of the lack of an immediate bronchodilator effect.

9.2 Economic evaluation of screening

We should be careful not to conduct economic evaluations of pharmaceutical interventions only. There is a need for a better understanding of the relative cost-effectiveness of various asthma and COPD management strategies, including the less intensive ones. Most of the literature on asthma and COPD care addresses the effectiveness of various treatment and control strategies. Relatively little attention has been given to primary and secondary prevention strategies. For example, an intensified focus on the elimination of smoking is likely to be a very cost-effective intervention. It may also be possible to decrease the lifetime impact of asthma and COPD by early detection of patients at risk and treatment at an earlier stage in the

course of the disease. The majority of patients in the early stages of the disease are treated by general practitioners, who therefore have the greatest potential to alter the natural course of the disease and make long-lasting improvements in disease morbidity. However, it is difficult for the general practitioner to identify the patients at risk. Facilities for diagnosis should be available. The question is whether the currently available diagnostic 'instruments', such as anamnesis, respiratory symptoms, physical examination of the chest and peak-flow meters are sufficient to detect patients at a very early stage of the disease. Perhaps a supplementary spirometer is necessary to measure and monitor the disease process in terms of decline in FEV₁ or related measures. Monitoring may raise both the physician's and patient's awareness of the problems, and could stimulate them to search for environmental factors or patterns and factors in the patient's behaviour which affect their symptoms.

Early detection of patients at risk, followed by monitoring, is only useful if an effective intervention is available that reduces or delays subsequent morbidity. The most effective therapy currently available is the combination of inhaled corticosteroids and β -agonists. However, relatively little is known about the effects of this combination therapy in very early stages of the diseases.

Extending the diagnostic facilities calls for considerable investments in technology for the general practitioner. These costs of case-finding (both screening and monitoring) should be weighed against the potential effects and in relation to total health care expenditure. A potential disadvantage of early detection is that it may cause medicalization and the unnecessary use of health care resources. We need to be careful not to create unnecessarily costly solutions for the majority of patients with asthma or COPD. On the other hand the concept of early detection, monitoring and intervention may produce returns analogous to the economic concept of return on a capital investment.¹⁶ Investments in better diagnostic facilities in general practice may yield a flow of health benefits to the patient, which in the long run could reduce the use of health care resources. The return on the initial investment is determined by calculating the present value of these future benefits, using the principle of discounting. This however calls for additional research into the lifetime economic consequences of the illness.

Thus far, little is known about the cost-effectiveness of screening in asthma and COPD. Therefore we are currently conducting a long-term randomized controlled study of the cost-effectiveness of a prevention strategy consisting of screening, monitoring and early intervention with inhaled corticosteroids, for asthma and COPD (the DIMCA policy). This strategy is being compared to current medical practice, in which patients are only identified and treated if they consult the general practitioner frequently with respiratory complaints. This cost-effectiveness study, which is being conducted in cooperation with the Department of General Practice in Nijmegen, is limited to patients aged 25 to 70 years. Since there is increasing evidence that chronic obstructive airways disease in adults begins in

childhood, an evaluation of asthma screening of children may be an option for future research.¹⁷

9.3 Population-based studies

Because economic evaluations need to be based on good clinical evidence, we conducted our cost-effectiveness studies alongside randomised controlled clinical trials. However, this raises a number of difficulties.¹⁸ The patients participating in drug trials may not be representative, especially due to the absence of women of child-bearing age, minorities and elderly patients with comorbidities who are receiving multiple medications not related to asthma.^{6,7} One strategy to gain additional information about these categories is through improved post-marketing surveillance of medications.

Another factor that may cause a discrepancy between cost-effectiveness determined in a clinical trial and cost-effectiveness in daily clinical routine is the expansion of the indications. Once a drug is on the market, the indications for its use may be expanded beyond the indications accepted by the Dutch Medicines Evaluation Board. The cost-effectiveness of, for example, an anti-asthmatic drug given to patients in the advanced stage of a disease may differ from the cost-effectiveness of that same drug when given to patients in the early stage of the disease. This was also demonstrated by the statistical analysis in chapter 4, which showed that inhaled corticosteroids were more beneficial in younger patients with asthma than in older patients with COPD.

The difference between effectiveness in use in the community and effectiveness in controlled clinical studies may also relate to the compliance of patients and physicians and the coverage of patients medically eligible for a particular treatment.¹⁹ The latter is a particularly relevant issue, given the under-diagnosis and under-treatment of asthma and COPD. Furthermore, when patients buy drugs but do not take them, or when physicians do not explain inhalation techniques appropriately, the effectiveness of the drugs decreases while the costs remain the same. In general, long-term studies such as those described in chapters 3, 4 and 5 are more likely to reflect compliance in real world situations than the short-term study described in chapter 6.

Finally, the time horizon of the economic evaluations performed alongside clinical trials may be too short to measure total cost savings due to long-term beneficial effects or the costs generated by the treatment of side-effects occurring only after prolonged treatment. To take these factors into account, population-based studies are needed that model chronic care management programs and the cost-effectiveness of continuity and discontinuity of these programs. To build these models, studies of the natural history of asthma and COPD and life-time calculations of the costs of medical consumption will be needed.

9.4 Further development of quality of life measures

Our studies showed that lung function measures and symptom scores are poor measures of the quality of life of asthma and COPD patients. Thus, there is certainly a place for quality of life as an outcome measure in clinical trials. However, with respect to asthma, future quality of life instruments should somehow take into account the fluctuations in the severity of this disease. For many patients, asthma is characterized by a relatively low impact on their well-being, except during acute exacerbations which range in severity from mild to life-threatening. This makes quality of life measurement in asthma more difficult than in other, non-fluctuating chronic diseases. Many of the quality of life instruments, including the ones we have used in our evaluation of salmeterol described in chapter 6, do not pay enough attention to the episodic character of asthma. For such instruments to be useful, the timing of the measurement should match the timing of the episodes as close as possible. In future, quality of life instruments should be developed that reflect both the episodic nature of the disease and the ongoing morbidity component that results from having asthma, such as chronic fatigue or low energy. If this can be achieved, quality of life could be used as an outcome measure for continuous monitoring purposes, as well as for clinical and economic evaluation.

