Underestimated uncertainties

Hospital-at-home for COPD exacerbations and methodological issues in the economic evaluation of healthcare

Lucas Goossens

Funding

The GO AHEAD trial (chapters 3, 4 and 5) was funded by from the Netherlands Organisation for Health Research and Development (ZonMw), grant numbers 945-50-773 and 945-07-7309, and health insurers VGZ and CZ. The studies on quality-of life measurement and database research in COPD (chapters 6 and 10) were funded by Boehringer Ingelheim GmbH. The study on statistical methods in observational cost-effectiveness analysis (chapters 8 and 9) was funded by the Netherlands Organisation for Health Research and Development (ZonMW), grant number 152002017.

ISBN 978-94-6169-338-9

Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

Underestimated uncertainties

Hospital-at-home for COPD exacerbations and methodological issues in the economic evaluation of healthcare

Onderschatte onzekerheden

Vervroegd begeleid ontslag na een COPD-exacerbatie en methodologische aspecten van economische evaluaties in de gezondheidszorg

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. dr. H.G. Schmidt en volgens besluit
van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 18 januari 2013 om 11.30 uur

door

Lucas Marcus Anthonius Goossens
geboren te Oss

Promotiecommissie

Promotores
Prof. dr. M.P.M.H. Rutten-van Mölken
Prof. dr. C.P. van Schayck (Universiteit Maastricht)

Overige leden
Prof. dr. C.A. Uyl-de Groot
Prof. dr. P.N.R. Dekhuijzen (Radboud Universiteit Nijmegen)
Prof. dr. J.L. Severens

Copromotor
Dr. W.K. Redekop

For Isis, Elvie and Zeb

Contents

General introduction

1

1.1 Introduction

Economic evaluation has been defined as `the comparative analysis of alternative courses of action in terms of both their costs and consequences' [1]. In an economic evaluation in healthcare, two or more interventions are compared in terms of costs and health outcomes. This results in estimates of the incremental health effects and the incremental costs. If one intervention leads to higher costs as well as more health benefits, an incremental cost-effectiveness ratio (ICER) can be calculated. This ratio expresses the cost per additional unit of health gain. An economic evaluation can play a role in policy makers' decisions on whether the health benefits of an intervention are worth the costs.

The scientific field of economic evaluation in healthcare has gained prominence since the 1980s. The number of publications has increased enormously [2] and the methodology has become more sophisticated [3]. The role of results of economic evaluations in policy decisions has increased markedly [4] – especially in the United Kingdom, but also in continental European countries, such as the Netherlands, although important barriers still hamper their use in decision-making [3,4].

Many economic evaluations are based on or conducted in combination with medical studies, in particular phase III randomised controlled trials, also in COPD [5]. These trials are set up to demonstrate the effectiveness of a specific new intervention compared to a conventional treatment or to placebo. However, an economic evaluation is more than a medical study with costs as an additional outcome measure. It is more complicated than combining the effectiveness results from a trial with the invoices from physicians, hospitals and pharmacies. Economic evaluations are often characterized by challenges that do not occur in purely clinical trials. A number of these challenges are central to this thesis.

Firstly, in many cases costs cannot be observed directly. They are generally different from payments or tariffs. There is a consensus that, from a societal point of view as well as from a healthcare perspective, the relevant costs of a resource are the opportunity costs: the value of the forgone benefits because the resource cannot be applied for the next best alternative purpose [1]. This means that all relevant resources that are used in and after the intervention should be identified and valued separately. Payments do not always reflect all resources at proper societal values, especially when insurance payments are involved or when tariffs apply to a combination of resources and actions (e.g. diagnosis-treatment combination), possibly averaged across a heterogeneous patient population.

When the benefits of treatment stretch into the future, costs that are incurred in later times must be included in the analysis as well. When a societal perspective is taken, all costs must be taken into account, whoever bears them. This may imply that informal care is valued, even if it is not paid.

Secondly, the objective of many, although not all, clinical trials are to find statistical proof that one intervention is superior to another one. In an economic evaluation the *size* of

the treatment effect is essential, not merely its existence. If the effect is estimated in an RCT that was designed to establish superiority, the estimated size may not be generalizable to the full patient population for whom the intervention is relevant. RCTs often include relatively homogenous, precisely selected groups of patients, who may not be completely representative of the patients who will later be treated in real life. In order to achieve external validity, an observational study can be conducted on patients in real life. However, this study design lacks the benefit of randomisation, which is aimed at balancing treatment groups in terms of prognostic characteristics. Adjustment for possible differences between treatment groups is required.

Thirdly, the size of the treatment effect must be estimated in terms of natural units of health in order to be useful in an economic evaluation. Possible outcome measures include absolute differences, risk differences and risk ratios. However, the results of medical and epidemiological studies are often expressed in hazard ratios – in time-to-event analyses – or odds ratios – for binary count data. These ratios are sufficient when merely the existence of a treatment effect is investigated, but they are purely statistical measures without a clear intuitive interpretation and, more importantly, not directly transferable to the size of the relevant treatment effect. They cannot be used in economic evaluations, in which outcome measures such as survival time and numbers of successes and failures are required instead.

Fourthly, in order to make the results of economic evaluations comparable across interventions and diseases, health gains must be expressed in a common outcome measure. Health-related quality-adjusted life years (QALYs) encompass both survival time and quality of life, which enables them to reflect changes in morbidity as well as mortality. In general, they are calculated by multiplying time spent in a certain health state with quality weights. The scale of the QALY weights included are anchored by the health states perfect health (weight=1) and death (weight=0). Weights for other health states may be below 0 but not higher than 1. A year spent in perfect health equals 1 QALY.

The papers in this thesis address these challenges. The thesis consists of two parts. The calculation of treatment costs is a major issue in this part, which is dedicated to the GO AHEAD study (Assessment of GOing Home under Early Assisted Discharge). This multicentre randomised controlled trial, examined several aspects of an early assisted discharge program for patients who were hospitalized for a COPD exacerbation in the Netherlands.

The second part is centered on the remaining three methodological issues mentioned above: natural units of health gain, quality-of-life measurement and observational studies.

1.2 Part I. Hospital-at-home after a COPD exacerbation

COPD is a chronic disease that is currently ranked the fourth cause of death globally, and the sixth in high income countries [6]. The World Health Organization has estimated that

worldwide 64 million people suffer from the disease [7]. A Dutch study found a prevalence of 11.6% in the population of 55 years and older [8], which would mean that 567,000 people had the disease by the end of 2011. Many of these cases are undiagnosed [9,10].

Due to an aging population and late effects of smoking, the prevalence is expected to rise in the coming decades [11]. COPD is characterised by airflow limitation which is not fully reversible. The symptoms, which include sputum production, cough and dyspnoea, are chronic and progressive over time [12]. The disease cannot be cured, but it can be treated to limit symptoms, prevent exacerbations and improve quality of life.

An acute exacerbation of COPD is a sudden, temporary deterioration of symptoms. Depending on the severity, exacerbations strongly affect the health-related quality of life of patients [13,14]. They lead to faster lung function decline [15] and a higher risk of mortality [16].

Most exacerbations can be treated with medication. However, in many cases, patients are admitted to the hospital. Indeed, exacerbations are the main cause of hospitalisation for COPD. The average annual frequencies have been estimated to vary from 0.11 for patients with mild COPD (GOLD stage I, as defined by airflow obstruction [12] and 0.16 for moderate disease (GOLD II), to 0.22 and 0.28 for severe and very severe COPD (GOLD III and IV) [17]. Nevertheless, the extent to which patients are prone to exacerbations varies substantially within GOLD stages [18].

With a mean length of stay in the hospital of 8.5 to 10 days [16,19], the large number of hospital admissions for exacerbations among COPD patients results in a high pressure on scarce hospital beds and high healthcare costs. Across several countries, hospitalisations have been shown to be a main driver for costs of COPD treatment [20-26]. In the Netherlands, they were estimated to account for 27% of all costs, including for account for 18 to 50% of total healthcare expenditure for COPD, including the costs of homecare and nursing homes [27]. Only medication accounted for a larger share of expenditures.

Several studies have shown that some patients with an exacerbation, who would otherwise be admitted to the hospital, can be treated at home safely after examination in the emergency department or a short hospital admission [28-32]. This is called hospital-at-home. Hospital-at-home aims to avoid admission, or reduce length of stay (early assisted discharge schemes). While no differences were found in the number of readmissions, mortality and disease-specific quality of life between hospital-at-home and usual hospital care [28-32], the results on costs and cost-effectiveness have been mixed. While some studies found cost savings [33-36], others observed increased costs [37], or yielded inconclusive results [32,38].

Hence, treating COPD exacerbations at home has the potential to reduce costs. To what extent it does so in the Dutch setting, remains to be demonstrated. Apart from the exact description of the intervention, the health economic impact of a therapy depends heavily on national and local treatment patterns, healthcare delivery structures, funding and reimbursement systems, absolute and relative differences in unit costs of resource use and drug prices.

There is a risk that cost savings are overestimated in hospital-at-home studies [39]. Firstly, the transfer of part of the patient care to unpaid, informal caregivers, does not mean that it is costless. Secondly, inpatient hospital days are not a homogenous commodity. They consist of many actions and provisions by nurses, physicians, laboratory personnel, administrative staff and others. The combination of resource that is contained in an inpatient day, may vary per patient, as well as across diseases and interventions. Crucially for hospital-at-home studies, treatment intensity may decrease over the course of an admission. This makes it likely that the savings from a reduction of one inpatient day are likely to be lower than the average daily cost [1,39].

Aim and research questions part 1

The objective of the first part of the thesis was to perform a full economic evaluation of the GO AHEAD trial of early assisted discharge of COPD patients after an exacerbation and to investigate treatment preferences of patients and informal caregivers.

The following research questions were formulated to achieve this aim:

- What evidence on cost savings for hospital-at-home services is currently available and what is its quality?
- What is the effectiveness, cost-effectiveness and cost-utility of early assisted discharge after a COPD exacerbation in the Netherlands from a healthcare perspective and a societal perspective?
- How can an early assisted discharge program for COPD patients be designed in order to optimize its acceptability to patients and informal caregivers?

1.3 Part II. Methodological issues

The second part of this thesis is devoted to the three remaining methodological issues that were mentioned at the beginning of this chapter: natural units of health gain, quality-of-life measurement and observational studies.

Natural outcome measures and quality-of-life

Medical interventions are aimed at improving quality of life, prolonging life, or at the combination of both. In the first case, if the effects of treatments are to be completely appreciated, health-related quality-of-life gains must be determined in order to calculate the gain in QALYs. Among the most widely used instruments to measure these is the EQ-5D [40]. It has been validated and applied in the stable phase of COPD [41] as well as in severe exacerbations of the disease, which require hospital admission [13,42]. It has been criticized for not being sensitive enough to changes in health status of COPD patients. However, it has

been shown to be able to distinguish between different GOLD stages of airflow obstruction [43]. This thesis contains an evaluation of the EQ-5D to detect the recovery from a moderate COPD exacerbation, which is defined as requiring antibiotic or systemic steroid therapy but no hospital admission [44]. It results in the tentative estimation of the health loss due to a moderate exacerbation.

When the prolonging of life is concerned, economic evaluations require an estimate of the difference in survival time. However, in time-to-event (or survival) analyses, the treatment effect is often expressed in a hazard ratio: the time-dependent probability of instantly experiencing the event of interest for subjects in one treatment group, as a proportion of this probability in the reference group. Parametric survival models can be used to estimate projected mean time-to-event [45], but their outcomes may suffer from non-collapsibility and from bias due to censoring [46-49]. Earlier studies have often confused these two phenomena, but it is essential that they are distinguished and addressed properly.

Bias in observational studies: potential solutions

Randomised controlled trials (RCTs) have long been viewed as the gold standard for estimating treatment effects of healthcare interventions. In ideal randomised experiments (i.e. large sample size, perfectly randomised, no loss to follow up, full adherence to assigned treatment and no measurement error) internal validity is assured. Association is causation: association measures can be directly interpreted as effect measures.

However, patients enrolled in clinical trials may not be representative of the population to which the therapy will eventually be applied. Treatment effectiveness and healthcare resource use in daily practice may be different from those in clinical trials. This is increasingly recognized by researchers, policy makers and decision makers responsible for pricing and reimbursement of healthcare interventions. Evaluations in 'real world' daily practice can provide policymakers with results that are more relevant and applicable.

As an alternative to RCTs, the use of observational data allows investigators to estimate the (cost)-effectiveness of treatment in day-to-day practice. Yet this approach has its own challenges. In daily clinical practice treatments are not randomly assigned. Treatments are assigned based on patient prognosis, preferences, regional or institutional conventions, time or other factors. Consequently, the prognosis of patients receiving a treatment will often differ systematically from that of patients not receiving a treatment. In other words, treatment assignment and prognostic factors may be associated. If this confounding is not removed or reduced, the treatment effect will be either underestimated or overestimated.

In COPD, the disease severity stage is a crucial factor for which adjustment is required. The most widely used severity classification is the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification [12]. It is based on airflow obstruction, which is associated with mortality [50], exacerbations [51,52], healthcare utilization and costs [53]. An association with quality of life exists as well, although this is less strong [54]. Because

it is increasingly recognized that COPD is not only a lung disease but a multi-dimensional disease with many systemic, extra-pulmonary effects, composite measures of COPD severity have been proposed, but almost all of these are still partly based on the forced expiratory volume in 1 second as a percentage of the predicted value (FEV_1-%predicted) [55-59].

However, measurements of airflow obstruction and information on disease severity are not routinely included in the databases and registries that are available for retrospective observational studies. It appears obvious that COPD severity is related to healthcare resource use and demographic data (such as age, sex and smoking status). However, this association must be quite strong if resource use is to be used as proxy for disease severity. It is not clear that this is the case and that an algorithm can be derived that strongly links the severity of COPD as classified by GOLD to variables that are commonly available in routine databases.

Another issue in observational studies is the technique that is applied in order to adjust for observed prognostic factors. Traditionally, regression methods have been used for this purpose. More recently, methods based on propensity scores have become increasingly popular. They are aimed at removing the association between treatment and prognostic factors and to create a RCT-like design. The propensity score is defined as a subject's probability of receiving a specific treatment assignment, based on certain observed characteristics of that subject. This score, which is usually derived from on a logit or probit regression model, can be used in several ways to address confounding. Methods include matching, weighting and stratification on the propensity score and regression adjustment with the propensity score as a covariate [60].

Examples of the application of propensity-score methods in cost-effectiveness analyses are still scarce and their relative performance in this context has not been investigated. More knowledge about the value of these methods in cost-effectiveness analyses would be useful. One of the reasons is that cost data have different properties than data on clinical effectiveness. More importantly, health economic studies examine two outcomes simultaneously, incremental costs and incremental effects, and combine them into incremental cost-effectiveness ratios (ICERs).

Aim and research questions part 2

The objective of the second part of this thesis is to contribute to methodological advancement in the field of health economics and economic evaluations, in particular with regard to observational studies and the estimation of health outcomes.

The following research questions are posed:

- Can the EQ-5D be used to measure changes in health-related quality of life during the recovery from a moderate COPD exacerbation?
- Is it possible to adjust for COPD severity in database studies by using variables that are routinely available, when no spirometry data have been recorded?

- How should propensity-score based adjustment methods be applied in observational cost-effectiveness studies?
- What is the impact of non-collapsibility and censoring bias on hazard ratios and predicted mean survival in Weibull models?

1.4 Outline of this thesis

This thesis is structure as followed. The next four chapters deal with hospital-at-home and the GO AHEAD trial. **Chapter 2** presents the results of a systematic review of cost studies of hospital-at-home for several diseases and interventions. The focus is on the quality of the cost calculations. **Chapter 3** contains the backgrounds and the study protocol of the GO AHEAD trial. In **chapter 4**, the incremental effectiveness, costs and cost-effectiveness of the program are evaluated. The preferences of patients and informal caregivers for different characteristics of aspects of hospital-at-home are investigated in a discrete choice experiment in **chapter 5**.

The next five chapters focus on a number of methodological issues in economic evaluations. In **chapter 6**, the usefulness of the EQ-5D questionnaire for measuring health-related quality of life during moderate COPD exacerbations is determined. In **chapter 7**, the concepts of non-collapsibility, confounding and omitted-variable bias are disentangled and demonstrated in the context of mean predicted survival in parametric proportional hazard survival models. **Chapter 8** compares the validity and accuracy of the results of several propensity-score-based statistical techniques in survival models for observational studies in a simulation study. In **chapter 9** the reliability of these methods in the context of economic evaluations is investigated. The focus of **chapter 10** is whether the severity of COPD can be derived from databases by relying on other patient characteristics and healthcare resource use. Finally, **chapter 11** is the general discussion.

1.5 References

1 Drummond MF, Sculpher MJ, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.
2 Leeson GW. The demographics and economics of UK health and social care for older adults. Oxford Institute of Ageing Working Paper. Oxford, 2004.
3 Brousselle A, Lessard C. Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. Soc Sci Med. 2011; 72: 832-839.
4 Eddama O, Coast J. A systematic review of the use of economic evaluation in local decision-making. Health Policy. 2008; 86: 129-141.
5 Rutten-van Molken MP, Goossens LM. Cost effectiveness of pharmacological maintenance treatment for chronic obstructive pulmonary disease: a review of the evidence and methodological issues. Pharmacoeconomics. 2012; 30: 271-302.