Many of the currently available instruments, disease-specific and even generic, are capable of discriminating between different levels of severity of illness. But often they are not able to detect a difference between the changes in clinical outcomes between groups. It is then not clear whether there is no real difference in the effects of the treatment or whether the instruments are not sensitive enough to detect the difference. It may be that, since asthma is usually reasonably well controlled, the beneficial impact on quality of life of adding new interventions to a patient's routine therapy is relatively small and very large samples are required to demonstrate them in a statistical sense.²⁰ The insensitivity of disease-specific measures may result from the strong focus of these measures on the quality of life between exacerbations and their weak focus on the number, duration and severity of quality of life impairment during illness episodes. Although asthma-specific measures often focus on the frequency of asthma symptoms, some fail to address the effect of these symptoms on physical, psychological, social and role functioning during episodes. The insensitivity of some generic instruments such as the NHP or the SIP may result from their strong focus on musculoskeletal problems. Furthermore, it is unclear why many generic instruments contain questions about the severity and frequency of pain and not about the intensity of breathlessness or other complaints. As a consequence, asthma and even COPD patients' scores on these generic instruments concentrate on the less impaired end of the scale, leaving little room to detect improvement.

Similar problems were encountered in the quality of life measurements undertaken by Kaptein and others as part of the two SGO trials.²¹ This explains why we could not report quality of life outcomes in our cost-effectiveness studies in

chapter 3 and 4. Quality of life was assessed using the instruments available in Dutch at that time and which had been used in asthma or COPD studies. These were 1) the anxiety, depression and sleep disorders subscales of the Symptom Check List - 90 items,²² 2) Optimism and Stigma subscales of the Respiratory Illness Opinion Survey,²³ and 3) 11 items from the Activity of Daily Living scale.²⁴ Measurements were scheduled at baseline and annual follow-ups. At entry into the study, quality of life was found to be mildly impaired and not significantly different from a healthy reference population. Since many patients were lost to follow-up during the first year of treatment from the two groups receiving bronchodilator drugs only, but not from the group receiving a corticosteroid, the first follow-up measurement could only measure the quality of life of patients who were clinically stable enough to remain on bronchodilators only. Thus it was not possible to draw any conclusions about differences in the quality of life between the groups. However, even without this selective treatment-related drop-out, it is very unlikely that these instruments would have been able to detect any differences in quality of life because, at baseline, the patients already scored on the healthy end of the scales.

There are other reasons for not detecting change in quality of life. Factors such as a patient's attitude, expectations, emotions and ability to adapt may complicate the relation between changes in clinical outcomes measures and changes in perceived quality of life. It is important to know how patients' attitudes toward their illness can change over time. We need to know more about the impact of the patient's age at onset, duration of illness and major life events on the patient's attitude toward the illness. Future research should address the reaction of patients to their illness and the way this illness behaviour influences medical outcomes.

Finally, since the prevalence of asthma among children is high, instruments to measure children's health status have to be developed. These instruments should focus especially on activities at school, as children spend a large part of the day there. The development of health status measures for children lags far behind that for adults. To a large extent this is a consequence of doubts about the capability of children to respond to health status questions. For very young children, parents are considered the most valid source of information. However, older children may be perfectly capable of providing the information needed. The validity, reliability and responsiveness of these new instruments should be assessed in an ongoing process in different settings, age groups and levels of disease severity.

9.5 Linking utility measures and generic quality of life measures

Some consensus has emerged that asthma and COPD research should include both disease-specific and generic quality of life measures. The principal advantage of the first type is their higher responsiveness to changes in the dimensions of most interest, and the principal advantage of the latter is the comparability of the results across different diseases. However, in addition to a comprehensive description of the important dimensions of quality of life using both disease-specific and generic instruments, we also need to be able to weight this dimension-specific information and integrate it into an overall score. Such a score is obtained through utility measurement and is needed for use in cost-utility studies. Currently available instruments are strong in either describing health states or weighting health states, but not in both.²⁵ A utility measurement instrument such as the Maastricht Utility Measurement Questionnaire used in chapter 6 and 7 is strong in weighting but does not describe health states very precisely. Most generic quality of life instruments give a comprehensive description of a health state but do not assign values to the overall health states. Some generic instruments, such as the NHP and the SIP, use a weighting system when calculating the overall score. However, using weights assigned to individual items in order to calculate the overall score is fundamentally different from asking respondents to assign an overall weight to a health state. Efforts to combine precise domain measurement with satisfactory methods of obtaining a summary score could be useful,²⁵ particularly to get a better understanding of the meaning of utility scores. Therefore, an appropriate approach is needed for "translating" detailed dimension-specific descriptions into the less detailed dimensions underlying utility measurement instruments.

9.6 Validity of the utility measures

During a workshop on Asthma Outcome Measures for Research Studies held in the US in 1992, Richards and Hamstreet recommended postponing the use of utility measurements in asthma research until the many methodological issues are resolved.³¹ They point out that any inadequacies in the utility measurement procedures are likely to exacerbate the negative view of asthma treatment as requiring life-long medication, with many of the medications being expensive. I doubt whether this is true, especially when asthma patients themselves are asked to assign values to the health states. Moreover, solutions to these methodological issues of utility measurement in general and in asthma and COPD in particular, can be found only through more research.

As far as we know, the study described in chapter 6 is the first to use patient utility measurement in evaluations of asthma treatment. A disadvantage of the methods used was that they were not adjusted for application in evaluating asthma

treatments. As said before, the health status of patients with asthma is generally characterized by relatively normal levels of health with periods of acute exacerbations that vary in severity. However, the utility values do not reflect these acute episodes, because they apply to the general level of functioning during the 2 weeks preceding the measurement. This may partly explain the relatively low sensitivity to change. It seems possible to obtain patient preferences about acute asthma episodes of varying severity and then to apply these values to estimate the effect of different treatments on the health status of patients with asthma.²⁶ To achieve this, health state descriptions should be developed that would reflect patient functioning during asthma attacks of various severities and durations. Utility values for the health states during these attacks could be obtained from groups of asthma patients who would assign values to these descriptions. By weighting the duration and frequency of these acute events according to their utility values, it would be possible to develop health state indexes that incorporate the frequency and duration of asthmatic episodes.

Although this adjustment to the methods we used for utility measurement might improve responsiveness to change, further research into the validity of the standard gamble, the rating scale and other methods is required even more. With respect to the standard gamble, one could accept the Von Neumann-Morgenstern utility theory as a normative standard, in which case one accepts that a person's risk attitude contributes to the utility of the certainty equivalent. In that case we need to study whether the levels of risk that persons are prepared to take correspond to the levels of risk taken in real medical decision situations.²⁷ However, one might also want to correct, in some fashion, for the influence of risk attitude on utilities elicited via the standard gamble method. In that case we should develop methods to measure risk attitude with respect to health. In both cases we should be aware of the common biases in people's perceptions of risk, as recently described by Redelmeier *et al.*²⁸ With respect to the rating scale, we have to question whether the value obtained is essentially different from the one-question global assessments of health which are frequently used in clinical trials. These validity studies should also produce better insights into the influence of socio-demographic factors on utilities, because current evidence in this area is scarce and conflicting.^{29,30}

9.7 Generalizability of the approach suggested for analyzing skewed cost data

It is increasingly recognised that health care utilization and its costs are important outcome measures in assessing the impact of asthma and COPD management interventions, especially in long-term studies. Several clinical trials have now used these measures. However, these studies were not able to provide

precise estimates of the difference in costs between treatment regimes, because the distribution of costs of medical consumption was very skewed and the studies did not have sufficient power to assess the impact of treatment on costs. In chapter 8 we have proposed an approach to estimating and statistically testing differences in cost outcomes between groups in the presence of heavily skewed distributions and treatment-related missing cost data. It is necessary to establish whether this approach is generally applicable in economic evaluations alongside clinical trials in which costs are measured on a patient-specific basis. Further adjustment of the approach to deal with situations in which a large proportion of the patients has no health care utilization at all may be necessary.

9.8 References

1. Eddy D. Selecting Technologies for Assessment. *Int. J. Technology Assessment in Health Care* 1989; 5: 485-501
2. Phelps CE, Parente ST. Priority Setting in Medical Technology and Medical Practice Assessment. *Medical Care* 1990; 28: 703-723
3. Donaldson M, Sox H. (ed). *Setting priorities for Health Technology Assessment: a Model Process*. Institute of Medicine, National Academy Press, Washington, 1992
4. Ziekenfondsraad. Advies inzake kosten-effectiviteitsevaluatie van bestaande verstrekkingen. No. 597, Ziekenfondsraad, Amstelveen 1993
5. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J* 1992; 4: 473-481
6. Ernst P, Spitzer WO, Suissa S, Cockcroft DW, Buist AS. Beta-agonists and asthma research: an international consultation. *Eur Respir J* 1993; 6: 273-278
7. Asthma care in the US: conference proceedings. Working Group Reports. VIB Research Agenda Development. *Medical Care* 1993; 3: MS111-MS116
8. Dompeling E, Van Grunsven PM, Van Schayck CP, Molema J, Akkermans R, Folgering H, Van Weel M. Treatment of mild asthma and chronic bronchitis with inhaled corticosteroids: is discontinuation of therapy possible? In: Domeling E. *Diagnosis and treatment of patients with progressive airflow obstruction*. Ph.D. thesis, Meditekst, Lelystad, 1993
9. Waalkens HJ, Van Essen-Zandvliet EEM, Hughes MD, Gerritsen J, Duiverman EJ, Knol K, Kerrebijn KF and the Dutch CNSLD Study Group. Cessation of long-term treatment with inhaled corticosteroid (budesonide) in children with asthma results in deterioration. *Am Rev Respir Dis* 1993; 148: 1252-1257
10. Anonymous. Low Potential for new asthma drugs? Marketletter July 13, 1992 page 19
11. Castle W, Fuller R, Hall J, Palmer J. Serevent nationwide surveillance study: comparison of salmeterol with salbutamol in asthmatic patients who require regular bronchodilator treatment. *British Medical Journal* 1993; 306: 1034-1037
12. Bunney R. Study too small to detect increase in deaths. Letter to the editor. *British Medical Journal* 1994; 306: 1610

13. Crompton G. Regular treatment with β agonists remains unevaluated. Letter to the editor. *British Medical Journal* 1994; 306: 1610
14. Sears MR, Taylor DR. Increase in deaths during salmeterol unexplained. Letter to the editor. *British Medical Journal* 1994; 306: 1610-1611
15. Barnes PH. New drugs for asthma. *European Respiratory Journal* 1992; 5: 1126-1136
16. Weiss KB, Sullivan SD. The Economic Costs of Asthma. A Review and Conceptual Model. *PharmacoEconomics* 1993; 4: 14-30
17. Jones A. Screening for asthma in children. *British Journal of General Practice* 1994; 44: 179-183
18. Leidl R. Some factors to consider when using the results of economic evaluation studies at the population level. *International Journal of Technology Assessment in Health Care* 1994 (forthcoming)
19. Tugwell P, Bennet KJ, Sackett DL, Haynes RB. The measurement iterative loop: a framework for the critical appraisal of need, benefits and costs of health interventions. *Journal of Chronic Diseases* 1985; 38: 339-351
20. Richards JM, Hemstreet MP. Measures of Life Quality, Role Performance and Functional Status in Asthma Research. *Am J Respir Crit Care Med* 1994; 149: S31-S39
21. Kaptein AA, Brand PLP, Dekker FW, Kerstjens HAM, Postma DS, Sluiter HJ, and the Dutch CNSLD study group. *Eur Respir J* 1993; 6: 1479-1484
22. Arrindell WA, Ettema JHM. Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de Symptom Checklist (SCL-90); gegevens gebaseerd op een fobische en een "normale" populatie. *Ned Tijdschr Psychol* 1981; 36: 77-108
23. Kaptein AA. Illness behavior of patients with asthma. Meppel, Krips 1982 (Ph.D. thesis)
24. Schrier AC, Dekker FW, Kaptein AA, Dijkman JH. Quality-of-life in elderly patients with chronic nonspecific lung disease seen in family practice. *Chest* 1990; 98: 894-899
25. Hays RD, Stewart AL, Shervourne CD, Marshall GN. The 'states versus weights' dilemma in quality of life measurement. *Quality of Life Research* 1993; 2: 167-168
26. Rothman ML, Revicki DA. VD. Issues in the Measurement of Health Status in Asthma Research. Asthma care in the US: conference proceedings. *Medical Care* 1993; 3: MS82-MS96
27. Buckingham K. Risks in utility assessment and risks of medical interventions. Letter to the editor. *Medical Decision Making* 1993; 2: 167-168
28. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. *Journal of the American Medical Association* 1993; 270: 72-76
29. Katz JN, Phillips CB, Fossel AH, Liang MH. Stability and Responsiveness of Utility Measures. *Medical Care* 1994; 32: 183-188
30. Bakker C, Rutten-van Mólken M, Hidding A, Van Doorslaer E, Bennett K, Van der Linden S. Patient utilities in ankylosing spondylitis and the association with other outcome measures. *Journal of Rheumatology* 1994; 21: 1298-1304