6 World Health Organization. World Health Statistics 2008. Geneva, 2008.
7 World Health Organization. The global burden of disease: 2004 update. Geneva, 2008.
8 van Durme YM, Verhamme KM, Stijnen T, et al. Prevalence, incidence, and lifetime risk for the development of COPD in the elderly: the Rotterdam study. Chest. 2009; 135: 368-377.
9 Pena VS, Miravitlles M, Gabriel R, et al. Geographic variations in prevalence and underdiagnosis of COPD: results of the IBERPOC multicentre epidemiological study. Chest. 2000; 118: 981-989.
10 Vandevoorde J, Verbanck S, Gijssels L, et al. Early detection of COPD: a case finding study in general practice. Respir Med. 2007; 101: 525-530.
11 Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. Lancet. 2007; 370: 765-773.
12 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of COPD. Available at: http://www.goldcopd.com. 2011.
13 O'Reilly JF, Williams AE, Rice L. Health status impairment and costs associated with COPD exacerbation managed in hospital. Int J Clin Pract. 2007; 61: 1112-20.
14 Seemungal TA, Donaldson GC, Paul EA, et al. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998; 157: 1418-22.
15 Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 2002; 57: 847-852.
16 Donaldson GC, Wedzicha JA. COPD exacerbations .1: Epidemiology. Thorax. 2006; 61: 164-168.
17 Hoogendoorn M, Feenstra TL, Hoogenveen RT, et al. Association between lung function and exacerbation frequency in patients with COPD. Int J Chron Obstruct Pulmon Dis. 2010; 5: 435-444.
18 Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med. 2010; 363: 1128-1138.
19 Saynajakangas O, Kinnunen T, Tuuponen T, Keistinen T. Length of stay and interval to readmission in emergency hospital treatment of COPD. Age Ageing. 2004; 33: 567-570.
20 Nielsen R, Johannessen A, Benediktsdottir B, et al. Present and future costs of COPD in Iceland and Norway: results from the BOLD study. Eur Respir J. 2009; 34: 850-857.
21 Miravitlles M, Murio C, Guerrero T, Gisbert R. Costs of chronic bronchitis and COPD: a 1-year follow-up study. Chest. 2003; 123: 784-791.
22 Koleva D, Motterlini N, Banfi P, et al. Healthcare costs of COPD in Italian referral centres: a prospective study. Respir Med. 2007; 101: 2312-2320.
23 Dal Negro RW, Tognella S, Tosatto R, et al. Costs of chronic obstructive pulmonary disease (COPD) in Italy: the SIRIO study (social impact of respiratory integrated outcomes). Respir Med. 2008; 102: 92-101.
24 Detournay B, Pribil C, Fournier M, et al. The SCOPE study: health-care consumption related to patients with chronic obstructive pulmonary disease in France. Value Health. 2004; 7: 168-174.
25 Nowak D, Dietrich ES, Oberender P, et al. [Cost-of-illness Study for the Treatment of COPD in Germany]. Pneumologie. 2004; 58: 837-844.
26 Bilde L, Rud Svenning A, Dollerup J, et al. The cost of treating patients with COPD in Denmark – a population study of COPD patients compared with non-COPD controls. Respir Med. 2007; 101: 539-546.
27 Hoogendoorn M, Feenstra TL, Rutten-van Molken MP. Projections of future resource use and the costs of asthma and COPD in the Netherlands. Ned Tijdschr Geneeskd. 2006; 150: 1243-1250.
28 Cotton MM, Bucknall CE, Dagg KD, et al. Early discharge for patients with exacerbations of chronic obstructive pulmonary disease: a randomized controlled trial. Thorax. 2000; 55: 902-906.

29 Davies L, Wilkinson M, Bonner S, et al. “Hospital at home” versus hospital care in patients with exacerbations of chronic obstructive pulmonary disease: prospective randomised controlled trial. BMJ. 2000; 321: 1265-1268.
30 Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003; 21: 58-67.
31 Ojoo JC, Moon T, McGlone S, et al. Patients’ and carers’ preferences in two models of care for acute exacerbations of COPD: results of a randomised controlled trial. Thorax. 2002; 57: 167-169.
32 Skwarska E, Cohen G, Skwarski KM, et al. Randomized controlled trial of supported discharge in patients with exacerbations of chronic obstructive pulmonary disease. Thorax. 2000; 55: 907-912.
33 Frick KD, Burton LC, Clark R, et al. Substitutive Hospital at Home for older persons: effects on costs. Am J Manag Care. 2009; 15: 49-56.
34 Nicholson C, Bowler S, Jackson C, et al. Cost comparison of hospital- and home-based treatment models for acute chronic obstructive pulmonary disease. Aust Health Rev. 2001; 24: 181-187.
35 Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003; 21: 58-67.
36 Puig-Junoy J, Casas A, Font-Planells J, et al. The impact of home hospitalization on healthcare costs of exacerbations in COPD patients. Eur J Health Econ. 2007; 8: 325-332.
37 Shepperd S, Harwood D, Gray A, et al. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. BMJ. 1998; 316: 1791-1796.
38 Aimonino Ricauda N, Tibaldi V, Leff B, et al. Substitutive «hospital at home» versus inpatient care for elderly patients with exacerbations of chronic obstructive pulmonary disease: a prospective randomized, controlled trial. J Am Geriatr Soc. 2008; 56: 493-500.
39 Jonsson B, Lindgren B. Five common fallacies in estimating the economic gains of early discharge. Soc Sci Med. 1980; 14: 27-33.
40 Brooks R, Rabin R, De Charro F. The measurement and valuation of health status using EQ-5D: A European perspective. Evidence from the EuroQol BIOMED Research Programme . Dordrecht: Kluwer Academic Publishers; 2003., 2003.
41 Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004; 13: 873-84.
42 Menn P, Weber N, Holle R. Health-related quality of life in patients with severe COPD hospitalized for exacerbations – comparing EQ-5D, SF-12 and SGRQ. Health Qual Life Outcomes. 2010; 8: 39.
43 Rutten-van Molken MP, Oostenbrink JB, Tashkin DP, et al. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? Chest. 2006; 130: 1117-1128.
44 Cazzola M, MacNee W, Martinez FJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. Eur Respir J. 2008; 31: 416-469.
45 Tappenden P, Chilcott J, Ward S, et al. Methodological issues in the economic analysis of cancer treatments. Eur J Cancer. 2006; 42: 2867-2875.
46 Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. Stat Sci. 1999; 14: 29-46.
47 Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1984; 71: 431-444.
48 Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. J Clin Epidemiol. 1991; 44: 77-81.
49 Kalbfleisch JD, Kalbfleisch JD. Misspecified proportional hazard models. Biometrika. 1986; 73: 363-369.

50 Anthonisen NR, Wright EC, Hodgkin JE. Prognosis in chronic obstructive pulmonary disease. Am Rev Respir Dis. 1986; 133: 14-20.
51 Burge PS, Calverley PM, Jones PW, et al. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. BMJ. 2000; 320: 1297-303.
52 Dewan NA, Rafique S, Kanwar B, et al. Acute exacerbation of COPD: factors associated with poor treatment outcome. Chest. 2000; 117: 662-671.
53 Andersson F, Borg S, Jansson SA, et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). Respir Med. 2002; 96: 700-8.
54 Ferrer M, Alonso J, Morera J, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. The Quality of Life of Chronic Obstructive Pulmonary Disease Study Group. Ann Intern Med. 1997; 127: 1072-1079.
55 Briggs A, Spencer M, Wang H, et al. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. Arch Intern Med. 2008; 168: 71-79.
56 Celli BR, Marin JM, Cote CG, et al. Prognostic assessment of patients with COPD. Lancet. 2009; 374: 1885; author reply 1886-7.
57 Eisner MD, Omachi TA, Katz PP, et al. Measurement of COPD severity using a survey-based score: validation in a clinically and physiologically characterized cohort. Chest. 2010; 137: 846-851.
58 Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. Eur Respir J. 2009; 34: 648-654.
59 Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. Lancet. 2009; 374: 704-711.
60 Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2007; 15: 1-21.

A systematic review of hospital-at-home care

Cost savings are overestimated

2

Background The concept of hospital-at-home means that home treatment is provided to patients who would otherwise have been treated in the hospital. Admissions may be shortened (early assisted discharge, EAD) or avoided. This may lead to lower costs, but savings are overestimated if relevant costs are not taken into account appropriately, or if average or generic prices per inpatient day are assumed. The objective of this study was to assess the quality of cost analyses of hospital-at-home studies for acute conditions published from 1996 through 2011 and to present an overview of evidence.

Methods The Medline and NHS HEED databases were searched. Cost calculations were considered incorrect if they failed to meet four criteria. Costs per inpatient hospital day had to be disease-specific. The decreasing intensity of care over the course of an admission had to be reflected in costs of inpatient days. In studies from the societal perspective, informal care costs had to be included. Violating any of these three criteria leads to overestimation of savings. Finally, follow-up had to be at least one month in order to capture relevant downstream costs.

Results Most studies found cost differences in favour of hospital-at-home; the range varied from a saving of €3020 to a cost increase of €1835 per patient. Five out of 29 studies met all criteria. The most frequent problems were the use of average costs per inpatient hospital day and too short follow-up. Informal care costs were wrongly ignored in one study.

Conclusion Many cost savings were probably overestimated.

Lucas M.A. Goossens, Pepijn Vemer, Maureen P.M.H. Rutten-van Mölken

Submitted for publication.

2.1 Introduction

The concept of hospital-at-home means that home treatment is provided to patients who would otherwise have been treated in the hospital. The hospital admission may be avoided completely (admission avoidance, AA) or shortened by early assisted discharge (EAD). Hospital-at-home can be attractive from different points of view. Patients may prefer home treatment to staying in the hospital, they may have a smaller risk of nosocomial infection, beds in the hospital may be used more efficiently and the total costs of care may be lower.

However, there is a risk that cost savings are overestimated in cost(-effectiveness) studies [1]. Firstly, part of the care of patients is transferred to relatives, friends, neighbours and other informal caregivers, even if patients receive formal professional care at home. The fact that care is delivered outside of the health care system, or even that it is unpaid, does not mean that no costs are involved. From a societal perspective, these informal care costs should not be ignored, as is stated in many international guidelines for economic evaluation studies, which recommend all costs to be included, whoever bears them [2,3]. Secondly, since the reduction of hospital days is the principal driver of potential cost savings, the calculation of the costs per hospital day is crucial. Not all hospital days are the same. The intensity of care may be different across diseases and interventions. Standard or reference costs per day may not be sufficiently representative for the use in a particular study [4]. Furthermore, since treatment intensity may decrease over the course of an admission, the savings from a reduction of one inpatient day are likely to be lower than the average daily cost [1,3]. This is likely to have a larger influence on the cost estimates in early discharge studies than in admission avoidance studies.

Three systematic reviews of hospital-at-home studies which paid attention to costs have been published in recent years. A Cochrane review of early assisted discharge that included 12 randomised controlled trials concluded that there is no 'compelling evidence' that hospital-at-home produces cost savings [5]. A second Cochrane review of admission avoidance found, after reviewing five randomised trials, that 'in most instances' estimated costs were lower in the admission avoidance scheme than in conventional hospital treatment [6,7]. A third review of hospital-at-home for exacerbations of chronic obstructive pulmonary disease (COPD) observed that 'substantial' savings had been found in four studies and did not mention any studies indicating cost increases [8]. However, these reviews discussed only a selection of all cost studies of hospital-at-home. They included only randomised controlled trials which mainly focused on the effectiveness of the intervention. Costs and costing methodology were not discussed extensively.

The current systematic review aimed to cast a wider net in order to comprise all papers on the costs of hospital-at-home for acute conditions published from 1996 to the end of 2011. Cost-minimization studies and cost-consequence studies were included as well as full

economic evaluations. Modelling studies, observational studies and pragmatic trials were reviewed as well as randomised controlled trials.

The first objective was an assessment of the quality of the cost analyses of the hospital-at-home studies, with an emphasis on the method of calculating the costs of hospital days, and the inclusion of the costs of readmissions and the costs of informal care. The second objective was to present an overview of currently available evidence on cost savings of hospital-at-home services. Finally, we investigated what proportions of patients were eligible for the various hospital-at-home programs.

2.2 Methods

Data sources

We searched MEDLINE and NHS HEED for studies published between January 1st 1996 and December 31st 2011 using predefined search strategies. Search terms for MEDLINE were (1) cost, costs or cost-effectiveness and (2) hospital-at-home, early (supported/assisted) discharge, home hospitalization/hospitalization, in-home healthcare or hospital in the home. Search terms for additional studies in NHS HEED were 'hospital-at-home' or 'early discharge' or 'early assisted discharge' or 'early supported discharge'. Systematic reviews were checked for additional studies.

Study selection

Two reviewers independently selected studies for inclusion or exclusion. A preliminary selection was made based on titles and abstracts. In case of initial disagreements, studies were included or excluded based on consensus. A second selection round took place after retrieving the full papers. Early discharge programs as well as hospital avoidance programs were selected.

To be considered for inclusion, studies had to comply with the following criteria:

- a comparison between treatment at home and conventional inpatient hospital treatment
- conventional hospital treatment had to include at least a one-night stay and more nights than the alternative;
- treatment of an acute condition or acute worsening of a chronic disease with a planned treatment period lasting 14 days maximum (rehabilitation programs were excluded);
- costs had to be reported in monetary terms;
- patients in the hospital-at-home scheme had to receive formal care at home, in the form of visits by nurses, paramedics and/or physicians;
- interventions involving children, child birth and mental healthcare were not included;
- papers had to have been written in English.

Assessment of quality

The methodological quality of each study was assessed by two reviewers using the Quality of Health Economic Studies instrument (QHES). This method uses the answers to 16 questions to express the overall quality of each study in one score on a 0 to 100 scale, with a higher score indicating a better quality [9]. Each study was independently assessed by two reviewers. Since the QHES contains questions on both empirical health economic studies and economic modelling as well as on both costs and health outcomes, not all questions are applicable to each study. After the two reviewers had reached consensus on the applicability of the questions, the final score was calculated as the average of the two reviewers' total scores, expressed as a percentage of the potential maximum score on the applicable questions.

In addition to the overall quality assessment, a more in-depth examination of the cost analysis of each study was performed. The first focal point was the reduction of hospital costs in early discharge studies: were all hospital days assumed to be equally costly (average costing) or were the avoided days analysed separately (marginal costing). The second point was the use of disease-specific or general hospital costs. Thirdly, the time horizon of the cost analysis was recorded. The fourth issue was the inclusion of costs of informal care.

The QHES contains a question on the correctness and clarity of the cost calculations. No points were awarded for this question when an early discharge study used average costs of hospital days instead of marginal costs. Also, no points were granted when generic costs were applied instead of disease-specific costs in early assisted discharge as well as admission avoidance studies. Other reasons for assigning zero points included a time horizon of less than one month, the use of tariffs instead of opportunity costs, except when the payer perspective was taken, and a lack of explanation, which made the assessment of the cost estimate impossible. For observational studies, the calculations were not considered correct if no adjustment was made for possible differences in baseline characteristics. The exclusion of informal care cost did not automatically lead to a lower QHES score, since the cost estimate could still be valid in the health care or payer perspectives. Studies were not excluded from further analysis based on the QHES score.

Data extraction

The study characteristics and outcomes of interest were the diagnosis group, intervention, country, study perspective, the estimated costs of each treatment, the cost difference, the length of hospital stay for each group, the duration of home care, the total duration of care in the hospital followed by the home care, and the proportion of patients who were recruited in the study. We distinguished three study perspectives. A perspective was labelled "societal" when all costs were taken into account, whoever bore them. These included costs within the healthcare sector and outside of it, such as productivity losses, travel costs and, potentially, informal care costs. Resource use had to be valued in terms of opportunity costs, not tariffs, charges, or payments to healthcare providers, because they may not reflect the real costs

in terms of the value of resources used. A perspective was labelled "healthcare" when only costs within the healthcare sector were taken into account and resource use was valued by the same method as in the societal perspective. Under the payer perspective, all costs were expressed as tariffs, charges, or payments by an insurer or other payer.

In order to facilitate comparisons, all cost savings and cost increases were converted into euros. This was done by using purchasing power parities (PPP) for the average price level of the Eurozone in the year to which the study applied [10]. That means that costs that were already presented in euros were adjusted to the Eurozone-wide PPP. No adjustment for inflation was made.

2.3 Results

Search results

The search resulted in a total of 423 potentially eligible papers. After inspection of the abstracts, 63 remained. Examination of the full text led to the exclusion of 32 additional papers. Two studies were presented in two papers, which were counted as one. The reasons for excluding initially selected articles are presented in table 2.1. All studies on stroke were excluded because they included a rehabilitation period that exceeded 14 days.

Table 2.1 Search results.

First selection round (after reading abstracts)													
Source	**Search result**	**Excluded**	Comparison not right	No formal care at home	No cost analysis	Paediatrics/ obstetrics	Long-term care	Mental health	Language	Review	Study protocol	Double report	**Remaining**
Pubmed	**397**	**341**	208	195	186	44	1	2	25	36	3	0	**56**
NHS Heed	**44**	**42**	21	30	12	5	3	0	0	10	0	0	**2**
Reviews	**43**	**35**	10	4	25	1	1	0	1	0	0	0	**8**
Total	**484**	**418**	239	229	223	50	5	0	26	46	3	0	**66**
Second selection round (after reading full texts)													
Pubmed		**29**	8	3	5	3	13	0	0	0	0	2	**25**
NHS Heed		**1**	1	1	0	0	0	0	0	0	0	0	**1**
Reviews		**5**	0	0	0	1	4	0	0	0	0	0	**3**
Total		**35**	8	4	5	4	16	0	0	0	0	2	**29**

Of the 29 studies that were finally included, 27 covered a single intervention or diagnosis: 5 addressed COPD exacerbations [11-15], 3 joint replacement [16-19][1], 2 breast cancer [20,21], 2 coronary bypass surgery [22,23], 2 (elective) surgery [24-26][2], 7 hospital-at-home in general for acute diseases in elderly patients [27-33], 2 congestive heart failure [34,35] and 4 other diseases (pneumonia [36], cystic fibrosis [37], diabetes [38], deep venous thrombosis [39].