Samenvatting

In dit proefschrift worden diverse aspecten belicht die samenhangen met de kosten-effectiviteit van behandelingen voor CARA. Het bevat hoofdstukken die ingaan op methodologische aspecten van kwaliteit van leven meting, utiliteitsmeting en kostenrekening. Andere hoofdstukken beschrijven de resultaten van empirische implementatie van de bovengenoemde metingen in klinische trials van geneesmiddelen voor CARA. De grote schaal waarop CARA in Nederland voorkomt en de daarmee verbonden maatschappelijke kosten rechtvaardigen de keuze om deze ziekte in dit proefschrift centraal te stellen. Pharmacotherapie is de voornaamste vorm van therapie voor CARA. Toch werd tot op heden aan de kosten-effectiviteit van de beschikbare geneesmiddelen weinig aandacht besteed. Recente Nederlandse voorstellen om de invloed van doelmatigheidsoverwegingen op beleidsbeslissingen te doen toenemen onderstrepen het belang van methodologische en empirische studies op dit terrein.

In *hoofdstuk 2* worden de resultaten beschreven van een literatuurstudie naar de kosten-effectiviteit van behandelingen voor CARA. Ook wordt ingegaan op de methodologische aspecten van de studies.

Ondanks het feit dat de maatschappelijke kosten van deze ziekte zeer aanzienlijk zijn en ondanks de noodzaak om de verdeling van financiële middelen meer en meer te baseren op doelmatigheidsoverwegingen, werden tussen 1980 en 1991 slechts 20 economische evaluaties van CARA-interventies gepubliceerd. De meeste economische evaluaties die gedaan zijn hadden betrekking op gezondheidsvoorlichting en opvoedings-programma's (GVO-programma's), zowel voor volwassenen als kinderen. Over het algemeen leken deze programma's te leiden tot lagere medische consumptie, een verbeterde attitude ten opzichte van astma, een verbeterde compliance en verbeterde zelfzorg vaardigheden. De resultaten van de evaluaties die betrekking hadden op thuiszorg-programma's leken het meest gunstig wanneer deze programma's opgezet en uitgevoerd werden vanuit een ziekenhuis en gericht waren op de ernstig zieke CARA-patiënten. De studies die de kosten van verschillende toedieningswijzen voor bronchodilatoren en zuurstof bekeken pleitten respectievelijk voor het gebruik van de 'metered dose inhaler' en de 'concentrator'. Er werd geen enkele kosten-effectiviteitsstudie van geneesmiddelen voor CARA gevonden. Dit is zeer opmerkelijk omdat medicamenteuze behandeling toch de voornaamste vorm van behandeling voor CARA is. Economische evaluaties van diagnostiek bleken eveneens volledig te ontbreken.

De kwaliteit van de gevonden economische evaluaties was voor verbetering vatbaar. De resultaten van de diverse studies waren moeilijk vergelijkbaar omdat de gebruikte methoden nogal uiteenliepen.

In *hoofdstuk 3* worden de kosten en effecten vergeleken van twee verschillende pharmacotherapieën voor de behandeling van kinderen met astma in de leeftijd van 7 tot 16 jaar. De eerste behandeling is een combinatietherapie bestaande uit een inhalatie β_2 -agonist en een inhalatie corticosteroid (BA+CS) en de tweede behandeling is een monotherapie die alleen de inhalatie β_2 -agonist omvat (BA+PLacebo). Naast de effecten worden de kosten van de astma-gerelateerde medische consumptie van 116 kinderen bepaald. Voor het berekenen van de kosten werden twee verschillende methoden gebruikt: 1) de cumulatieve kostenbenadering en 2) de patiëntjaar benadering.

BA+CS therapie verbetert de longfunctie en de hyperreactiviteit en resulteert in minder symptomen en minder schoolverzuim. Op jaarbasis is BA+CS therapie echter 480 gulden per patiënt duurder. Daar staat tegenover dat BA+CS tot besparingen op de overige medische consumptie leidt. Met behulp van de, methodologisch meest juiste, patiëntjaar benadering (met regressie op getransformeerde kosten) werden deze besparingen geschat op 43% van de kosten van medische consumptie in de BA+PL groep. De besparingen werden aanzienlijk groter wanneer rekening werd gehouden met de kosten van productiviteitsverlies dat ontstaat wanneer ouders door het schoolverzuim van hun kind hun gebruikelijke dagelijkse activiteiten niet meer kunnen doen.

Geconcludeerd kan worden dat het toevoegen van een inhalatie corticosteroid aan een inhalatie β_2 -agonist tot dusdanig betere resultaten leidt dat de geringe extra kosten die mogelijk ontstaan gerechtvaardigd lijken. Deze combinatietherapie zou zelfs tot netto besparingen kunnen leiden.

In *hoofdstuk 4* worden de resultaten beschreven van een 2.5 jaar durende gerandomiseerde klinische trial waarin 3 verschillende inhalatie pharmacotherapieën met elkaar worden vergeleken: 1) β_2 -agonist + corticosteroid (BA+CS), 2) β_2 -agonist + anticholinergicum (BA+AC) en 3) β_2 -agonist + placebo (BA+PL). Het onderzoek werd uitgevoerd bij 174 volwassen CARA patiënten in de leeftijd van 18-60 jaar met een matige luchtwegobstructie. De voornaamste klinische effectmaten waren FEV₁ (Forced Expiratory Volume in 1 second), hyperreactiviteit uitgedrukt in PC₂₀ (Provocatieve Concentratie histamine die leidt tot 20% daling in FEV₁), symptoomvrije dagen en dagen waarop men niet in staat is de gebruikelijke dagelijkse activiteiten (inclusief betaald werk) uit te voeren. De economische effectmaten waren de kosten van medische consumptie bestaande uit de kosten van de te evalueren geneesmiddelen, 'rescue' therapie, additioneel geneesmiddelengebruik, ziekenhuisopnamen, additionele specialistencontacten en -verrichtingen.