Two studies included multiple, diagnosis groups for which the results were presented separately: COPD, hip and knee replacement, hysterectomy and elderly in general [40]; and COPD, pneumonia, congestive heart failure and cellulitis [41].

The majority was conducted in two countries: England (9) and Australia (6). The United States and Spain each were the focus of three studies and there were two papers from New Zealand. The other papers came from Italy, Sweden, Iceland, Hong Kong, Canada and the Netherlands.

Assessment of methodological quality

Design

The results of the quality assessment are presented in table 2.2. Most studies, 20 out of 29, were randomised controlled trials.

Three studies were designed as sequential controlled trials. The papers of Frick et al. and Leff et al. are based on the same trial [27,41]. Frick et al. performed analyses per diagnosis, Leff et al. calculated costs for the entire sample. These three trials prospectively followed control group patients before the introduction of a hospital-at-home program and recruited the intervention group afterwards. Hardy et al. and Evans et al. used historical controls [21,33] and Campbell used a self-selected control group [30], which means that patients could choose whether they were treated at home or in the hospital. The control group of Hensher consisted of a small number of patients who were medically eligible for hospital-at-home but who lived outside the scheme's catchment area [17].

Eligibility criteria were the same for both groups in all non-randomised controlled trials. None of them applied a form of regression analysis or matching in order to adjust for possible differences between the baseline characteristics when the cost differences were calculated. Some studies did mention that the characteristics were 'similar' [33] or 'not statistically significantly' different [25]. However, this does not mean that possible confounding was addressed. Frick et al. did perform regression adjustment in order to present efficient p-values, though not for point estimates for cost differences.

Two studies did not include an actual control group. Ting et al. applied the average cost per inpatient hospital day on the treatment duration of the intervention group in order to estimate the costs for the hospital treatment in similar patients [39]. Fleming et al. mentioned

1. The papers by Jester et al. report on the same study. They were counted as one.
2. The papers by Caplan et al. and Board et al. report on the same study. They were counted as one.

Table 2.2 Quality assessment.

First author	Year	Design	Sample size[a]	Perspective	Type	QHES	Disease-specific hospital costs	Average/ marginal hospital costs/day	Tariffs/ costs	Informal care	Costs calculation correct and clear	Time horizon
COPD												
Shepperd	1998	RCT	15, 17	Healthcare	EAD	91	Yes	Marginal	Costs	No	Yes	3 months
Nicholson	2001	RCT	13, 12	Healthcare	AA	58	Yes	Average	Tariffs	No	No	Intervention
Aimonino	2008	RCT	52, 52	Healthcare	AA	75	Yes	Average	Costs	No	*Yes*	Intervention
Frick	2009	Obs.	48, 92	Payer	AA	69	Yes	Unclear	Both	No	No	8 weeks
Skwarska	2000	RCT	122, 62	Healthcare	AA	75	Yes	Average	Costs	No	No	Int./8 weeks
Hernandez	2003	RCT	121, 101	Payer	Both	90	Yes	Average	Both	No	Yes	8 weeks
Puig-Junoy	2007	RCT	103, 77	Payer	Both	81	Yes	Average	Both	No	Yes	8 weeks
Pneumonia												
Richards	2005	RCT	24, 25	Payer	AA	49	Yes	Unclear	Tariffs	No	No	Intervention
Frick	2009	Obs.	54, 89	Payer	AA	69	Yes	Unclear	Both	No	No	8 weeks
Congestive heart failure												
Mendoza	2009	RCT	37, 34	Healthcare	*AA*	79	No	Average	Tariffs	No	No	12 months
Frick	2009	Obs.	37, 71	Payer	AA	69	Yes	Unclear	Both	No	No	8 weeks
Patel	2008	RCT	13, 18	Healthcare	Both	84	Unclear	Unclear	Tariffs	No	No	12 months
Hip replacement												
Sigurdsson	2008	RCT	27, 23	Societal	EAD	82	Yes	Marginal	Costs	No	No	6 months
Hensher 1	1996	Obs.	42, 10	Healthcare	EAD	63	Yes	Average	Costs	No	No	Intervention
Hensher 2	1996	Obs.	61, 6	Healthcare	EAD	51	Yes	Average	Costs	No	No	Intervention
Shepperd	1998	RCT	35, 49	Healthcare	EAD	91	Yes	Marginal	Costs	No	Yes	3 months
Knee replacement												
Shepperd	1998	RCT	45, 39	Healthcare	EAD	91	Yes	Marginal	Costs	No	Yes	3 months

Table 2.2 Quality assessment (continued)

First author	Year	Design	Sample size[a]	Perspective	Type	QHES	Disease-specific hospital costs	Average/ marginal hospital costs/day	Tariffs/ costs	Informal care	Costs calculation correct and clear	Time horizon
Knee or hip replacement												
Jester	2003	RCT	64, 45	Healthcare	EAD	49	Yes	Average	Costs	No	No	Intervention
Lumpectomy												
Evans	2000	Model		Healthcare	AA	46	Yes	Average	Tariffs	No	No	60 days
Mastectomy												
Evans	2000	Model		Healthcare	EAD	46	Yes	Average	Tariffs	No	No	60 days
Lumpectomy or mastectomy												
Bonnema	1998	RCT	62, 63	Societal	EAD	85	Yes	Average	Costs	Yes	No	4 months
Coronary bypass surgery												
Booth	2004	RCT	65, 32	Healthcare	EAD	79	Unclear	Average	Costs	No	No	12 weeks
Penque	1999	RCT	25, 25	Unclear	EAD	68	Unclear	Unclear	Costs	No	No	2 weeks
Hysterectomy												
Shepperd	1998	RCT	108, 122	Healthcare	EAD	91	Yes	Marginal	Costs	No	Yes	3 months
Herniorrhaphy												
Caplan/ Board	1998/ 2000	Obs.	44, 58	Healthcare	EAD	51	Yes	Marginal	Costs	No	Yes	Intervention
Cholystectomy												
Caplan/ Board	1998/ 2000	Obs.	57, 65	Healthcare	EAD	51	Yes	Marginal	Costs	No	Yes	Intervention
Fleming	2000	Obs.	37, na	Healthcare	AA	33	Yes	Unclear	Unclear	No	No	Intervention
Acute care for elderly patients												
Coast	1998	RCT	158, 78	Healthcare	EAD	62	Yes	Average	Costs	No	No	3 months
Leff	2005	Obs.	169, 286	Payer	AA	74	Yes	Average	Both	No	No	8 weeks

Table 2.2 Quality assessment (continued)

First author	Year	Design	Sample size[a]	Perspective	Type	QHES	Disease-specific hospital costs	Average/ marginal hospital costs/day	Tariffs/ costs	Informal care	Costs calculation correct and clear	Time horizon
Board	2000	RCT	50, 47	Healthcare	AA	58	Yes	Average	Costs	No	Yes	Intervention
Jones	1999	RCT	102, 97	Healthcare	AA	86	Yes	Average	Costs	No	Yes	3 months
Campbell	2001	Obs.	30, 21	Healthcare	Both	75	Yes	Marginal	Costs	No	No	3 months
Harris	2005	RCT	143, 142	Healthcare	Both	63	Yes	Average	Costs	No	No	30 days
Shepperd	1998	RCT	50, 44	Healthcare	Both	91	Yes	Marginal	Costs	No	Yes	3 months
Hardy	2001	Obs.	149, nr	Healthcare	EAD	35	Yes	Average	Costs	No	No	Intervention
Diabetes												
Wong	2005	RCT	52, 49	Unclear	EAD	57	No	Average	Tariffs	No	No	Intervention
Deep venous thrombosis												
Ting	1998	Obs.	100	Healthcare	EAD	22	Yes	Average	Costs	No	No	Intervention
Cystic fibrosis												
Wolter	1997	RCT	13, 18	Healthcare	EAD	61	Yes	Average	Costs	No	No	Intervention
Cellulitis												
Frick	2009	Obs.	30, 34	Payer	AA	74	Yes	Unclear	Both	No	No	8 weeks

[a]intervention group, control group

Abbreviations: RCT, randomised controlled trial; Obs., observational study; EAD, early assisted discharge; AA, admission avoidance.

that they estimated the relative costs of the intervention group compared with a 'representative sample' of routine treatment patients. No more details were presented [24].

Study perspective

Most studies, 20, were performed from the healthcare perspective. Sigurdsson et al. and Bonnema et al. adopted a societal perspective [16,20] and five studies explicitly took a payer perspective [12,13,27,36,41]. For the studies by Penque et al. and Wong et al., the perspective could not be determined [23,38].

Generic or disease-specific inpatient hospital costs

Almost all studies, 24, used disease-specific costs per inpatient hospital day. Mendoza et al. and Wong et al. applied generic costs [35,38]. The papers by Booth et al., Penque et al. and Patel et al. did not make clear what kind of inpatient hospital costs they used [22,23,34].

Average or marginal inpatient hospital costs

The large majority, 20 studies, used average costs per hospital day. For the early-assisted-discharge studies, for which we considered using marginal costs as essential for a correct cost calculation, the number was 12 out of 19. For two of this studies, however, the tariff per day was justified by their payer perspective [12,13]. For one EAD study, three AA studies and one combined program, it is not clear whether average of marginal costs per inpatient hospital day were used.

Opportunity costs or tariffs

In 5 studies that took the societal or healthcare perspective, tariffs instead of (or in addition to) opportunity costs were used for important costs categories [14,21,34,35,38]. All studies that took the payer perspective used tariffs for the appropriate cost categories.

Informal care

Only Bonnema et al. included the costs of informal care in their calculations [20]. The other study a societal perspective, Sigurdsson et al., did not do this.

Correct cost calculation

Based on the criteria set out in the Methods section and the results in the preceding subsections, the cost calculations of 7 studies were considered correct [11-13,25,26,28,31,40].

Time horizon

For 14 studies, the follow-up period was not sufficiently long to capture all relevant costs. They only recorded costs for the initial treatment episode. Of the 7 studies with a correct cost calculation, 5 also had an appropriate follow-up period [12,13,28,31,42].

QHES scores

The average QHES score was 65. Six studies scored less than 50 out of the potential 100 points [19,21,24,33,36,39]. The highest score was achieved by Shepperd et al., 91 [40]. The lowest score, 22, was for Ting et al. [39].

Estimated cost differences

All estimates of the cost differences between hospital-at-home and conventional hospital treatment are presented in figure 2.1. Most (sub)studies, 19 out of 29, found cost savings. The largest savings were estimated by Sigurdsson et al. for early assisted discharge after hip replacement in Iceland: €3020. On the other extreme, Harris et al. found cost increases of €1835 for a combined early assisted discharge and admission avoidance scheme for elderly patients in New Zealand.

COPD exacerbation

Results for the treatment of COPD exacerbations are presented in table 2.3. The cost calculations in three studies, by Shepperd et al., Hernandez et al. and Puig-Junoy et al. were considered correct in our quality assessment.

Shepperd et al. is the only COPD study on a pure early-assisted-discharge scheme and the only study that found significantly higher costs for the hospital-at-home group [40]. The difference between the medians is €1603. Although no explanation for this result is given in the paper, it appears to be caused by the high number of days of readmissions to the hospital in hospital-at-home group compared to the usual care group (median 5 vs. 0). The total length of stay for the initial treatment is similar in both groups.

The studies by Hernandez and Puig-Junoy are based on different analyses of the same data sample of a mixed AA and EAD scheme in Spain, which explains why their estimated savings for hospital-at-home are very similar (€931 and €969). In addition to calculating mean costs per patient, Puig-Junoy estimated the expected cost per patient for different levels of disease severity through a regression model [12,13].The savings for patients with average characteristics, slight, moderate and severe COPD were €774, €548, €927 and €1698, respectively.

The remaining four studies investigated an admission avoidance program. Nicholson et al., again in a small sample, observed significant cost savings of €1204 per patient in Australia [14]. Since no details were given on length of stay, it is unclear what drove the savings.

Aimonino Ricauda et al. found non-significant savings of €172 per patient in Italy [11]. Costs per day of treatment were lower for the hospital-at-home group, but this advantage was largely but not entirely offset by the longer duration of care in this group. Furthermore, the control group experienced more readmissions, which were not included in the cost analysis since the time horizon of this analysis was limited to the first acute episode. Finally, 11.7% of the patients in the control group received a short period of care in a long-term

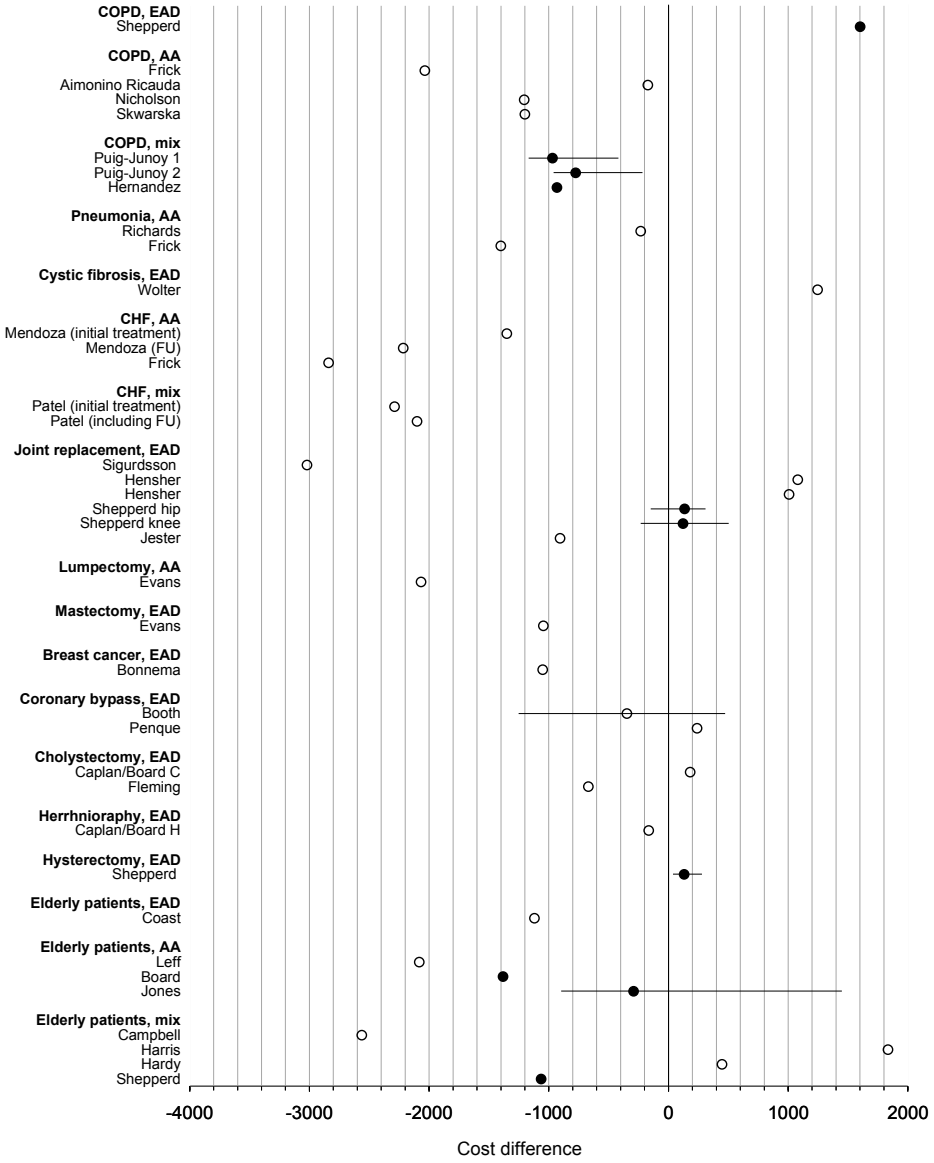

Figure 2.1: Summary of incremental cost estimates of early assisted discharge or admission avoidance versus regular hospital care.

care facility, whereas all hospital-at-home patients stayed at home. These costs were also excluded from the cost analysis.

The largest cost savings were found by Frick et al. for the USA [41]. However, part of this difference might be explained by residual confounding resulting from the observational study design. It is conceivable that patients in the hospital-at-home group were less severely ill than patients in the control group. This is consistent with the shorter duration of care, with less variation, in the hospital-at-home group.

Skwarska et al. found savings of €1163 per patient in England [15]. The uncertainty of hospital costs was addressed in a sensitivity analysis in which the assumed costs per in-patient hospital day were halved. This reduced the cost difference to €19 per patient. The reason for performing this sensitivity analysis is the authors' notion that average bed day costs overestimate the value of potential savings because fixed costs continue if beds are not occupied. However, in the longer run these resources can be expected to be put to use. Nevertheless, base-case cost savings may have been overestimated because of the use of average costs per hospital day.

Pneumonia

The cost calculations of neither study of the costs in hospital-at-home for pneumonia were considered correct in the quality assessment. Richards et al. found a saving of €232 per patient in a New Zealand admission avoidance scheme (see table 2.3) [36]. No patient-level costing was performed, and the statistical significance of this difference was not tested. The

Table 2.3. Estimated costs savings: COPD, pneumonia, cystic fibrosis, congestive heart failure

First author	Type	Year	Country	Costs, hospital-at-home	Costs, usual hospital treatment	Amount saved (original currency)	Amount saved (in PPP euros)[a]	Statistics[b]	Proportion recruited
COPD									
Shepperd	EAD	1998	England	£2380[c]	£1248[c]	-£1132[c]	-1603	0.01	nr
Nicholson	AA	2001	Australia	A$745	A$2543	A$1798	1204	<0.01	15%
Aimonino Ricauda	AA	2008	Italy	$1176	$1391	$215	172	0.38	34%
Frick	AA	2009	USA	$4928	$7274	$2346	2035	<0.05	nr
Skwarska	AA	2000	Scotland	£877	£1753	£876	1199	nr	26%
Hernandez	Both	2003	Spain	1255	2034	€ 778	931	0.003	39%
Puig-Junoy	Both	2007	Spain	1154	1964	€ 810	969	€418-€1169	39%
				1154	1801	€ 647	774	€217-€959	na
Pneumonia									
Richards	AA	2005	New Zealand	NZ$158	NZ$1556	NZ$398	232	nr	27%
Frick	AA	2009	USA	$6072	$7686	$1614	1400	not sign.	nr
Cystic fibrosis									
Wolter	EAD	1997	Australia	A$2476	A$ 5028	A$ 2552	1247	nr	nr
Congestive heart failure									
Mendoza, initial treatm.	AA	2009	Spain	€2541	€4502	€1961	2216	<0.001	nr
Mendoza, follow-up				€3425	€4619	€1194	1350	0.83	na
Frick	AA	2009	USA	$3959	$7232	$3273	2839	<0.001	nr
Patel, initial treatment	Both	2008	Sweden	€586	€3277	€2691	2287	nr	13%
Patel, including follow-up				€2680	€5150	€2470	2100	0.08	13%

[a] A negative amount means a cost increase ; [b] 95% confidence interval or p-value; [c] Medians

Abbreviation: nr, not reported.

American study by Frick et al. found savings of €1400 per patient, which were not statistically significant [41].

Congestive heart failure

All three studies – the cost calculations of which were not considered correct in the quality assessment – found significant cost savings for hospital-a-home, ranging from €2100 to €3566 (see table 2.3). For the Spanish admission avoidance study by Mendoza et al., the cost difference was statistically significant during the initial treatment but not during the follow-up [35].

Although Patel et al. combined admission avoidance and early assisted discharge, their Swedish hospital-at-home program was not more costly (€2287 during the admission phase) than the admission avoidance programs in the other studies (€2216 for the admission phase in Spain and $2839 with 8 weeks of follow-up in the United States) [34,41]. The duration of care is much longer in the study by Mendoza et al.

Joint replacement

The calculations of costs of joint replacement by Shepperd et al. were considered correct in the quality assessment. They found a small and insignificant increase for early assisted discharge for both hip (€135) and knee replacement (€122) in England (see table 2.4). [40]. Hospital-at-home patients received more days of care than hospital patients.

The two English schemes for early discharge after hip replacement that were investigated by Hensher et al. were thought to bring about significant cost increases as well: €1009 and €1081, respectively [17]. Treatment duration was longer than for usual hospital care, while costs per day were equally high or only slightly lower in the hospital-at-home group.

According to the Iceland study by Sigurdsson et al., early assisted discharge after hip replacement led to substantial and significant savings: €3020 from the societal perspective, €2257 from the healthcare perspective [16]. Half of the cost savings from the healthcare perspective were due to the days that patients in the usual care group spent in a convalescent home after leaving the hospital. Furthermore, no patient in the hospital-at-home group had to be readmitted, whereas there were readmissions in the usual care group.

Jester et al. found savings of €905 for the hip and knee replacement in England, but no significance testing was performed [19].

Breast cancer surgery

The cost calculations of neither study of the costs in hospital-at-home for breast cancer surgery were considered correct in the quality assessment. In the Canadian study by Evans et al., calculations were made separately for patients with mastectomy and lumpectomy (see table 2.4). Compared to usual hospital care, net savings for lumpectomy patients who avoided hospital admission were estimated to be €2067. Savings due to inpatient days avoided were offset partly by higher costs for surgery and counselling [21].

Table 2.4 Estimated cost savings: surgery

First author	Type	Year	Country	Costs, hospital-at-home	Costs, usual hospital treatment	Amount saved (original currency)	Amount saved (in PPP euros)[a]	Statistics[b]	Proportion recruited
Hip replacement									
Sigurdsson	EAD	2008	Iceland	$8550	$11,952	$3402	3020	nr	nr
				$5720	$8263	$2543	2257	nr	nr
Hensher	EAD	1996	England	£2780	£2012	-£768	-1081	<0.05	nr
Hensher	EAD	1996	England	£2023	£1306	-£717	-1009	<0.05	nr
Shepperd	EAD	1998	England	£911	£816	-£96	-135	0.59	20%
Knee replacement									
Shepperd	EAD	1998	England	£1462	£1375	-£86	-122	0.55	25%
Knee or hip replacement									
Jester	EAD	2003	England	£2332	£2984	£652	905	nr	nr
Lumpectomy									
Evans	AA	2000	Canada	C$6050	C$8836	C$2786	2067	nr	nr
Mastectomy									
Evans	EAD	2000	Canada	3424	6046	$2622	1945	nr	nr
Lumpectomy or mastectomy									
Bonnema	EAD	1998	Netherlands	$3062	$4382	$1320	1050	0.0007	nr
Coronary bypass surgery									
Booth	EAD	2004	England	£6127	£6381	£254	346	-€474 - €1252	nr
Penque	EAD	1999	USA	$17,749	$17,480	-$269	-239	nr	nr
Cholystectomy									
Caplan/Board	EAD	1998/2000	Australia	A$2887	A$3282	A$265	182	nr	nr
Fleming	AA	2000	Australia	A$3145	A$4129	A$984	668	nr	56%
Herrhnioraphy									
Caplan/Board	EAD	1998/2000	Australia	A$2082	A$2321	A$239	164	nr	nr
Hysterectomy									
Shepperd	EAD	1998	England	£772	£679	-£93	-131	<0.01	35%

[a] A negative amount means a cost increase; [b] 95% confidence interval or p-value.

Abbreviation: nr, not reported.

The estimated savings for mastectomy patients were €1945. For this patient group, a risk of readmission was assumed, whereas this was disregarded for mastectomy patients. No significance testing was performed.

Both patient groups were combined in an early assisted discharge program investigated by Bonnema et al. [20], who found significant savings of €1050 due to hospital-at-home care for breast cancer surgery in the Netherlands. This might be an overestimation because the inpatient days that were substituted for care at home, were less costly than the average costs that were used in this study.

Coronary bypass surgery

The cost calculations in both studies of hospital-at-home after coronary bypass surgery were considered incorrect in the quality assessment.

Penque et al. found that an early assisted discharge program for coronary bypass surgery in Minnesota, United States, was €239 per patient more expensive than regular treatment (see table 2.4) [23]. This difference was not tested statistically. However, the authors argue that patients in both trial arms were discharged earlier than 'usual'. Therefore, they also compared the costs in the hospital-at-home group with an 'average' of costs for patients undergoing this form of surgery, for which no source was given. This difference was said to be significant and in favour of hospital-at-home. It is not clear whether a comparison of patients in the trial with the average patient was justified. In the English study by Booth et al. the total healthcare costs per patient in the three months after surgery were €346 lower in the hospital-at-home group than in the usual hospital care group [22]. This difference was not statistically significant. It was caused almost entirely by the higher costs of readmissions in the usual hospital care group. The costs during the initial episode were almost the same in both groups.

Elective surgery

The calculations of costs of hysterectomy by Shepperd et al. were considered correct in the quality assessment [40]. The calculations in the other studies of elective surgery were not. Shepperd et al. investigated early assisted discharge after hysterectomy in England and found significantly higher costs for the hospital-at-home group. The cost difference was €131 per patient (see table 2.4). Total treatment duration was longer for hospital-at-home care than for usual care.

Two papers reported on one study on assisted discharge after herniorrhaphy or cholystectomy in Australia and are discussed together [25,26]. They found that per-patient costs were €164 and €182 lower than usual hospital care, respectively. However, part of these differences is due to the fact that patients in the usual care group spent more time in the hospital before the operation.

According to Fleming et al. avoiding hospital admission for cholystectomy in England led to costs savings of €668 per patient. These savings were achieved by reducing the length of stay with merely one day. However, hospital-at-home patients recovered in a day-surgery unit, which was less expensive than the recovery room of the main theatre suite [24].

Acute diseases in elderly patients

For three of the studies of hospital-at-home for elderly patients, the cost calculations were considered correct in the quality assessment [28,31,40].

In a pure admission avoidance scheme, Jones found statistically insignificant savings of €291 per patient when the intention-to-treat principle was applied (see table 2.5). Patients who refused to be treated at home, were still counted as hospital-at-home patients anyway.

When these patients were left out, the savings were larger (€1505 per patient) and statistically significant.

Board et al. reported significant savings of €1270 per patient for an Australian admission avoidance program [27,31]. In the English study by Shepperd et al., the costs were €448 per patient higher than for usual hospital care in England, but this difference was not statistically significant [40]. Duration of treatment was longer in the hospital-at-home group.

Three of the other studies reported savings for hospital-at-home. The largest savings were calculated by Campbell et al. in a program for early assisted discharge and admission avoidance in an English hospital (€2562 per patient) [30]. The difference was statistically significant.

Two other English studies found savings as well, but they did not provide statistical testing results. Coast et al. reported that early assisted discharge was less costly than regular hospital treatment, a difference of €1119 per patient [29]. Hardy et al. found savings for a mix of assisted discharge and admission avoidance: €1063 [33].

The study by Leff et al. was based on the same data as the study by Frick et al., in which treatments for different diseases were assessed separately. The estimated cost savings for admission avoidance were €2081 and were statistically significant [27].

Harris et al. found a significant cost increase for hospital-at-home: €1835 per patient in New Zealand [32]. In the latter study, the duration of treatment in the hospital-at-home group was much longer than the length of stay in the regular hospital care group. The authors note that the program was terminated shortly after the study was completed.

Different interventions

The cost calculations in the studies for the remaining interventions were not considered correct in the quality assessment. Wolter et al. examined the costs of an early discharge hospital-at-home program for patients with an infective exacerbation of cystic fibrosis in Australia [37]. In case of a recurrent episode, patients alternated treatment arms. A possible effect of the first treatment on later exacerbations was not taken into account. They found a healthcare cost difference of €1247 in favour of hospital-at-home, which was not tested statistically (table 2.5).

Wong et al. investigated early assisted discharge for patients with diabetes type I or II who were admitted to achieve glycaemic control in Hong Kong (table 2.5) [38]. Treatment at home consisted of telephone calls by nurses to monitor glycaemic levels until these were stable. If necessary, patients were asked to return to the clinic. The costs for these returns were not included in the analysis. The authors found that the costs in the hospital-at-home program – after the first days of admission to the hospital – were less than 3% of the costs in the regular hospital care group, saving €1647 per patient.

Table 2.5. Estimated cost savings: elderly patients, other diseases.

First author	Type	Year	Country	Costs, hospital-at-home	Costs, usual hospital treatment	Amount saved (original currency)	Amount saved (in PPP euros)[a]	Statistics[b]	Proportion recruited
Acute care for elderly patients									
Coast	EAD	1998	England	£2545	£3336	£791	1119	nr	nr
Leff	AA	2005	USA	$5081	$7480	$2399	2081	<0.001	25%
Board	AA	2000	Australia	AS$1764	A$3775	A$2011	1381	<0.001	nr
Jones	AA	1999	England	£3671	£3877	£206	291	0.647	nr
				£3697[c]	£4761[c]	£1064[c]	1505[c]	0.025	nr
Campbell	Both	2001	England	£2864	£4748	£1884	2562	<0.001	42.2%, 50.1%[d]
Harris	Both	2005	New Zealand	NZ$6524	NZ$3525	-NZ$2999	-1835	<0.0001	nr
Shepperd	Both	1998	England	£1705	£1388	-£317	-448	0.09	nr
Hardy	EAD	2001	England	nr	nr	£782	1063	nr	nr
Diabetes									
Wong	EAD	2005	Hong Kong	HK$322	HK$12,210	HK$11,888	1647	<0.001	nr
Deep venous thrombosis									
Ting	EAD	1998	Australia	AS$2613	AS$3852	AS$1239	838	nr	nr
Cellulitis									
Frick	AA	2009	USA	$4927	$8011	$3084	2675	not sign.	nr

[a] A negative amount means a cost increase; [b] 95% confidence interval or p-value; [c] Excluding patients who refused their assigned treatment ; [d] Admission avoidance and early assisted discharge, respectively.

Abbreviations: PPP, purchasing power parity; nr, not reported.

Ting et al. assessed cost savings of early assisted discharge for patients with lower-limb deep venous thrombosis in Australia [39]. The estimated cost savings per patient were €838 (table 2.5). No significance testing was performed.

Frick et al. estimated the mean savings, when avoiding admission of cellulitis patients in the US, at €2675 per patient (table 2.5) [41]. This difference was statistically significant.

Proportion of recruited patients

The authors of 15 of the selected (sub)studies presented the proportion of patients that underwent the intervention and who were recruited in the study. For COPD exacerbations, proportions varied from 14.9% [14] to 38.8% [12,13] of all patients hospitalized for a COPD exacerbation. In acute care for elderly patients, 25.2% to 50.1% was considered eligible. The lowest number, 13.0% was found in the study by Patel et al. [34]. Fleming et al. recruited the highest proportion of admitted patients, 56% [24].

2.4 Discussion

This systematic review provided an overview and quality assessment of all cost- studies of hospital-at-home programs published between 1996 and 2011. Cost calculations were considered correct only if they met a number of explicit criteria. Firstly, the unit costs of inpatient hospital days had to be disease-specific. Secondly, in studies of early discharge programs, these unit costs were also required to take into account that the intensity of care might decrease in the course of an admission. Thirdly, in studies that took the societal perspective, the costs of informal care had to be included. Fourthly, the time horizon had to be at least one month after admission in order to capture costs and effects after the initial treatment episode. This particularly includes the costs of re-admissions in patients that were discharged early or the costs of hospital admissions in patients that were part of an admission avoidance program.

Only 5 out of 29 studies, containing 10 out of 43 sub studies, met all criteria and had a sufficiently clear explanation of their methodology [12,13,28,31,40]. The most frequent problem was the use of average costs per inpatient hospital day, which was problematic in at least 10 studies of early discharge schemes [17-22,29,32,37-39]. This is unclear for one study [23].

Generic instead of disease-specific inpatient costs were applied in at least two studies [35,38], while this was unclear for three others [22,23,34]. Informal care costs were left out of one study that stated to adopt a societal perspective [16].

The follow-up period was considered too short in 13 studies to include all relevant down-stream costs of the hospital-at-home program, in particular costs of (re-) admissions [11,14,17-19,23-26,33,36-39]. In one study, the explicit criteria were met but insufficient information on the calculation of costs was presented [15]. None of the observational studies adjusted for possible baseline differences between the treatment groups. [21,24,27,30,33,41].

Violating any of first three explicit criteria leads to an overestimation of the savings (or underestimation of the cost increases) from a hospital-at-home program compared to regular hospital care. The last days in the hospital tend to be the less costly since patients require less care than the average over the regular admission period [1,3]. Exchanging these days for care at home leaves the more expensive hospital days in place. Generic, non-disease-specific costs per inpatient hospital day are based on the average patient across several conditions and diseases. Patients who are eligible for hospital-at-home are likely to need less care than this average patient. Even if disease-specific costs are used, when they are not tailored to the specific target group of the hospital-at-home program, they may lead to overestimation of the costs of inpatient hospital days, and of the cost savings, because the severely ill patients are usually excluded from the program. Finally, when the costs of informal care are not included in the analyses, part of the care for the patient in a hospital-at-home program wrongly appears to bear no opportunity costs. Violation of the criterion on the time horizon does not have to

lead to overestimation of the savings. It only does so when the need for re-hospitalization or other forms of care is higher in the hospital-at-home group than in the usual care group.

This leads to the conclusion that, while most studies found that hospital-at-home was cost-saving, these savings were probably overestimated. Savings may not disappear after adjusted calculations, but they will be considerably smaller.

In some cases, however, hospital-at-home led to cost increases [17,23,25,26,32,33,40]. There were two reasons for these results. In many studies, the treatment period was not fixed and patients in the regular hospital care group were discharged from hospital earlier than patients in the hospital-at-home group stopped home care. The threshold for adding another day to the treatment period appeared to have been lower in hospital-at-home. In these studies, the lower costs per day were offset by a higher number of days in treatment. The second reason is that patients in the hospital-at-home group sometimes had more readmissions.

There was a correlation between the criteria for unbiased calculation of cost savings and the QHES score. The three studies that met our criteria for a correct cost calculation scored very high on the QHES score, while studies with low QHES scores also scored badly on our criteria, especially the time horizon. At the same time, some studies with biased estimates still achieved high QHES scores. Since violation of our criteria led to an overestimation of cost savings, better designed studies were less likely to find large savings. This difference was enhanced by the fact that Shepperd et al. found cost increases for hospital-at-home for several diseases.

One more issue with the cost estimates of inpatient hospital days requires discussion. In all papers, as in other cost-effectiveness studies, they were treated as fixed unit costs. However, an inpatient hospital day is a complex combination of staff time, hotel costs, equipment, medication, overhead costs. This combination is different for each patient, which means that uncertainty surrounding the mean of the cost estimate should be taken into account. This would require patient-level data on each separate inpatient hospital day, the collection of which would not be feasible in most studies. Instead, the uncertainty could be acknowledged by performing sensitivity analyses with different costs per hospital admission.

We did not perform a meta-analysis, which would involve pooling results from different studies in order to combine evidence by statistical methods. The question whether hospital-at-home in general is cost-saving may not be very meaningful. Hospital-at-home is not one uniform treatment. It is an umbrella term which covers a great deal of combinations of treatment characteristics. Treatment patterns, health systems, the organisation that delivers the homecare (hospital or a separate organisation), price levels, price differentials differ across jurisdictions, time and hospital-at-home programs. Costs and cost savings depend on the exact specifications of the program. Furthermore, many of the selected paper did not contain the information that is required to perform a meta-analysis.