In vergelijking met BA+PL leidde BA+CS tot significante verbeteringen in FEV₁, PC₂₀ en symptoomvrije dagen. Op deze parameters evenals op de economische parameters werden geen significante verschillen gevonden tussen BA+PL en BA+AC. De kosten van BA+CS, BA+AC en BA+PL bedragen respectievelijk 532 US\$, 277 US\$ en 156 US\$ per patiënt per jaar. BA+CS is dus

376 US\$ duurder dan BA+PL. Daar staat tegenover dat BA+CS ten opzichte van BA+PL tot besparingen op de overige medische consumptie leidt die geschat worden op 175 US\$ per patiënt per jaar. Ongeveer de helft van de extra kosten van BA+CS wordt dus gecompenseerd door besparingen. Als de indirecte kosten worden meegenomen tegen een bedrag van 42 US\$ of meer dan zal BA+CS leiden tot netto besparingen.

Geconcludeerd kan worden dat toevoeging van een inhalatie corticosteroid aan een inhalatie β_2 -agonist leidt tot verbeteringen in longfunctie en meer symptoomvrije dagen die de geringe extra kosten lijken te rechtvaardigen. De toevoeging van een inhalatie anticholinergicum is duurder en lijkt niet te leiden tot extra effecten.

In *hoofdstuk 5* worden de effecten vergeleken van het gebruik van inhalatie bronchodilatoren op de longfunctie en de kwaliteit van leven. Het onderzoek betrof een gerandomiseerde klinische trial bij volwassen CARA patiënten met lichte luchtwegobstructie. Twee verschillende therapieën werden met elkaar vergeleken: 1) continu gebruik van inhalatie bronchodilatoren versus 2) alleen gebruik van bronchodilatoren bij klachten (symptomatisch gebruik). Binnen deze 2 groepen kreeg de helft van de patiënten het eerste jaar salbutamol voorgeschreven en het tweede jaar ipratropium bromide terwijl de andere helft deze geneesmiddelen in omgekeerde volgorde kreeg voorgeschreven. Naast longfunctie (FEV_1) werd kwaliteit van leven gemeten met behulp van de NHP en de VOEG.

Van 144 patiënten was een complete follow-up beschikbaar. De kwaliteit van leven van deze patiënten was duidelijk slechter dan in de algemene bevolking. Er werd geen significant verschil gevonden tussen salbutamol en ipratropium bromide, niet wat betreft longfunctie en niet wat betreft kwaliteit van leven. Er werd wel een verschil gevonden tussen continu en symptomatisch gebruik. De achteruitgang in FEV_1 was significant groter bij continu gebruik dan bij symptomatisch gebruik. Dit werd echter niet weerspiegeld in een significante achteruitgang in kwaliteit van leven gemeten met de NHP en de VOEG. Achteruitgang in FEV_1 correleerde niet met veranderingen in de scores op de kwaliteit van leven vragenlijsten. Hiervoor zijn 3 redenen mogelijk. Ten eerste kunnen de patiënten zich snel aanpassen aan een dalende longfunctie, waardoor dit geen direct effect heeft op de *ervaren* kwaliteit van leven. Ten tweede kan continu gebruik van bronchodilators en het daardoor continu onderdrukken van symptomen de verergering van de ziekte die aan de symptomen ten grondslag ligt maskeren, waardoor patiënten zich er simpelweg niet van bewust zijn. Tenslotte kan het zijn dat de NHP en de VOEG niet gevoelig genoeg zijn om de wel optredende veranderingen in kwaliteit van leven te registreren.

In *hoofdstuk 6* worden de resultaten beschreven van een onderzoek naar de psychometrische eigenschappen van 4 verschillende instrumenten voor het meten van kwaliteit van leven. De instrumenten werden vergeleken wat betreft hun

validiteit, betrouwbaarheid en gevoeligheid voor verandering. Dit onderzoek vond plaats in de vorm van een gerandomiseerde dubbelblinde klinische trial waarin 120 patiënten met matig ernstige astma in de leeftijd van 18 tot 70 jaar behandeld werden met ofwel salmeterol, een nieuwe, langwerkende β_2 -agonist, ofwel salbutamol, een reeds langer bestaande, kortwerkende β_2 -agonist. Naast klinische effectparameters werd kwaliteit van leven gemeten met behulp van: 1) Astma Quality of Life Questionnaire (AQLQ), 2) Living with Astma Questionnaire (LWA), 3) Sickness Impact Profile (SIP) en 4) rating scale (RS) en standard gamble (SG) utiliteiten.

Vergeleken met salbutamol leidde salmeterol tot significant grotere verbeteringen in nagenoeg alle klinische effectmaten. Hoewel alle kwaliteit van leven instrumenten deze trend weerspiegelden waren alleen de AQLQ en de RS utiliteiten in staat aan te tonen dat salmeterol tot een significant grotere verbetering in kwaliteit van leven leidde dan salbutamol. De correlatie tussen verandering in longfunctie en verandering in kwaliteit van leven was laag voor alle instrumenten behalve voor de AQLQ. Terwijl de AQLQ goed correleerde met de globale beoordeling van de effectiviteit gegeven door de patiënt vertoonden de andere kwaliteit van leven instrumenten geen significante correlatie met deze beoordeling. Van alle instrumenten correleerde de AQLQ het beste met de symptoomscores. De cross-sectionele correlatie tussen de AQLQ en de LWAQ bedroeg 0.73, terwijl de longitudinale correlatie tussen deze instrumenten slechts 0.29 bedroeg. De SG utiliteiten correleerden over het algemeen het slechtst met de andere effectmaten, inclusief de RS utiliteiten. De 6-weeken test-hertest betrouwbaarheid was relatief hoog voor de AQLQ, de LWAQ en de SIP, maar lager voor de utiliteiten.

Geconcludeerd kan worden dat salmeterol tot een significant grotere verbetering in kwaliteit van leven lijkt te leiden dan salbutamol. Zowel de AQLQ als de LWAQ zijn relatief betrouwbare ziektespecifieke instrumenten, maar in de hier onderzochte setting is de AQLQ gevoeliger voor verandering en meer valide. Van de generieke instrumenten is de SIP het betrouwbaarst, maar de RS het meest gevoelig voor verandering. De SG was minder betrouwbaar en minder valide.