The answer to the question on the effectiveness and costs of these programs always depends on who is treated. Hospital-at-home treatments are probably acceptable to patients and care providers only if their health outcomes are at least equivalent to those of regular hospital admissions. Patients who are most at risk of needing intensive treatment at home – where it may be more expensive as well as less effective or more hazardous than in the hospital – may not be included in the program. Different assessments of the suitability of patients may be made in different programs. The proportions of eligible patients in the selected studies were quite low, probably because physicians wanted to avoid risks.

As in any review, the risk of publication bias must be acknowledged. It is conceivable that negative results of hospital-at-home schemes were not published. Indeed, the expectation of savings was the rationale for the implementation of many programs. The hospital-at-home scheme reported by Harris et al. was terminated after the study ended with negative conclusions [32].

This review discussed more studies than earlier reviews [5-8]. We aimed at including observational studies as well as randomised controlled trials. Even if RCTs are considered the gold standard in obtaining evidence on the efficacy of treatments, well-executed studies in daily practice – which cannot always be randomised – are useful.

If the value of the current review is not in combining all available evidence in one conclusion about the costs of hospital-at-home, it is in identifying the problems in existing studies and in presenting an overview of the evidence per disease area. The results of most of the studies should at best be regarded as the upper limit of potential cost benefits: especially the savings due to the reduction of inpatient hospital days are overestimated.

2.5 References

1 Jonsson B, Lindgren B. Five common fallacies in estimating the economic gains of early discharge. Soc Sci Med. 1980; 14: 27-33.

2 College voor zorgverzekeringen (Health Care Insurance Board). Guidelines for pharmacoeconomic research. Diemen, 2006.

3 Drummond MF, Sculpher MJ, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.

4 Tan SS, Van Gils CW, Franken MG, et al. The unit costs of inpatient hospital days, outpatient visits, and daycare treatments in the fields of oncology and hematology. Value Health. 2010; 13: 712-719.

5 Shepperd S, Doll H, Broad J, et al. Early discharge hospital at home. Cochrane Database Syst Rev. 2009; (1): CD000356.

6 Shepperd S, Doll H, Angus RM, et al. Admission avoidance hospital at home. Cochrane Database Syst Rev. 2008; (4): CD007491.

7 Shepperd S, Doll H, Angus RM, et al. Avoiding hospital admission through provision of hospital care at home: a systematic review and meta-analysis of individual patient data. CMAJ. 2009; 180: 175-182.

8 Ram FS, Wedzicha JA, Wright J, Greenstone M. Hospital at home for patients with acute exacerbations of chronic obstructive pulmonary disease: systematic review of evidence. BMJ. 2004; 329: 315.
9 Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care. 2003; 41: 32-44.
10 Organisation for Economic Co-operation and Development (OECD). PPPs and exchange rates. stats.oecd.org 2011
11 Aimonino Ricauda N, Tibaldi V, Leff B, et al. Substitutive "hospital at home" versus inpatient care for elderly patients with exacerbations of chronic obstructive pulmonary disease: a prospective randomized, controlled trial. J Am Geriatr Soc. 2008; 56: 493-500.
12 Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003; 21: 58-67.
13 Puig-Junoy J, Casas A, Font-Planells J, et al. The impact of home hospitalization on healthcare costs of exacerbations in COPD patients. Eur J Health Econ. 2007; 8: 325-332.
14 Nicholson C, Bowler S, Jackson C, et al. Cost comparison of hospital- and home-based treatment models for acute chronic obstructive pulmonary disease. Aust Health Rev. 2001; 24: 181-187.
15 Skwarska E, Cohen G, Skwarski KM, et al. Randomized controlled trial of supported discharge in patients with exacerbations of chronic obstructive pulmonary disease. Thorax. 2000; 55: 907-912.
16 Sigurdsson E, Siggeirsdottir K, Jonsson H,Jr, et al. Early discharge and home intervention reduces unit costs after total hip replacement: results of a cost analysis in a randomized study. Int J Health Care Finance Econ. 2008; 8: 181-192.
17 Hensher M, Fulop N, Hood S, Ujah S. Does hospital-at-home make economic sense? Early discharge versus standard care for orthopaedic patients. J R Soc Med. 1996; 89: 548-551.
18 Jester R, Hicks C. Using cost-effectiveness analysis to compare Hospital at Home and in-patient interventions. Part 1. J Clin Nurs. 2003; 12: 13-19.
19 Jester R, Hicks C. Using cost-effectiveness analysis to compare Hospital at Home and in-patient interventions. Part 2. J Clin Nurs. 2003; 12: 20-27.
20 Bonnema J, van Wersch AM, van Geel AN, et al. Cost of care in a randomised trial of early hospital discharge after surgery for breast cancer. Eur J Cancer. 1998; 34: 2015-2020.
21 Evans WK, Will BP, Berthelot JM, et al. Breast cancer: better care for less cost. Is it possible? Int J Technol Assess Health Care. 2000; 16: 1168-1178.
22 Booth JE, Roberts JA, Flather M, et al. A trial of early discharge with homecare compared to conventional hospital care for patients undergoing coronary artery bypass grafting. Heart. 2004; 90: 1344-1345.
23 Penque S, Petersen B, Arom K, et al. Early discharge with home health care in the coronary artery bypass patient. Dimens Crit Care Nurs. 1999; 18: 40-48.
24 Fleming WR, Michell I, Douglas M. Audit of outpatient laparoscopic cholecystectomy. Universities of Melbourne HPB Group. Aust N Z J Surg. 2000; 70: 423-427.
25 Caplan G, Board N, Paten A, et al. Decreasing lengths of stay: the cost to the community. Aust N Z J Surg. 1999; 69: 433-437.
26 Board N, Caplan G. Implications of decreasing surgical lengths of stay. Aust Health Rev. 2000; 23: 62-76.
27 Leff B, Burton L, Mader SL, et al. Hospital at home: feasibility and outcomes of a program to provide hospital-level care at home for acutely ill older patients. Ann Intern Med. 2005; 143: 798-808.
28 Jones J, Wilson A, Parker H, et al. Economic evaluation of hospital at home versus hospital care: cost minimisation analysis of data from randomised controlled trial. BMJ. 1999; 319: 1547-1550.

29 Coast J, Richards SH, Peters TJ, et al. Hospital at home or acute hospital care? A cost minimisation analysis. BMJ. 1998; 316: 1802-1806.
30 Campbell H, Karnon J, Dowie R. Cost analysis of a hospital-at-home initiative using discrete event simulation. J Health Serv Res Policy. 2001; 6: 14-22.
31 Board N, Brennan N, Caplan GA. A randomised controlled trial of the costs of hospital as compared with hospital in the home for acute medical patients. Aust N Z J Public Health. 2000; 24: 305-311.
32 Harris R, Ashton T, Broad J, et al. The effectiveness, acceptability and costs of a hospital-at-home service compared with acute hospital care: a randomized controlled trial. J Health Serv Res Policy. 2005; 10: 158-166.
33 Hardy C, Whitwell D, Sarsfield B, Maimaris C. Admission avoidance and early discharge of acute hospital admissions: an accident and emergency based scheme. Emerg Med J. 2001; 18: 435-440.
34 Patel H, Shafazand M, Ekman I, et al. Home care as an option in worsening chronic heart failure – a pilot study to evaluate feasibility, quality adjusted life years and cost-effectiveness. Eur J Heart Fail. 2008; 10: 675-681.
35 Mendoza H, Martin MJ, Garcia A, et al. 'Hospital at home' care model as an effective alternative in the management of decompensated chronic heart failure. Eur J Heart Fail. 2009; 11: 1208-1213.
36 Richards DA, Toop LJ, Epton MJ, et al. Home management of mild to moderately severe community-acquired pneumonia: a randomised controlled trial. Med J Aust. 2005; 183: 235-238.
37 Wolter JM, Bowler SD, Nolan PJ, McCormack JG. Home intravenous therapy in cystic fibrosis: a prospective randomized trial examining clinical, quality of life and cost aspects. Eur Respir J. 1997; 10: 896-900.
38 Wong FK, Mok MP, Chan T, Tsang MW. Nurse follow-up of patients with diabetes: randomized controlled trial. J Adv Nurs. 2005; 50: 391-402.
39 Ting SB, Ziegenbein RW, Gan TE, et al. Dalteparin for deep venous thrombosis: a hospital-in-the-home program. Med J Aust. 1998; 168: 272-276.
40 Shepperd S, Harwood D, Gray A, et al. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. BMJ. 1998; 316: 1791-1796.
41 Frick KD, Burton LC, Clark R, et al. Substitutive Hospital at Home for older persons: effects on costs. Am J Manag Care. 2009; 15: 49-56.
42 Shepperd S, Harwood D, Gray A, et al. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. BMJ. 1998; 316: 1791-1796.

3

Effectiveness and cost-effectiveness of early assisted discharge for Chronic Obstructive Pulmonary Disease exacerbations

The design of a randomised controlled trial

Background Exacerbations are the main cause for hospitalization of Chronic Obstructive Pulmonary Disease (COPD) patients. Hospitalisations result in a high pressure on hospital beds and high health care costs. Because of the increasing prevalence of COPD this will only become worse. Hospital-at-home is one of the alternatives that has been proved to be a safe alternative for hospitalisation in COPD. Most schemes are early assisted discharge schemes with specialised respiratory nurses providing care at home. Whether this type of service is cost-effective depends on the setting in which it is delivered and the way in which it is organised.

Methods GO AHEAD (Assessment Of Going Home under Early Assisted Discharge) is a 3-months, randomised controlled, multi-centre clinical trial. Patients admitted to hospital for a COPD exacerbation are either discharged on the fourth day of admission and further treated at home, or receive usual inpatient hospital care. Home treatment is supervised by general nurses. Primary outcome is the effectiveness and cost effectiveness of an early assisted discharge intervention in comparison with usual inpatient hospital care for patients hospitalised with a COPD exacerbation. Secondary outcomes include effects on quality of life, primary informal caregiver burden and patient and primary caregiver satisfaction. Additionally, a discrete choice experiment is performed to provide insight in patient and informal caregiver preferences for different treatment characteristics. Measurements are performed on the first day of admission and 3 days, 7 days, 1 month and 3 months thereafter. Ethical approval has been obtained and the study has been registered.

Conclusion This article describes the study protocol of the GO AHEAD study. Early assisted discharge could be an effective and cost-effective method to reduce length of hospital stay in the Netherlands which is beneficial for patients and society. If effectiveness and cost-effectiveness can be proven, implementation in the Dutch health care system should be considered.

Cecile M.A. Utens, Lucas M.A. Goossens, Frank W.J.M. Smeenk, Onno C.P. van Schayck, Walter van Litsenburg, Annet Janssen, Monique van Vliet, Wiel Seezink, Dirk R.A.J. Demunck, Brigitte van de Pas, Peter J. de Bruijn, Anouschka van der Pouw, Jeroen M.A.M. Retera, Petra de Laat-Bierings, Loes van Eijsden, Maria Braken, Riet Eijsermans, Maureen P.M.H. Rutten-van Mölken

A version of this chapter was published in BMC Public Health, 2010; 10:618.

3.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a chronic disease that is currently ranked as fourth major cause of death globally [1]. Due to an aging population and late effects of smoking, the prevalence of COPD will increase in the following 20 years [2]. Projections for the year 2030 indicate that COPD will be third major cause of death, as a result of the projected increase of tobacco use, especially among women and low-and middle income countries [1]. COPD is characterised by an airflow limitation which is not fully reversible. Symptoms include sputum production, cough and dyspnoea. These symptoms are chronic and progressive over time [3].

Acute exacerbations of COPD can be defined as 'an event in the natural course of the disease characterized by a change in a patients' baseline dyspnoea, cough and/ or sputum production that is beyond the day-to-day variations, is acute in onset, and may warrant a change in regular medication in a patient with underlying COPD' [3]. The exacerbation frequency is dependent on several factors including disease severity and number of exacerbations in the previous year [4,5]. Although most exacerbations are treated in the community [5], exacerbations are the main cause for hospitalisations in COPD patients [5,6]. Studies have shown that exacerbations and hospital admissions negatively influence patient outcomes, by increasing lung function decline [5,6], decreasing quality of life [7,8], increasing mortality [9] and increasing readmission risk [8,10].

With a mean length of stay in the hospital of 9 days [6,11] the large number of hospital admissions for exacerbations among COPD patients result in a high pressure on scarce hospital beds and high health care costs, accounting for up to 70% of total expenses for COPD [12,13]. Even without intervention, hospital costs will rise as a result of the increasing prevalence of COPD, especially among women. To reduce health care costs, alternatives for hospital treatment have been developed. One alternative that has gained popularity in the last 15 years are hospital-at-home schemes [14,15]. These schemes aim at reducing the pressure on hospital beds and overall health care costs without negatively influencing the patient outcomes and increase patient satisfaction[16] [16].

Hospital-at-home and early assisted discharge

Hospital-at-home is defined as 'a service that provides active treatment by health care professionals in the patient's home for a condition that otherwise would require acute inpatient care, and always for a limited time period' [17]. Hospital-at-home is also known as 'home hospitalisation' or 'hospital in the home'. Depending on the target population of the scheme and the type of care provided, schemes vary in organisational structure and may involve different professionals [17]. Hospital-at-home schemes can be divided in admission avoidance schemes, (early) assisted or supported discharge schemes and combined schemes. Depending on who bears the financial and management responsibilities, the

schemes can further be divided in community resourced or hospital resourced. Community resourced schemes commonly built on the existing infrastructure for care provision in the community, whereas hospital resourced schemes work on an outreach basis and home care is provided by hospital staff.

Hospital-at-home for COPD exacerbations

Hospital-at-home schemes for the treatment of COPD exacerbations specifically, have been studied in several randomised controlled trials [18-24] and various nonrandomised studies including observational studies [25-30] and studies with retrospective analysis [31]. Studies were performed in the United Kingdom [19,20,23-28], Ireland [29], Australia [22], Italy [18] and Spain [21,30]. These studies showed that approximately 25% of all patients with an acute exacerbation of COPD can be treated at home safely with no negative effects on their health outcomes and with great patient satisfaction [16]. These results triggered the wide implementation of hospital-at-home schemes for COPD exacerbations in the United Kingdom over the last 10 years [14,15]. In 2007, the British Thoracic Society developed the Hospital-at-Home in Chronic Obstructive Pulmonary Disease guideline providing a framework for the development and adjustment of hospital-at-home schemes [32].

Most hospital-at-home schemes active in the United Kingdom are assisted discharge schemes, with specialised respiratory nurses providing home care on an outpatient basis [14,15]. However, it remains unknown whether this is the most effective model for hospital-at-home care. The use of generic community district nurses or telephone monitoring might be an option that increases the capacity of the hospital-at-home schemes for COPD exacerbation. Supported by their positive results, Davison et al. [27] suggest that the use of generic community nurses in hospital-at-home schemes should be studied more intensively.

Hospital-at-home for COPD exacerbations initially requiring hospital admission has also been the subject of several cost or cost-effectiveness studies [18,21,22,24,33-35]. Statistically significant and substantial cost savings were found in Australia (€1200 per episode) [22], Spain (€800) [21,34] and the United States (€1700) [33]. No significant cost savings were found for England and Italy [18,35]. (Amounts were converted to euros, using exchange rates of March 2010.) All studies were performed from a health care or payer perspective, meaning that they recorded only costs in the health care sector and not included costs of informal care. Although treating COPD exacerbations at home has the potential to reduce costs, whether and the extent to which it does so in the Netherlands is unknown. Apart from the exact organisation of the hospital-at-home scheme, its health economic impact depends heavily on national and local treatment patterns, health care delivery structures, funding and reimbursement systems, absolute and relative differences in unit costs of resource use and drug prices. The limited transferability of cost-effectiveness results to other settings stresses the need for setting-specific cost-effectiveness studies.

This contribution presents the design of the GO AHEAD trial (GO AHEAD is an acronym for Assessment Of Going Home under Early Assisted Discharge). In this Randomised Controlled Trial (RCT) patients admitted to the hospital for an exacerbation of their COPD are discharged early and monitored at home by nurses.

Research questions

Our primary research question is: "What is the effectiveness and cost-effectiveness of an early assisted discharge intervention compared to hospital care as usual for patients hospitalised with an exacerbation of their COPD." The primary measure of effectiveness will expressed by the change in health status, measured by the change in Clinical COPD Questionnaire (CCQ) [36] scores between randomisation and day 7, while costs include COPD-related health care costs, patients' and informal caregivers' out-of-pocket costs and patients' and informal caregivers' costs of production loss.

The following secondary research questions will be addressed:

1) What is the long-term effectiveness of early assisted discharge compared to hospital care as usual?
2) What is the difference in treatment failures between the early assisted discharge scheme and usual care in hospital? Treatment failure in the intervention group is defined as readmission before day 7 or death before day 7. In the control group treatment failure is defined as death before day 7 or clinical deterioration leading to prolongation of hospital stay after day 7.
3) What is the effect of early assisted discharge on readmission rates after discharge from hospital or the early assisted discharge scheme, in comparison with usual care in hospital?
4) What is the effect of early assisted discharge on mortality after discharge from hospital or early assisted discharge scheme, in comparison with usual care in hospital?
5) What is the effect of early assisted discharge on patients quality of life in comparison with usual care in hospital?
6) What is the effect of early assisted discharge on primary informal caregiver burden in comparison with usual care in hospital?
7) How is patient and primary informal caregiver satisfaction with the early assisted discharge scheme compared with usual care in hospital?

Additionally a discrete choice experiment (DCE) is performed in order to provide more insight in patient and primary informal caregiver preferences for different treatment characteristics.

3.2 Methods

Design, recruitment and outcome measures

The GO AHEAD study is a randomised controlled, multi-centre trial comparing two management strategies for patients admitted to the hospital for a COPD exacerbation. The intervention strategy is early assisted discharge, which implies that patients are discharged early from hospital with a package of home care. Recovery is monitored while patients are further treated at home. This management strategy is compared to usual hospital care, where patients remain hospitalised and are monitored in hospital. The total length of the active, supervised treatment phase for both groups is planned to be seven days. The follow up period of the trial is three months. Figure 3.1 gives a complete overview of the study design. Main focus of the study is not only to perform an effect evaluation, but also a cost evaluation

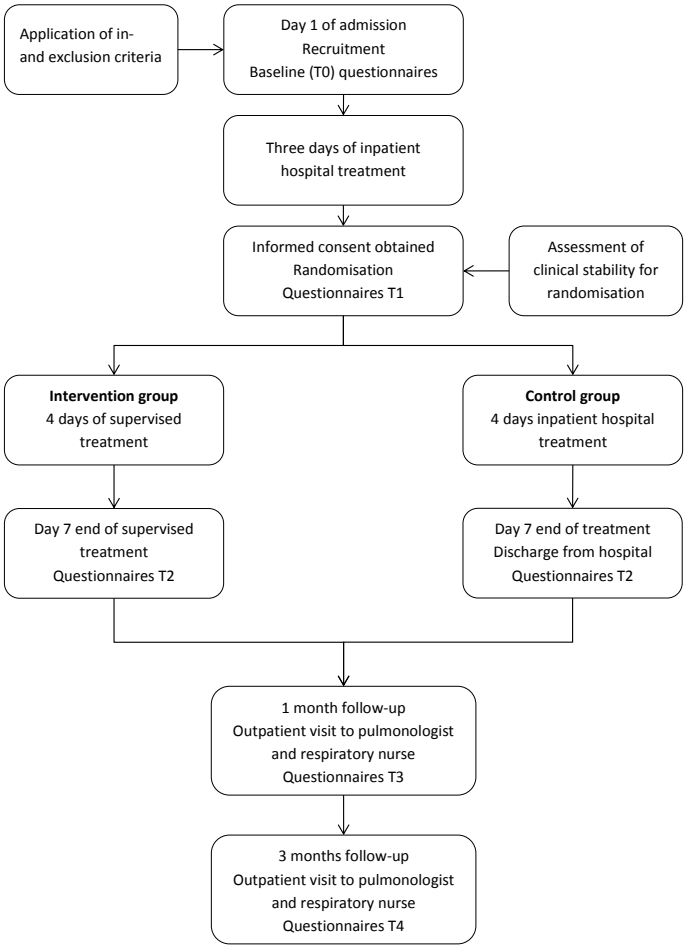

Figure 3.1 Trial design.

and a discrete choice experiment. This trial was approved by the Medical Ethics Committee of the Catharina-hospital Eindhoven, the Netherlands and this approval was reconfirmed by the Medical Ethics Committees of the other participating hospitals.

Setting and recruitment

Patients admitted to one of the participating hospitals because of an exacerbation of their COPD, through either the emergency department or after an unscheduled outpatient visit to the pulmonologist, are screened for eligibility. Patients are assessed for eligibility at two time points. On day one, the pulmonologist and research nurse screen the patient for eligibility according to the inclusion and exclusion criteria listed in table 3.1.

On day 1 patients are considered eligible for potential early discharge if they meet the following inclusion criteria: aged 40 or over, competent, diagnosed with at least COPD GOLD stage 1 (post-bronchodilator FEV1/FVC < 70%[3]) and a smoking history of minimally 10 pack years (PY), hospitalised with a moderate to severe exacerbation and finally, completed informed consent on day three of admission. Patients are excluded if they meet any of the following exclusion criteria: major uncontrolled comorbidity, mental disability, active alcohol- or drug abuse, inability to understand the program, living outside the region of the participating home care organisation, indication for admission to the intensive care unit or non-invasive ventilation and insufficient availability of informal care at home.

On day 1, all patients considered eligible for potential early discharge are invited to participate in the controlled clinical trial in which they either continue their hospital admission or are discharged early to home care. These patients are informed that, if they fulfil the in- and exclusion criteria and are still willing to participate on day 3, they will be randomised on

Table 3.1 Inclusion and exclusion criteria

Inclusion criteria
Age 40 years and over
Competent
Diagnosed with at least COPD GOLD I and 10 Pack Years or grounded susceptibility for COPD
Hospitalised with COPD exacerbation
Completed informed consent on day 3 of admission
Exclusion criteria
Major uncontrolled comorbidity
Mental disability
Active alcohol abuse and/or drug abuse
Inability to understand the program
Living outside care region of the home care organisation
Indication for admission to intensive care unit or non-invasive ventilation
Insufficient availability of informal care at home

day 3 and discharged on day 4. For patients admitted before 12:00 pm, the day of admission is considered as day 1, otherwise, the following day is considered as day 1.

Patients must have signed the informed consent form before randomisation on day 3, which means they have three days to decide whether they want to participate in the trial or not. This procedure is possible because the treatment in the first three days is not different from that of patients not participating in the trial. Data collected before the third day of admission will be destroyed if the patient does not give informed consent on that day. This procedure has been approved by the ethics committee.

Patients who refuse to participate in the trial are invited to participate only in the discrete choice experiment study without participating in the RCT. These patients are contacted by telephone one month after admission and asked to give informed consent for this part of the study. This informed consent form is different from the form that is used in the RCT.

Randomisation procedure

On the third day of admission clinical stability is assessed in order to determine whether patients can be randomised. This design for randomisation is adapted from the Spanish study performed by Diaz et al. [37] although the criteria for clinical stability were adapted to represent the current practice in the participating hospitals and because of the difference in treatment package patients receive at home. Patients need to meet the following criteria to be randomised: (1) acceptable general health defined as decrease of physical complaints, non-dependence on therapies that cannot be given at home and being able to visit the toilet independently; (2) normal or moderately increased blood sugar levels, defined as ≤ 15 mmol/L or ≥ 15 mmol/L while the patient is capable to regulate blood sugars independently at home; and respiratory complaints of dyspnoea, wheezing and rhonchi must have decreased in comparison with day one of admission. A special symptoms scoring list adapted from the one used by Ojoo et al.[23] is used for this. This scoring chart scores the major exacerbation symptoms such as dyspnoea, coughing, mucus production and colour and oedema. By scoring these symptoms daily, improvement or deterioration in comparison with the previous days becomes more visual and supports the pulmonologist when applying the randomisation criteria to the patient.

Randomisation is performed on a 1:1 scale using a computer-generated randomisation list that is placed in sealed envelopes containing the allocation sequence of the two treatment groups. Randomisation is performed per participating location of the hospitals. We chose this procedure to ensure all participating hospitals have a similar proportion of patients in the intervention group and patients in the control group and that the burden for each participating home care organisations is similar. Furthermore, a block size of 6 is applied to create equal numbers in both groups.

The treatment protocol for the following four days is started after randomisation and for the intervention group the process of discharge planning is started.

Treatment protocol day 1-3 and day 4-7

The treatment protocol during the seven days of supervised treatment can be divided in the treatment before randomisation and the treatment after randomisation.

During the first three days of the treatment all patients receive usual care. The pharmacological part of this treatment consists of systemic corticosteroids (10 days in total, first 3 days 50 mg of oral or intravenous prednisolone or other corticosteroid with an equivalent dose, following 7 days of 30 mg oral prednisolone or other corticosteroid with an equivalent dose), nebulised bronchodilators (ipratropium 500 μg/salbutamol 2.5 mg, 4-6 times per day), sub cutaneous thrombosis prophylaxis, stomach protection – because of the high dosage of corticosteroids – and, if necessary, oxygen therapy. Antibiotics are prescribed if patients meet any of the following criteria: increase of the amount of mucus, mucus purulence or CRP > 50 for which no other cause can be determined. First choice of antibiotics is co-amoxiclav. However, if previous mucus cultures show sensitivity for different antibiotics, or patients are allergic, different antibiotics can be prescribed. Antibiotics are prescribed for at least 7 days.

Non-pharmacologic usual care consists of physiotherapy for all patients and dietary advice upon indication [38]. The physiotherapist instructs the patient in breathing and coughing techniques and reactivation. A standardised (additional) written instruction was developed ensuring identical instruction in the participating hospitals. Dietary advice is indicated in case of a Body Mass Index ≤ 21 or 10% unintended weight loss in the six months prior to admission [38].

After randomisation systemic corticosteroids are continued and patients start with pressure metered dose inhaled medication via spacer (at least an β_2-antagonist or anticholinergic with inhaled gluco-corticosteroid). Patients receive inhalation instruction on the day before starting with these inhalations. Patients already using nebulised inhalation medication prior to admission, may continue this after randomisation.

The physiotherapist instructs patients to follow the written instructions at home and dietary consultation is continued as in usual care and at the dietician's judgement.

On the fourth day of admission the intervention group is transferred home and undergoes the early assisted discharge intervention. The control group remains hospitalised and receives usual care in hospital.

In both groups, patient recovery progress is monitored daily using the translated and adapted exacerbation symptom scoring chart.

Early assisted discharge intervention

Patients randomised into the early assisted discharge group are transferred home on day four of admission. The previously described treatment is continued at home and supervised by nurses. These nurses have daily contact with the patient for four consecutive days. Main objective of the supervision of the home treatment is the observation of the patient's recovery and providing counselling and reassurance to the patient and the primary informal

caregiver. The nurses also address medication compliance and inhalation techniques, provide support in applying breathing- and coughing techniques and, if applicable, provide support in adhering to dietary advices. If necessary patients, can be supported in their daily life activities (e.g. washing and dressing) by the home care organisation. During the four days of home treatment, the emphasis lies on the recovery of the exacerbation. Secondary objectives like disease management and smoking cessation are addressed during the first follow up moment one month after randomisation.

General practitioners are informed about patients' participation in the trial but are not directly involved in these patients' home care. In cases of deterioration of the patient, the patient is discussed with the treating hospital pulmonologist and if necessary the patient is readmitted to the hospital. Patients can contact the hospital 24 hours a day, 7 days a week with questions or in case of an emergency.

Follow-up visits

For both treatment groups two follow up visits at the outpatient clinic are scheduled at one month and three months after randomisation. During these visits patients are seen by their own pulmonologist and a respiratory nurse. The visits to the pulmonologist are as in usual care, the visits to the respiratory nurse have a twofold purpose. Firstly, these visits focus on the different aspects of disease management. It is at the discretion of the respiratory nurse which aspects need to be addressed in each specific patient. Secondly, these visits are used to collect, dispense and administer the questionnaires and cost diaries. Additional visits can be planned at discretion of the pulmonologist (e.g. for additional testing) but do not fall under the study protocol. At the three month follow up visit lung function testing and a six minute walking distance test are performed as well.

Data collection

Data are collected on five time points: on the first day of admission (T0), on the third day of admission (randomisation, T1), at the end of the supervised treatment (day 7, T2) and one month (T3) and three months after randomisation (T4). We use self-administered questionnaires and cost diaries to obtain data. The questionnaires are administered when supervision is available. Cost diaries are supplied at two time points (T2 and T3) for the upcoming period, collected at the end of each follow up period and if necessary completed under supervision.

Effect evaluation

Table 3.2 provides an overview of the measures of the effect evaluation and the economic evaluation, and at which time point the measurements are performed.

Table 3.2 Overview of measurements per time point

Measurement	T0	T1	T2	T3	T4
Demographic characteristics	x				
Smoking	x				
Body Mass Index	x				
Living situation	x				
Comorbidity	x				
Coping style (UCL)		x			
Medical treatment prior to admission	x				
Exacerbation severity	x				
Indication for admission	x				
Clinical COPD Questionnaire (CCQ)	x	x	x		x
EuroQol 5D (EQ-5D)		x	x		x
Satisfaction					
patient satisfaction			x		x
primary caregiver satisfaction			x		x
Caregiver Strain Index			x		x
Treatment Failures			x		
Readmissions					x
Mortality					x
Cost diary			x (IG)	x	x
Discrete Choice Experiment				x	
Lung function testing					x
6 Minute Walking Distance					x

Abbreviations: T0 = baseline; T1 = randomisation; T2 = end of treatment; T3 = follow -up 1; T4 = follow-up 2; IG = intervention group only.

Primary outcome measures

Primary outcome measure in this study is the effectiveness of early assisted discharge compared to usual care expressed by the change on the Clinical COPD Questionnaire (CCQ) [36] between the third day of admission (T1) and the last day of supervised treatment (T2 = day 7). The CCQ is a disease specific, ten-item questionnaire that calculates an overall score and three domain scores: symptoms, functional state and emotional state. All items are scored on a seven point scale with 0 representing the best possible score and 6 representing the worst possible score [36]. In this study the version with a 24-hour recall period is used, reflecting the health status of the past 24 hours. The CCQ is responsive to change [36] and a study in patients admitted to the hospital with an acute exacerbation of COPD, indicated that the minimal clinical important difference (MCID) of the CCQ is 0.4 [39].

Secondary outcome measures

Main secondary outcome measurement of the study is the cost-effectiveness. This will be discussed in the economic evaluation section. The following other secondary measure-

ments will be performed. These correspond with the secondary research questions from the last paragraph of the introduction section:

1. Long-term effectiveness, measured with the CCQ change over the time points from day 1 to the end of the follow-up period (T4).
2. Number of treatment failures.
3. Number of readmissions to hospital and time to readmission during the three months follow-up period.
4. Mortality and time to death during the three months follow-up.
5. Generic health related quality of life, measured with the EuroQol (EQ-5D) 5D [40]. The EQ-5D will also be used to calculate quality adjusted life year (QALYs), discussed in the economic evaluation section.
6. Effects on primary informal caregiver burden measured by the Caregiver Strain Index [41].
7. Patient and primary informal caregiver satisfaction with the program. We use a translated version of the satisfaction questionnaire used by Ojoo et al.[23] and extended it with additional questions.

Patient characteristics

Patients are characterised using the following variables: demographic factors (age, gender, socioeconomic status measured through level of education and income), smoking, Body Mass Index, living situation, comorbidity measured with the Charlson Comorbidity Index (CCI) [42], Coping Style measured with the Utrecht Coping List [43], medical treatment at home prior to the admission, severity of the exacerbation, indication for admission and finally severity of the disease are measured as well. In addition, severity of COPD is measured three months after admission at the end of the follow up period by performing lung function testing and a six minute walking test.

Economic evaluation

In accordance with the broad international consensus that economic evaluations should be conducted from a societal perspective [44] this cost-effectiveness analysis will include all costs, irrespective of who actually bears them.

All direct health care and non-health care costs as well as the costs of productivity losses of patient and caregiver within the three months after randomisation will be taken into account.

The following types of resource use will be recorded to calculate direct health care costs: number and length of hospital admissions and readmissions, total amount of time of community nursing care (distinguished by nurse grade), number of visits to the emergency department, number of contacts with pulmonologist, other specialist physicians, general practitioner, respiratory nurse, dietician, physiotherapist, and social worker, number of ambulance rides and medication use. These will be recorded in cost diaries and obtained

from hospital records. Costs of organisational arrangements of the early discharge scheme will also be included.

Direct non-health care costs primarily include paid and unpaid household help, including the time spent by the primary informal caregiver.

Indirect costs are costs of productivity losses. We record the days a patient is absent from paid work. We also ask informal caregivers to record the number of days off work due to caring for the patient. Costs are calculated by multiplying the volume of resource use (such as hospital days, physician visits, time spent by formal and informal caregivers) by a price per unit that includes total, not marginal costs.

In addition to the societal perspective, we will calculate the costs from the financial hospital perspective, the financial perspective of the organisation providing the home care and the perspective of the health care sector. This includes costs covered by the hospital budget, the budget of the homecare organisation and the health care sectors budget, respectively.

The principal health outcome measures in the economic evaluation are the number of patients with a clinically relevant improvement in CCQ between day of randomisation and day 7, and between day of randomisation and month 3, the change in CCQ score between day of randomisation and day 7 and day of randomisation and month 3, the number of QALYs randomisation and the end of the three-month follow-up period. The latter is calculated using the Dutch EQ-5D tariff [45].

Health outcomes will be related to cost outcomes. If one of the treatment options is more effective but also more costly, results will be presented as an incremental cost-effectiveness analysis: the additional cost per additional unit of health gain, which is calculated as the difference in mean costs between early discharge and usual inpatient hospital care divided by the in mean health effects.

Data analysis

Data analysis will be performed according to the intention to treat principle. Data from patients who die, quit participation or are otherwise lost to follow up will be included in the analysis up until the point of drop out. Missing observations will be imputed or weighted appropriately. All primary and secondary outcome measurements will be analysed using analysis of covariance. In order to control for dependency between the repeated measurements within one patient, and for the dependency between patients from the same hospital, multilevel analyses will be performed as well. We set the significance level at $\alpha = 0.05$.

Primary outcome measure

The changes on the primary outcome measurement, the CCQ, will be analysed with in a repeated measures model. The dependent variable is the change in CCQ score from baseline (T1) to the end of the supervised treatment (T2) and to the end of follow-up (T4). Baseline CCQ score (T1) and centre of treatment are considered as covariates. Age, gender and

severity of the disease will also be included in the model as covariates. If necessary, other covariates will be included in the model.

Secondary outcomes measurements

All time-to-event outcomes (i.e. time to readmission and time to death) will be analysed using Kaplan-Meier curves and Cox proportional hazards regression model. Event rates (i.e. treatment failures, readmission rates and mortality rates) will be analysed using an appropriate model for count data (e.g. poisson regression or binomial regression).