In *hoofdstuk 7* worden diverse methodologische aspecten van utiliteitsmetingen met de rating scale en de standard gamble methode behandeld. Deze methodologische aspecten werden gedestilleerd uit twee gerandomiseerde klinische trials waaraan respectievelijk 85 fibromyalgie patiënten en 144 spondylitis ankylopoëtica patiënten deelnamen. In beide trials vond 1 baseline meting en 2 follow-up metingen plaats waarin patiënten gevraagd werden hun waarderingen uit te spreken voor hun eigen gezondheidstoestand en enkele hypothetische gezondheidstoestanden (licht matig en ernstig ziek).

De standard gamble bleek consistent tot hogere scores te leiden dan de rating scale, zowel voor de eigen gezondheidstoestand als voor de hypothetische gezondheidstoestanden. De 3-maand test-hertest betrouwbaarheid voor de hypothetische toestanden uitgedrukt in intraclass correlatie coëfficiënten varieerde

van 0.24 tot 0.33 voor de rating scale en van 0.43 tot 0.70 voor de standard gamble. Hoewel de reproduceerbaarheid niet hoog was, bleven de groepsgemiddelden in de loop der tijd relatief constant. De gemiddelde standard gamble score lijkt afhankelijk te zijn van de wijze waarop de metingen worden uitgevoerd. Utiliteiten gemeten met zogenaamde 'chained gambles' waren significant hoger dan utiliteiten gemeten met de 'basic reference gamble'. Op het individuele niveau werden inconsistente antwoorden gevonden. Meer dan 70% van deze inconsistente antwoorden viel echter binnen de grenzen van de meetfout, die varieerde van 0.11 tot 0.13 op de standard gamble (0-1 schaal) en van 8 tot 10 op de rating scale (0-100 schaal). Het grote aantal negatieve utiliteiten voor de hypothetische toestand van een ernstig zieke -de toestand die als ankerpunt in de 'chained gambles' was gebruikt- en de waarde van deze negatieve utiliteiten (tot -19) onderstrepen de noodzaak betere methoden te ontwikkelen voor het meten van negatieve utiliteiten en het hanteren van deze utiliteiten in QALY berekeningen.

In *hoofdstuk 8* wordt een methode beschreven waarmee verschillen in kosten tussen behandelingen kunnen worden geschat. Voorwaarde voor deze methode is dat de kosten longitudinaal en op patiëntniveau verzameld worden. Dit is vaak het geval wanneer parallel aan een klinische trial een economische evaluatie wordt uitgevoerd. In die situatie wordt de schatting van de kostenverschillen vaak bemoeilijkt door 2 problemen. Ten eerste kunnen de percentages uitvallers nogal verschillen tussen de groepen ten gevolge van selectieve, vroegtijdige uitval van ernstig zieken uit de groep die de minst effectieve behandeling krijgt. Dit betekent dat vooral de patiënten met de hoogste kosten vroegtijdig de studie verlaten, waardoor er een aanzienlijk aantal missing values ontstaat. Ten tweede is de frequentieverdeling van kosten vaak zeer scheef door enkele ziekenhuisopnamen of andere relatief dure interventies. Om de schattingsmethode te illustreren zijn data gebruikt uit de trial beschreven in hoofdstuk 4.

Om de kosten per patiëntjaar te bepalen werden de kosten van patiënten met een incomplete follow-up geëxtrapoleerd. Als gevolg van de scheve verdeling konden deze kosten per patiëntjaar niet met behulp van de gebruikelijke parametrische statistische methoden worden geanalyseerd. Om een normale verdeling te benaderen, de invloed van extreme waarden te reduceren en vergelijkbare varianties te verkrijgen werden de data logaritmisch getransformeerd. Daarna werd een multiple regressie uitgevoerd met de getransformeerde data als afhankelijke variabele. Met behulp van deze regressie kon gestandaardiseerd worden voor verschillen in patiëntkenmerken. Voor de retransformatie werd de zogenaamde 'smearing estimator' gebruikt. Deze 'transformatie-standaardisatie-retransformatie' procedure leidt tot consistente en efficiënte schattingen van kostenverschillen.

Het laatste hoofdstuk, *hoofdstuk 9*, bevat een algemene discussie van de voornaamste conclusies uit de hiervoor gepresenteerde collectie van studies. Tevens worden een aantal lacunes in de huidige kennis gesignaleerd die leiden tot aanbevelingen voor toekomstig onderzoek.

In dit proefschrift werd o.a. de behandeling bestaande uit continu gebruik van inhalatie corticosteroïden in combinatie met continu gebruik van kortwerkende inhalatie bronchodilatoren onderzocht. Of continu gebruik van inhalatie corticosteroïden niet beter aangevuld kan worden met bronchodilators 'on demand' of met langwerkende bronchodilators is nog een onbeantwoorde vraag. Datzelfde geldt voor de vraag of cromoglycaat -vaak de eerste keus voor onderhoudsbehandeling bij kinderen- kosten-effectiever is dan inhalatie corticosteroïden. Daarnaast worden er nieuwe vormen van inhalatie corticosteroïden ontwikkeld die evaluatie behoeven.

Echter niet alleen economische evaluaties van medicamenteuze behandelingen zijn noodzakelijk. Wellicht dat een groot deel van de levenslange invloed van CARA gereduceerd kan worden door vroegtijdige opsporing van patiënten met een verhoogd risico, zodat zij bij eventuele ontwikkeling van de ziekte ook vroegtijdig behandeld kunnen worden. Een kosten-effectiviteitsstudie van een dergelijke preventieve strategie wordt momenteel uitgevoerd, waarbij zorgvuldig de gevaren van medicalisering en het onnodig creëren van dure interventies in het oog worden gehouden.

In dit afsluitende hoofdstuk wordt ook gewezen op de mogelijke discrepanties tussen kosten-effectiviteit in de routine van de dagelijkse medische praktijk en kosten-effectiviteit gemeten in een klinische studie. Factoren die deze discrepantie veroorzaken zijn bijvoorbeeld de geringe representativiteit van de geselecteerde patiënten in de trials, verschillen in compliance, verschillen in bereikbaarheid van patiënten, uitbreiding van de indicaties waarvoor de geneesmiddelen worden gebruikt etc. Om met al deze factoren rekening te houden zijn modelmatige, populatiegebaseerde studies noodzakelijk.

Op het gebied van kwaliteit van leven bestaat de noodzaak instrumenten te ontwikkelen die niet alleen aandacht besteden aan het chronische karakter van astma maar vooral ook aan het episodische karakter. Dit geldt ook voor de utiliteitsmetingen. Door meer aandacht te besteden aan de fluctuaties in de ernst van de klachten zou wellicht ook de gevoeligheid van de meetinstrumenten voor verandering verbeterd kunnen worden. Verder zal gezocht moeten worden naar mogelijkheden om generieke kwaliteit van leven instrumenten en utiliteitsmetingen aan elkaar te relateren, zodat er niet langer alléén instrumenten zijn die óf goed zijn in het beschrijven van de gezondheidstoestand óf goed zijn in het waarderen, maar in beide. Daarnaast behoeft vooral de validiteit van de utiliteitsmetingen meer onderzoek, evenals de algemene toepasbaarheid van de ontwikkelde methode voor het schatten van kostenverschillen in economische evaluaties gekoppeld aan klinische trials.

Dankwoord

Heel veel mensen hebben aan de totstandkoming van dit proefschrift bijgedragen. Ik ben hen allemaal dank verschuldigd. In het bijzonder natuurlijk de patiënten die aan de onderzoeken hebben deelgenomen. Enkele mensen wil ik met name noemen.

Allereerst prof.dr. F.F.H. Rutten, die ervoor heeft gezorgd dat ik in het Medical Technology Assessment onderzoek ben gerold. Beste Frans, ik herinner me nog goed hoe blij ik was toen je me vroeg om bij het iMTA te komen werken. Jouw enthousiasme voor MTA heb ik overgenomen. Gelijk bij mijn eerste project gooide je me in het diepe en je leerde me ervaren dat ik onderzoek heel leuk vond en dat ik tot meer in staat was dan ik in het begin voor mogelijk hield. Soms moesten we heel wat moeite doen om projecten binnen te halen. Tegenwoordig gaat dat heel wat gemakkelijker. Wat dat betreft heb ik veel van je geleerd. Bedankt voor het feit dat je me altijd bent blijven stimuleren, zelfs na je vertrek uit Maastricht.

Prof.dr. E.K.A. Van Doorslaer heeft als mijn directe begeleider aan de basis gestaan van dit proefschrift. Beste Eddy, toen ik nog studeerde waren de studenten het erover eens dat er maar weinigen waren die beter college konden geven dan jij. Later heb ik ervaren dat dit ook voor gezondheidseconomisch onderzoek geldt. Ik voel me bevoorrecht dat ik met jou heb mogen samenwerken. We zijn goede vrienden geworden en hebben heel wat gediscussieerd en heel wat afgelachen. Jouw humor en verstrooidheid maakten onze samenwerking tot een groot plezier. Jouw ideeën en stimulans zijn zeer bepalend geweest voor de totstandkoming van dit proefschrift. Bedankt voor dit alles. Ik ben blij dat we, ondanks het feit dat we niet aan dezelfde universiteit werken, ook in de toekomst blijven samenwerken.

Jan van Emmerik heeft bij alle onderzoeken in dit proefschrift de computerondersteuning verzorgd en de lay-out van dit proefschrift voor zijn rekening genomen. Beste Jan, bedankt voor het feit dat je op ieder moment van de dag en zelfs in de avonduren en in het weekend beschikbaar was als weer eens een SPSS programma vastliep, als er dia's gemaakt moesten worden of mijn auto niet meer wilde starten. Onvermoeibaar veranderde je alle tabellen, figuren en referentienummers in dit proefschrift als ik weer eens iets had toegevoegd. Jouw flexibele houding en je kennis van zaken heb ik enorm gewaardeerd.

Prof.dr. Sj. Van der Linden en drs. C. Bakker hebben bijgedragen aan het ontwikkelen van de utiliteitsmetingen. Beste Sjef en Carla, bedankt voor de leuke besprekingen en discussies die we samen hebben gevoerd over de zinvolheid van utiliteitsmetingen, de moeilijkheid van de methoden en de vaak verrassende antwoorden van patiënten. De contacten van jou, Sjef, met McMaster University in Canada hebben Carla en mij in de gelegenheid gesteld de daar gebruikte methoden

van utiliteitsmeting ook in Nederland toe te passen. Onze reisjes naar Hamilton waren een waar genoegen.

Drs. C.C.M. Jansen en drs. L. Heurman waren verantwoordelijk voor het afnemen van de interviews met patiënten in het salmeterol onderzoek. Beste Linda en Margreet, bedankt voor al die keren dat jullie naar het De Wever ziekenhuis in Heerlen zijn gegaan om daar de vragenlijsten te laten invullen en interviews af te nemen. Een interview over kwaliteit-van-leven kan nog zo gestructureerd zijn opgezet, het nodigt mensen uit om hun hele levensverhaal te vertellen. Jullie waren altijd bereid om te luisteren. Margreet, jou wil ik ook graag bedanken voor je bijdrage aan de twee kosten-effectiviteitsstudies van inhalatiecorticosteroiden. Het was vaak één grote puzzel, die we uiteindelijk samen hebben opgelost. Linda, jou bedank ik ook voor al die keren dat je uren in de bibliotheek hebt gestaan om voor mij artikelen te kopiëren.

Ik bedank alle leden van de vakgroep Economie van de Gezondheidszorg voor de prettige sfeer waarin ik de afgelopen jaren heb kunnen werken en de belangstelling die er altijd werd getoond voor de vorderingen van mijn proefschrift. In het bijzonder bedank ik prof.dr. R. Leidl voor de gelegenheid die hij mij heeft gegeven om dit proefschrift te schrijven. Onze vakgroepsecretaresse, Maria Kalivas, heeft mij daarbij altijd ten dienste gestaan. Mariëlle Goossens en Guido van den Boom begeleid ik bij hun MTA onderzoeken van respectievelijk behandelingen voor chronische pijnpatiënten en screening, monitoring en vroegtijdige behandeling van CARA. Tijdens de laatste fase van mijn proefschrift had ik te weinig tijd voor jullie. Daar komt nu verandering in. Ik bedank Silvia, Mariëlle, Grant en Rob (ons sportclubje) omdat ze mij steeds wisten over te halen om tijdens de lunchpauze toch mee naar de conditietraining te gaan.