Differences in outcomes defined as the mean change from baseline (e.g. long-term effectiveness, primary informal caregiver burden, patient- and primary informal caregiver satisfaction and quality of life) will be analysed using repeated measures model..

Patient- and caregiver preference for place of treatment will be analysed using a logistic regression model.

Cost-effectiveness

In order to derive the total utility experienced over the course of the investigation, the number of QALYs per patient will be calculated as the area under the utility curve.

Uncertainty around the estimates of costs and health outcomes will be addressed by bootstrapping the data with bias correction and acceleration (BCa) [46]. The 95% confidence interval around the difference in mean costs and health outcomes will be determined by taking the 2.5th percentile and the 97.5th percentile of these bootstrap replications. The bootstrap replicates will be plotted in cost-effectiveness planes (CE-planes). A CE-plane is an x-y-diagram with the x-axis representing the difference in health outcome between the treatment and usual care group and the y-axis representing the difference in costs. By plotting all bootstrap replicates in this diagram the uncertainty around the point estimates of the ICERs will be displayed [47]. The information from the CE-planes will be summarised into cost-effectiveness acceptability curves, which represent the likelihood that early assisted discharge is the most cost-effective option at different values of the maximum acceptable willingness to pay (WTP) for a health outcome [48].

Sample size calculation

Primary outcome is the change in the CCQ score between baseline (day 3 of admission) and the end of the supervised treatment (day 7).

Before the start of the study, a preliminary sample size calculation for an independent samples t-test was performed based on the results of a pilot study, where the average CCQ decreased from 3.8 on the day of admission to 2.6 by the end of the supervised treatment The standard deviation of that change was 1.04. With a MCID of 0.4, the required Cohen's effect size d would be 0.385 [99]. For a risk of a type-I error of 5% ($\alpha = 0.05$) and a risk of a type-II error 20% ($1\text{-}b = 0.80$), the required sample size was 214.

However, primary outcome measure in this study is the change in the CCQ score from the third day of admission and the end of the supervised treatment (day 7), which is likely to have a stronger correlation with the baseline score. Therefore, a new sample size calculation for ANCOVA was performed after 85 patients had been treated, without breaking the randomisation code. Taking into account the correlation between the baseline score and the change ($\rho = 0.288$), as well as the standard deviations measured in the trial (0.988 for the intervention group and 0.922 for the control group), the required effect size f is 0.22 and the sample size is 165.

3.3 Discrete Choice Experiment

Background

As part of the GO AHEAD trial we perform a discrete choice experiment (DCE) to explore the preferences of patients and their informal caregivers for different treatment arrangements. The DCE provides quantitative information on the relative importance of the characteristics of the hospital treatment and the early assisted discharge scheme and the rate at which patients are willing to trade between them.

A DCE is a type of conjoint analysis used to determine individual preferences. In this study it involves presenting respondents with a series of choices between an early assisted discharge scenario and a usual hospital care scenario. Each scenario is described in terms of several characteristics, which are called attributes.

DCEs originate from mathematical psychology and have been most widely applied in market research to determine consumer preferences for goods and services and investigate the relative importance of the characteristics of these goods and services [49,50].

Design

A review of literature and conversations with patients and pulmonologists have lead to the selection of seven attributes with two or three levels each for the home treatment options, while the hospital option is kept constant and is not described by attributes. The attributes for the home treatments are: type of nurse (generic or respiratory), number of home visits (1, 2 or 3 per day), copayment (€0, €50 or €100), risk of readmission to hospital within treatment period (1%, 5% or 10%), whom to contact in case of emergency (general practitioner or pulmonary ward in the hospital), number of hours of informal care (1, 3 or 5 hours per day), number of different nurses visiting the patient (1-2 or more than two).

The questionnaire consists of 14 choice scenarios, two of which have a 'right answer' and aimed at testing if the respondent understands the task. There are three versions of the questionnaire, which add up to a D-optimal design of 36 different scenarios. Each respondent is asked to complete one version of the questionnaire.

In each scenario respondents indicate a preference for one of two home treatment options or the complete hospital treatment. Respondents who initially choose the hospital option, are subsequently asked to make a choice between the two home treatments. By using this forced choice question, we ascertain that all respondents provide information about their preferences for attributes of the home treatment.

Data analysis

Depending on the choice pattern of the respondents, the data will be analysed using conditional logit model with alternative-specific constants, a random parameter multinomial logit model (i.e. a mixed logit model) or a multilevel latent class conditional logit model, all with and without interaction effects.

This analysis results in a regression coefficient for each level of the attributes. The estimated coefficients allow conclusions to be drawn about the relative importance of the attributes and possible trade-offs between them.

Furthermore, we will test whether the patients who were assigned to the early assisted discharge scheme have different preferences for various characteristics of the care delivery than patients who were assigned to the conventional inpatient hospital care.

Finally, we will test for differences in preferences between patients, their informal and formal caregivers.

3.4 Discussion

We presented the protocol of the GO AHEAD study, which assesses the effectiveness and cost-effectiveness of an early assisted discharge intervention for patients admitted to the hospital for a COPD exacerbation. It is a multi-centre RCT comparing the early assisted discharge intervention with usual care in the hospital.

Despite research on the effectiveness of early assisted discharge for COPD exacerbations, several aspects of these hospital-at-home schemes remain unclear. This study will provide not only information on the effectiveness and cost-effectiveness, but also on which aspects of the intervention are important for patients to make a certain choice and secondary outcome measurements, namely effects on quality of life, effects on primary caregiver burden and patient and primary caregiver satisfaction.

Several critical success factors are to be mentioned. In the past two decades average length of hospital stay, for acute care of all conditions, has already decreased internationally from approximately 9 days to 6 days [51]. Average length of stay for COPD exacerbations follows these trends, with changes from approximately 8.5 days to 6.5 days [11,52,53]. Although this trend has occurred in the Netherlands as well, average length of hospital stay for COPD patients is still 10.5 days [54]. With the projected increasing prevalence of COPD,

especially for women, in the following years, this leaves room for interventions that reduce length of stay.

Prolonged hospital stay is associated with the presence of comorbidities [55], continuation of conservative therapy (e.g. therapies that can only be given in hospital or pulmonologists wish to observe stable patients) [52] and complex discharge planning that requires additional home care arrangements) [52,53] among others. Moreover, comorbidities are more present in patients with more severe COPD [56], and most patients hospitalised have COPD stages III or more [53,57]. This might suggest that the Dutch hospitalised population is different from that in the United Kingdom. However, the national UK audit from 2008 [57] showed that severity of the disease of patients admitted to the hospital has not changed between 2003 and 2008 (median FEV_1-% = 38% of predicted value, GOLD stage III) and 77% of all patients admitted to hospital have one or more comorbidities.

Early assisted discharge also anticipates for the need for social work involvement during hospitalisation. In the early assisted discharge scheme care at home is arranged for a certain number of days and patients are closely monitored at home. Because of the presence of nurses at home, the possible need for prolonged or extended home care is quickly identified. Arrangements can be made more easily because patients are already in the system of the home care organisation.

In the United Kingdom, early assisted discharge for COPD exacerbations is more common in hospitals that are characterised by greater number of hospital beds, higher numbers of hospital admissions and the presence of respiratory nurses [15]. The Dutch hospitals are of similar as or larger than those in the United Kingdom [51] [51]. The number of admissions is high (11.6 per 10.000 population in 2007 [54]) and respiratory nurses have an important role in patient care. The Dutch health care system also has a large network of primary care organisations that deliver home care that is easily accessible for the population. Therefore the use of generic community nurses in the early assisted discharge scheme is possible.

Despite similarities between the British and Dutch health care system, organisational and financial differences between these countries exist and results from studies performed in one country, with its own characteristics, cannot simply be translated to and implemented in other countries. Similarities and differences should be studied more intensively and taken into account before implementing early assisted discharge in the Dutch health care system. Possible boundaries to implementation in the Netherlands are the different reimbursement systems and budgets of hospital care and home care. An integrated financing structure may facilitate the implementation in the health care system.

In this trial supervision of the treatment at home is either performed by community based, generic nurses or hospital resourced specialised respiratory nurses (nurse practitioners). The use of hospital resourced specialised respiratory nurses is the most frequently described and studied form of supervision at home. Using generic community nurses, who are more available and less costly, could enable the development of more hospital-at-home services.

In this study both strategies for early assisted discharge (hospital resources or community based) and the different professionals involved in home care (generic nurses or specialised respiratory nurses) are being used. When sample sizes of both groups allow it, this study may provide more insight in which model for early assisted discharge is more preferable.

Compared to commonly applied measures of satisfaction, a DCE quantifies the relative importance of the characteristics and levels. It assesses the trade-offs that respondents make and provides an estimate of the overall value of early assisted discharge treatments and the usual inpatient hospital care option.

Because common satisfaction questionnaires do not quantify the relative importance of attributes and levels, it is likely that patient preferences are not represented correctly in organising the process of care delivery. This may lead to suboptimal decision-making and may impair the acceptance of early assisted discharge.

In summary, in this contribution we presented the research protocol of a multi-centre RCT studying the effectiveness and cost-effectiveness of an early assisted discharge scheme for COPD exacerbations in the Netherlands.

3.5 References

1 World Health Organization. World Health Statistics 2008. Geneva, 2008.
2 Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. Lancet. 2007; 370: 765-773.
3 Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2007; 176: 532-555.
4 Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. Eur Respir J Suppl. 2003; 41: 46s-53s.
5 Wedzicha JA, Donaldson GC. Exacerbations of chronic obstructive pulmonary disease. Respir Care. 2003; 48: 1204-13; discussion 1213-5.
6 Donaldson GC, Wedzicha JA. COPD exacerbations .1: Epidemiology. Thorax. 2006; 61: 164-168.
7 Seemungal TA, Donaldson GC, Paul EA, et al. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998; 157: 1418-22.
8 Wang Q, Bourbeau J. Outcomes and health-related quality of life following hospitalization for an acute exacerbation of COPD. Respirology. 2005; 10: 334-340.
9 Hoogendoorn M, Hoogenveen RT, Rutten-van Molken MP, et al. Case fatality of COPD exacerbations: a meta-analysis and statistical modelling approach. Eur Respir J. 2011; 37: 508-515.
10 Almagro P, Barreiro B, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. Respiration. 2006; 73: 311-317.
11 Saynajakangas O, Kinnunen T, Tuuponen T, Keistinen T. Length of stay and interval to readmission in emergency hospital treatment of COPD. Age Ageing. 2004; 33: 567-570.
12 Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. Chest. 2000; 117: 5S-9S.
13 Strassels SA, Smith DH, Sullivan SD, Mahajan PS. The costs of treating COPD in the United States. Chest. 2001; 119: 344-352.

14 Johnson MK, Flanigan U, Fuld J, et al. Hospital at home services for acute exacerbation of chronic obstructive pulmonary disease: a survey of British practice. Health Bull (Edinb). 2001; 59: 163-170.
15 Quantrill SJ, Lowe D, Hosker HS, et al. Survey of early discharge schemes from the 2003 UK National COPD Audit. Respir Med. 2007; 101: 1026-1031.
16 Ram FS, Wedzicha JA, Wright J, Greenstone M. Hospital at home for patients with acute exacerbations of chronic obstructive pulmonary disease: systematic review of evidence. BMJ. 2004; 329: 315.
17 Shepperd S, Doll H, Broad J, et al. Early discharge hospital at home. Cochrane Database Syst Rev. 2009; (1): CD000356.
18 Aimonino Ricauda N, Tibaldi V, Leff B, et al. Substitutive "hospital at home" versus inpatient care for elderly patients with exacerbations of chronic obstructive pulmonary disease: a prospective randomized, controlled trial. J Am Geriatr Soc. 2008; 56: 493-500.
19 Cotton MM, Bucknall CE, Dagg KD, et al. Early discharge for patients with exacerbations of chronic obstructive pulmonary disease: a randomized controlled trial. Thorax. 2000; 55: 902-906.
20 Davies L, Wilkinson M, Bonner S, et al. "Hospital at home" versus hospital care in patients with exacerbations of chronic obstructive pulmonary disease: prospective randomised controlled trial. BMJ. 2000; 321: 1265-1268.
21 Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003; 21: 58-67.
22 Nicholson C, Bowler S, Jackson C, et al. Cost comparison of hospital- and home-based treatment models for acute chronic obstructive pulmonary disease. Aust Health Rev. 2001; 24: 181-187.
23 Ojoo JC, Moon T, McGlone S, et al. Patients' and carers' preferences in two models of care for acute exacerbations of COPD: results of a randomised controlled trial. Thorax. 2002; 57: 167-169.
24 Skwarska E, Cohen G, Skwarski KM, et al. Randomized controlled trial of supported discharge in patients with exacerbations of chronic obstructive pulmonary disease. Thorax. 2000; 55: 907-912.
25 Callaghan S. ACTRITE: Acute Chest Triage Rapid Intervention Team. Accid Emerg Nurs. 1999; 7: 42-46.
26 Chetty M, MacKenzie M, Douglas G, Currie GP. Immediate and early discharge for patients with exacerbations of chronic obstructive pulmonary disease: is there a role in «real life»? Int J Chron Obstruct Pulmon Dis. 2006; 1: 401-407.
27 Davison AG, Monaghan M, Brown D, et al. Hospital at home for chronic obstructive pulmonary disease: an integrated hospital and community based generic intermediate care service for prevention and early discharge. Chron Respir Dis. 2006; 3: 181-185.
28 Gravil JH, Al-Rawas OA, Cotton MM, et al. Home treatment of exacerbations of chronic obstructive pulmonary disease by an acute respiratory assessment service. Lancet. 1998; 351: 1853-1855.
29 Murphy NM, Byrne CC, O'Neill SJ, et al. An outreach programme for patients with an exacerbation of chronic obstructive pulmonary disease. Ir Med J. 2003; 96: 137-140.
30 Sala E, Alegre L, Carrera M, et al. Supported discharge shortens hospital stay in patients hospitalized because of an exacerbation of COPD. Eur Respir J. 2001; 17: 1138-1142.
31 Ansari K, Shamssain M, Farrow M, Keaney NP. Hospital-at-home care for exacerbations of chronic obstructive pulmonary disease: an observational cohort study of patients managed in hospital or by nurse practitioners in the community. Chron Respir Dis. 2009; 6: 69-74.
32 British Thoracic Society Guideline Development Group. Intermediate care--Hospital-at-Home in chronic obstructive pulmonary disease: British Thoracic Society guideline. Thorax. 2007; 62: 200-210.
33 Frick KD, Burton LC, Clark R, et al. Substitutive Hospital at Home for older persons: effects on costs. Am J Manag Care. 2009; 15: 49-56.

34 Puig-Junoy J, Casas A, Font-Planells J, et al. The impact of home hospitalization on healthcare costs of exacerbations in COPD patients. Eur J Health Econ. 2007; 8: 325-332.
35 Shepperd S, Harwood D, Gray A, et al. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. BMJ. 1998; 316: 1791-1796.
36 van der Molen T, Willemse BW, Schokker S, et al. Development, validity and responsiveness of the Clinical COPD Questionnaire. Health Qual Life Outcomes. 2003; 1: 13.
37 Diaz LS, Gonzalez LF, Gomez Mendieta MA, et al. [Evaluation of a home hospitalization program in patients with exacerbations of chronic obstructive pulmonary disease]. Arch Bronconeumol. 2005; 41: 5-10.
38 Stichting Ketenkwaliteit COPD. Richtlijn Ketenzorg COPD [Guideline Transmural Care COPD] . Alphen aan den Rijn: Van Zuiden, 2005.
39 Kocks JW, Tuinenga MG, Uil SM, et al. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. Respir Res. 2006; 7: 62.
40 Brooks R. EuroQol: the current state of play. Health Policy. 1996; 37: 53-72.
41 Robinson BC. Validation of a Caregiver Strain Index. J Gerontol. 1983; 38: 344-348.
42 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40: 373-83.
43 Schreurs PJG, van de Willige G, Brosschot JF, Tellegen B, Graus GMH. Handleiding Utrechtse Coping Lijst UCL (herziene versie) [Manual of the Utrecht Coping List] . Lisse: Swets & Zeitlinger, 1993.
44 Drummond MF, Sculpher MJ, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.
45 Lamers LM, McDonnell J, Stalmeier PF, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006; 15: 1121-1132.
46 DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statist Sci. 1996; 189-228.
47 Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ. 1998; 7: 723-740.
48 van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ. 1994; 3: 309-319.
49 Hensher DA, Rose JM, Green WH. Applied choice analysis: a primer . Cambridge: Cambridge University Press, 2005.
50 Louviere JL, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications . Cambridge: Cambridge University Press, 2000.
51 OECD. OECD Health Data. http:/stats.oecd.org/index.aspx?lang=en] 2010
52 Mushlin AI, Black ER, Connolly CA, et al. The necessary length of hospital stay for chronic pulmonary disease. JAMA. 1991; 266: 80-83.
53 Wong AW, Gan WQ, Burns J, et al. Acute exacerbation of chronic obstructive pulmonary disease: influence of social factors in determining length of hospital stay and readmission rates. Can Respir J. 2008; 15: 361-364.
54 Centraal Bureau voor de Statistiek [Statistics Netherlands]. http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71859NED&D1=6,8&D2=0&D3=0,I&D4=106&D5=9-26&HDR=T&STB=G1,G2,G3,G4&VW=T 2010
55 Crockett AJ, Cranston JM, Moss JR, Alpers JH. An association between length of stay and comorbidity in chronic airflow limitation. Int J Qual Health Care. 2000; 12: 41-46.
56 Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. Eur Respir J. 2008; 32: 962-969.