Verder dank ik alle leden van de S.G.O. CARA groep bestaande uit hoogleraren longziekten, longartsen, longfunctie-deskundigen, longfunctie- en allergologie-assistenten en CARA-onderzoekers van 6 academische ziekenhuizen in Nederland. De programmacommissie van deze onderzoeksgroep bestaande uit prof.dr. K.F. Kerrebijn, prof.dr. Ph.H. Quanjer en wijlen prof.dr. H.J. Sluiter, ben ik zeer erkentelijk voor het feit dat ze mij de kans hebben gegeven om aan de twee klinische trials die zij uitvoerden een kosten-effectiviteitsstudie te koppelen. Voor de wereld der longziekten een volledig nieuwe maar achteraf succesvol bevonden ervaring. In het bijzonder wil ik dr. H.A.M. Kerstjens en dr. E.E.M. van Essen-Zandvliet bedanken. Beste Huib en Liesbeth, jullie waren altijd bereid mij de benodigde gegevens en bijbehorende toelichting te verstrekken. Wat de medische aspecten van CARA betreft heb ik veel van jullie geleerd. Huib, jouw inbreng bij de interpretatie van de kostenverschillen tussen de 3 behandelingen en de relatie van deze verschillen tot de effecten heb ik bijzonder waardevol en plezierig gevonden. Dezelfde blijken van waardering wil ik uiten richting prof.dr. D.S. Postma. Beste

Dirkje, jouw vermogen om mee te denken over de economische resultaten heb ik zeer bewonderd. Op jouw karakteristieke vriendelijke wijze heb je mij tijdens vergaderingen en congressen wegwijds gemaakt binnen de S.G.O. CARA groep.

De CARA onderzoeksgroep van de Katholieke Universiteit Nijmegen in samenwerking met het Medisch Centrum Dekkerswald en het Radboud Ziekenhuis Nijmegen, in het bijzonder prof.dr. C. van Weel, dr. C.P. van Schayck en prof.dr. H. Folgering dank ik voor de plezierige samenwerking gedurende de afgelopen jaren. Lilian Biermans en Agnes Raaijmakers hebben de interviews gedaan voor een pilot onderzoek naar de haalbaarheid van utiliteitsmetingen bij CARA patiënten.

Dr. F.P.V. Maesen, hoofd van de afdeling longziekten in Heerlen ben ik zeer erkentelijk voor de mogelijkheid die hij mij heeft geboden om op zijn afdeling uitgebreid onderzoek te doen naar de kwaliteit-van-leven effecten van salmeterol. De heer Sj. Smeets dank ik voor de organisatie van deze salmeterol trial en drs. M. Costongs en drs. F. Custers dank ik voor de uitvoering van het medisch gedeelte van deze trial. Onze samenwerking was zeer succesvol.

Alle leden van de Werkgroep Onderzoek Gezondheidstoestandmeting (WOG), onder voorzitterschap van achtereenvolgens prof.dr. P.J. van der Maas en prof.dr. J. Passchier dank ik voor de plezierige workshops die we de afgelopen jaren regelmatig hebben gehad. Dat ik als een van de weinige niet-gepromoveerden deel kon uitmaken van deze werkgroep was voor mij zeer waardevol. Ik denk dat we veel van elkaar hebben geleerd en toch een klein stukje standaardisatie binnen kwaliteit-van-leven onderzoek hebben weten te bereiken. Ik ben blij dat we ermee doorgaan. Drs. M-L. Essink-Bot dank ik voor het feit dat ze de WOG-workshops zo goed organiseerde en alle leden van de werkgroep zo fantastisch op de hoogte hield van nieuwe ontwikkelingen. Marie-Louise, wij hebben samen het rapport getiteld "Het meten van de gezondheidstoestand" geschreven. Ik vond onze samenwerking buitengewoon nuttig, maar vooral ook gezellig.

De firma Glaxo, het Nederlands Astma Fonds en het Stimuleringsprogramma Gezondheidszorg Onderzoek wil ik bedanken voor de financiële ondersteuning van de onderzoeken in dit proefschrift. Dr. H. Tamminga en drs. A. Bommer van de firma Glaxo dank ik voor hun waardevolle adviezen en ondersteuning bij de uitvoering van de onderzoeken. Besprekingen met jullie verliepen in een bijzonder plezierige sfeer.

Sen McGlinn wil ik hartelijk danken voor de deskundige manier waarop mijn Engelstalige teksten werden gecorrigeerd en dr. H. Schouten wil ik danken voor zijn deskundige statistische adviezen.

Mijn ouders dank ik dat ze me in de gelegenheid stelden om te studeren. Jullie interesse en die van Iris in mijn onderzoek en mijn proefschrift was heel fijn.

Wil, er bestaan niet genoeg woorden om jouw voor alles te bedanken. Ik probeer het dus maar niet. Jij hebt mij altijd gestimuleerd, je was mijn steun en toeverlaat. Als ik thuiskwam stond er vaak een heerlijke maaltijd voor me klaar en mopperen als ik weer eens dacht te moeten overwerken ligt niet in je aard. Zonder jou zouden de afgelopen jaren nooit zo leuk zijn geweest.

Curriculum Vitae

Maureen van Mólken werd geboren op 6 juli 1966 te Sittard. In 1984 behaalde zij het eindexamen atheneum aan de Scholengemeenschap Groenewald te Stein. In dat jaar startte zij de studie gezondheidswetenschappen aan de Rijksuniversiteit Limburg waar zij in 1988 afstudeerde in de richting Beleid en Beheer der Gezondheidszorgvoorzieningen.

Van 1988 tot 1991 was zij werkzaam als onderzoeker bij het Instituut voor Medische Technology Assessment (directeur: prof.dr. F.F.H. Rutten) aan de Rijksuniversiteit Limburg. Hier legde zij de basis voor het onderhavige proefschrift. Sedert 1991 werkt zij als universitair docent bij de vakgroep Economie van de Gezondheidszorg (hoofd: prof.dr. R. Leidl) aan de Faculteit der Gezondheids-wetenschappen van dezelfde universiteit.

Zij is betrokken bij vele wetenschappelijke onderzoeken en het onderwijs op het gebied van de medical technology assessment. Zij is lid van de Werkgroep Onderzoek Gezondheidstoestandmeting (WOG) een nationale werkgroep die zich bezig houdt met standaardisatie van methoden voor gezondheidstoestand-metingen binnen evaluatie onderzoek. Haar speciale aandachtsgebied betreft utiliteitsmetingen.

Maureen Rutten-van Mólken is getrouwd met Wil Rutten.