57 The Royal College of Physicians of London, British Thoracic Society, British Lung Foundation. Report of the National Chronic Obstructive Pulmonary Disease Audit 2008: clinical audit of COPD exacerbations admitted to acute NHS units across the UK. www.rcplondon.ac.uk/clinical-standards/ceeu/Current-work/ncrop/Documents/Report-of-The-National-COPD-Audit-2008-clinical-audit-of COPD-exacerbations-admitted-to-acutre-NHS-units-across-the-UK.pdf

Cost-effectiveness of early assisted discharge for COPD exacerbations in the Netherlands

4

Objectives Hospital admissions for exacerbations of chronic obstructive pulmonary disease (COPD) are the main cost drivers of the disease. An alternative is to treat suitable patients at home instead of in the hospital. This paper reports on the cost-effectiveness and cost-utility of early assisted discharge (EAD) in the Netherlands.

Methods In the multi-centre randomised controlled GO AHEAD trial (n=139), one group received 7 days of inpatient hospital treatment (HOSP), and one group was discharged after three days and treated at home by community nurses for four days. Healthcare resource use, productivity losses and informal care were recorded in cost questionnaires. Microcosting was performed for inpatient day costs.

Results Seven days after admission, mean change from baseline Clinical COPD Questionnaire (CCQ) score was better for HOSP, but not statistically significantly: 0.29 (95% CI: –0.04; 0.61). The difference in the probability of having a clinically relevant improvement was significant, in favour of HOSP: 19.0%-point (0.5%; 36.3%). After three months follow-up, differences in effectiveness had almost disappeared. The difference in quality-adjusted life years (QALYs) was 0.0054 (-0.021; 0.0095). From a healthcare perspective, EAD was cost-saving: -€ 244 (treatment phase, CI: -315; -€168) and -€168 (three months, CI: -€1253; €922). Societal perspective: -€65 (treatment phase, CI: -€152; €25) and +€908 (three months, CI: -€553; €2296). The savings per QALY lost were €31,111 from a healthcare perspective. From a societal perspective, HOSP was dominant.

Conclusions No clear evidence was found to conclude that either treatment was more effective or less costly.

Lucas M.A. Goossens, Cecile M.A. Utens, Frank W.J.M. Smeenk, Onno C.P. van Schayck, Monique van Vliet, Walter van Litsenburg, Maria Braken, Maureen P.M.H. Rutten-van Mölken

Submitted for publication.

4.1 Introduction

Hospital admissions for exacerbations of chronic obstructive pulmonary disease (COPD) are important drivers of the high treatment costs for the disease [1-5]. These admissions put great pressure on scarce hospital beds of respiratory wards, especially during winter months [6]. From an economic and organisational point of view, it may be attractive to treat suitable patients at home instead of in the hospital, if this is medically possible and responsible.

Treatment schemes in which patients are treated and supervised at home, as an alternative to usual hospital treatment, are often called hospital-at-home. [7,8] These schemes may either avoid admission completely or discharge patients from the hospital early and continue treatment at home.

Studies on the costs and cost-effectiveness of hospital-at-home services for patients with a COPD exacerbation have shown varying results. Shepperd et al. concluded that a particular scheme in England led to significantly higher costs [9], while Skwarska et al. found cost savings in a different scheme in the same country.[10] Significant cost savings were reported for hospital-at-home services in Australia [3], Spain [11,12] and the United States [13]. The results in an Italian study were inconclusive [14].

Although these studies were performed in different countries and in different healthcare systems, they had some aspects in common. They all took a healthcare perspective; the costs or value of resources used outside of the healthcare sector were not taken into account. Secondly, the length of treatment was variable in each study. Physicians and/or nurses decided on the timing of discharge from the hospital or from treatment at home, depending on the patient's recovery.

The current paper reports on the cost-effectiveness and cost-utility of an early discharge scheme that is different in the two aspects mentioned above. The study was performed in the Netherlands as part of the GO AHEAD trial (Assessment of GOing Home under Early Assisted Discharge). In this multi-centre randomised controlled trial, one group of COPD patients received usual inpatient hospital treatment for seven days. The other group was discharged after three days and was treated and supervised at home for the remaining four days. The Netherlands have a nation-wide infrastructure for community nursing provided by homecare organisations. Dutch hospitals do not deliver healthcare in the community. Therefore, the care at home in this trial was provided by community-based homecare organisations which employ mostly generically-trained nurses and few specialised nurses. The clinical results of this study have been presented in detail elsewhere [15].

4.2 Methods

Study design

The GO AHEAD study was a randomised, controlled, multi-centre trial comparing two management strategies for patients admitted to the hospital for a COPD exacerbation [16]. After three days of usual hospital treatment patients were randomised to be either discharged home with homecare or to continue hospital treatment. The total duration of this initial treatment phase was 7 days for both groups, unless the treatment failed and patients had to be either re-admitted or had to prolong their hospital stay. Patients were followed for 3 months, with outcome measurements scheduled after 7 days and 3 months.

Patients

Patients admitted to one of the participating hospitals because of an exacerbation of their COPD, were screened for eligibility. On the day of admission they were considered potentially eligible for early discharge if they met the following inclusion criteria: age ≥ 40, sufficiently competent to consider informed consent, and a smoking history of ≥ 10 pack years. In order to be randomised on day three of the admission, their physical and respiratory complaints (dyspnoea, wheezing and rhonchi) had to be improved compared to the day of admission, they should not be depending on therapies that could not be administered at home and they should be able to visit the toilet independently. Also, blood sugar level had to be normal or only moderately increased (≤ 15 mmol/L or regulated independently at home).

Exclusion criteria were: major uncontrolled comorbidity, mental disability, active alcohol or drug abuse, inability to understand the program, living outside the region of the participating homecare organisation, indication for admission to the intensive care unit or non-invasive ventilation and insufficient availability of informal care at home.

Intervention

During the first three days of the treatment all patients received usual hospital care. The pharmacological part of this treatment consisted of systemic corticosteroids (10 days), nebulised bronchodilators, subcutaneous thrombosis prophylaxis and stomach protection. If necessary, oxygen therapy and/or antibiotics were prescribed. Non-pharmacologic usual care consisted of physiotherapy for all patients for breathing- and coughing instructions and dietary advice if indicated (Body Mass Index ≤ 21 or 10% unintended weight loss in the six months prior to admission).

Patients randomised to early assisted discharge were discharged home on the fourth day of admission and further treated at home. Community nurses visited the patient once to three times on the day of discharge and the three following days. Main objective of the supervision of the home treatment was the observation of the patient's recovery and providing counselling and reassurance to the patient and the primary informal caregiver. The nurses

also addressed medication compliance and inhalation techniques, provided support in applying breathing- and coughing techniques and, if applicable, in adhering to dietary advices. If necessary patients, could be supported in their daily life activities (e.g. washing and dressing) by the home care organisation. During the four days of home treatment, the emphasis lies on the recovery of the exacerbation. In case COPD symptoms suddenly worsened, the patients could contact the respiratory hospital ward directly and round-the-clock. The general practitioner was informed of the early discharge but the respiratory physician of the hospital kept the final responsibility.

Effects

The following outcome measures were used: 1) the incremental change from day of randomisation in Clinical COPD Questionnaire (CCQ) score at day 7 and 3 months, 2) the incremental proportion of patients with a clinically relevant improvement in CCQ score (i.e. ≥ 0.4 units) [17] on day 7 and at 3 months, and 3) the gain in quality-adjusted life years (QALYs) after 3 months using utilities as measured by the EQ-5D questionnaire using the Dutch tariff for the valuation of health states.[18] The CCQ score can range from 0 (best possible score) to 6 (worst possible score). Based on the Dutch tariff, the EQ-5D score can range from -0.329 (worst possible utility) to 1 (perfect health).

Costs

Costs were calculated from two perspectives, the healthcare perspective and the societal perspective. The former included only the direct healthcare costs within three months after randomisation. The latter includes direct healthcare costs, non-healthcare costs and costs of productivity loss for the three months follow-up period. This is in accordance with the Dutch recommendations that economic evaluations should be conducted from a societal perspective [19].

In the 7-day treatment phase, the duration of hospital admission and the amount of community nursing care were recorded. Patients randomised to early discharge, were asked to record all additional formal healthcare as well as informal care and days of absence from paid work of the informal caregiver in a 4-day cost diary, a specially designed questionnaire on the amount of resources used on each day.

During the follow-up phase, the following types of resource use were recorded on a weekly basis in costs questionnaires that were distributed for each month of the trial: number and length of hospital readmissions, number of visits to the emergency department, number of contacts with pulmonologist and other specialist physicians, general practitioner, respiratory nurse, homecare, dietician, physiotherapist, and social worker, number of ambulance rides and medication use. Direct non-healthcare costs recorded in these questionnaires were paid and unpaid domestic help, including the time spent by the primary informal caregiver. In order to capture all informal care, respondents were asked to provide

information on help with domestic tasks, personal care and practical support. They were instructed only to consider time they would not have spent on these purposes if the patient had not experienced the exacerbation. Indirect costs were costs of productivity losses. The days a patient was absent from paid work were recorded in the cost questionnaires.

Costs (in 2009 euros) were calculated by multiplying the volume of resource use (such as hospital days, physician visits, time spent by formal and informal caregivers, production losses) by a cost per unit that includes total, not marginal costs. Except for inpatient hospital days, standard unit costs from the Dutch Manual for Costing Studies [20] were used for all healthcare resource use. The unit costs for production losses represented the mean of sex-specific mean wages per day, weighted for the sex distribution in our sample.

Unit costs are presented in table 4.1. Medication prices were based on the official list prices of drugs obtained from retail pharmacists which were published on the internet [21], including value added tax and increased by a standard prescription reimbursement for the pharmacist. Costing for permanent medication was done based on one prescription per three months.

Table 4.1 Unit costs (Euros 2009) [19]

Type of resource	Unit costs
GP, consultation	28
GP, home visit	43
GP, phone call	14
Specialist, consultation	64
Specialist, hour	135.50
Resident, hour	27.85
Nurse in hospital, hour	26.75
Physiotherapist, consult	36
Dietician, consult	27
Pulmonary nurse, consult	36
Social worker	36
Emergency room, visit	151
Ambulance transport	504
Community nurse, hour	65
Domestic community care, per hour	24
Informal care, hour	12.5
Production loss, hour (patient)	29.72
Inpatient hospital day, standard price	435

Costs of inpatient hospital days

Costs for each inpatient hospital day were estimated using the microcosting methodology, which provides cost estimates that most accurately reflect actual costs by identifying cost components at the most detailed level [22,23].

We interviewed ten nurses, three pulmonologists and one laboratory staff member who worked in one of hospitals that participated in the trial. They were required to have been involved in the treatment of at least three randomised patients in order to be acquainted with the disease severity of these patients and of the intensity of care that they needed.

First, during interviews with healthcare professionals participating in the trial, all steps in the treatment and nursing process were identified. Then, at the two hospitals which recruited the most patients, which were the Catharina Hospital in Eindhoven and Atrium Medical Centre in Heerlen, pulmonologists, residents, nurses and laboratory staff who participated in the trial and were very familiar with the type of COPD patients enrolled, were asked to provide best estimates of resource utilization. Separate estimates were made for each inpatient hospital day, in order to detect possible changes in care intensity over the course of the stay. Using standardised reporting templates, the participants were asked how many minutes they spent on each component of care per average patient. Additionally, nurses were asked what proportion of their yearly working hours were 'indirect treatment time', i.e. time not spent directly caring for patients, but for instance on trainings and department meetings. Indirect treatment time was then allocated to patients by adding a mark-up of 24.5% to the amount of direct treatment time they received. This mark-up was calculated as the proportion of time spent on indirect treatment (averaged over all nurses), divided by the proportion of time spent on direct treatment (averaged over all nurses) [23].

Labour time was valued using standardised costs per minute, which reflected national average incomes per profession (including social premiums, fees for irregular working hours, and the costs of replacement during illness) divided by the number of workable minutes per year. For consulting physicians, the standardised time-costs included a 43% mark-up for indirect treatment time. For residents, a mark-up of 30% was applied. The latter percentage was chosen to be in-between those for physicians and nurses.

For hotel and nutrition costs, the national reference costs from the Dutch Manual for Costing Studies was used. For the first and last day of hospital admission, only half of these costs were taken into account. Finally, labour and hotel costs for each hospital day were supplemented with a proportional mark-up for overhead and capital costs (42%), which was the national reference percentage [20]. In these cost calculations, the day of admission was considered as day one when patients were admitted before 12:00 pm. If the patient was admitted after 12 pm, the day of admission was considered day 0 and the following day as day one. Hence, costs of day 1 were calculated separately for patients who were admitted on day 1, and for patients who were admitted on day 0. For inpatient day 4, separate calculations were performed for the patients who were discharged and those who remained in hospital.

Statistical analysis

Effects

The change from day of randomisation in CCQ score and the EQ-5D score were analysed in repeated measures analyses. In these linear models with correlated errors, the covariance matrix was unstructured.

The final model was developed in a backward selection process, which started with the following covariates. In addition to time (i.e. measurement at days 7 (end of treatment) or at the end of follow-up) and the interaction of time and treatment, starting model contained the following variables were tested: baseline CCQ or EQ-5D score, treatment centre, age, gender, co-morbidity, smoking status, living situation, availability of informal caregiver, presence of home care prior to admission, course of oral corticosteroids and/or antibiotics prior to admission. In each step, the variable with the highest p-value was removed unless its exclusion led to a 10% change in the estimated treatment effect [24].

The explanatory variables in the final model were treatment, CCQ score or EQ-5D score at day of randomisation, time (moment of measurement), the interaction of treatment and time, and Charlson's comorbidity score (1 or >1, only in the EQ-5D model) [25].

In the CCQ model the coefficient for treatment could directly be interpreted as the marginal difference in change from the day of randomisation to day 7, i.e. the difference between the hypothetical situations in which all patients were treated at home or all received usual hospital treatment. The marginal difference in change from day of randomisation at three months was the sum of the coefficient for treatment and the coefficient for the interaction of treatment and measurement.

In order to calculate the mean marginal difference in EQ-5D score per measurement, scores were predicted for all patients for each measurement: one score for each treatment. To calculate QALYs the mean utility of two subsequent measurements was multiplied with the number of days between these measurements, the sum of which was divided by 365.25.

The probability of experiencing improvement ≥ 0.4 units in CCQ between the day of randomisation and day 7 and month 3 was analysed with logistic regression analyses. Explanatory variables were treatment and the CCQ score on day of randomisation. The results of these regression analyses were used to predict the probabilities in each treatment group, based on the CCQ score distribution of the full sample (two treatment groups combined).

Costs

In the usual care group, total costs during the 7-day treatment phase include all hospital costs. In the early discharge group, these costs were calculated as the sum of the hospital costs, the community care costs and the costs of healthcare utilization as recorded in the diaries for the 4-day period of homecare.

Total costs during the follow-up phase were calculated as the sum of the predicted monthly treatment costs, the predicted medication costs and the costs of readmissions. To obtain the predicted costs, the monthly cost questionnaires were analysed in a linear repeated measures model with correlated error terms and unstructured covariance. The dependent variable was the costs in a certain month. The explanatory variables were the time (first, second or third month) and the interaction of treatment and time of measurement. The results were used to predict the mean costs per treatment group for each month. Monthly medication costs were analysed in the same way. Because all explanatory variables were dummy variables, it was not necessary to apply a transformation to the cost variable in order to achieve a normal distribution. A generalized estimating equations (GEE) model, which could have been used for that purpose, would have led to the same results as the linear repeated measures model with correlated error terms. The reason for choosing the latter model is its more intuitive interpretation and the analogy with the effects models.

Intention to treat

Data analysis was performed according to the intention-to-treat principle. Data from patients who died, quit participation or were otherwise lost to follow-up were included in the analysis up to the point of drop-out.

Missing data were handled by the repeated measures models, which have the capacity to exploit the covariance structure of the existing data to adjust the results. This characteristic of the statistical models was used to achieve unbiased estimates of the treatment effect at each measurement and of the mean costs for each month during follow-up [26,27].

Cost-effectiveness and cost-utility

Health outcomes on day 7 were related to costs of the initial treatment phase; health outcomes after three months were related to total costs of the initial treatment phase and the follow-up period combined. If one of the treatment options was more effective and also more costly, results were presented in incremental cost-effectiveness ratios (ICERs): the additional cost per additional unit of health gain or the savings per unit of health loss, which was calculated as the difference in mean costs divided by the difference in mean health outcomes.

Uncertainty around the estimates of costs and health outcomes was addressed by bootstrapping the data [28]. All statistical analyses were performed on each of 1000 bootstrap replications. The mean values of incremental costs and effects from the bootstrap replications were used as the point estimates. The 95% confidence interval around the difference in mean total costs and health outcomes was determined by taking the 2.5th percentile and the 97.5th percentile of these bootstrap replications. The bootstrap replicates for the outcomes and costs after three months were plotted in cost-effectiveness planes (CE-planes). [29] The information from the CE-planes on incremental costs per QALY was summarised

in cost-effectiveness acceptability curves, which represent the likelihood that early assisted discharge is the most cost-effective option at different values of the maximum acceptable willingness to pay (WTP) for a health outcome [30].

Sensitivity analyses

Several sensitivity analyses were performed. Firstly, the costs of informal care in the follow-up period were left out of the total costs in the societal perspective (SA1). Secondly, a different unit cost per inpatient hospital day was used instead of the costs from the microcosting study (SA2). This cost was the standard unit price from the Dutch Manual for costing studies [20], which is based on a broad spectrum of diagnoses and is constant for all days during an admission.

Additionally, in order to express the uncertainty about the estimate of the costs per inpatient hospital day, sensitivity analyses were performed using the estimates of the respondent with the highest (SA3) and lowest (SA4) costs (most costly and least costly healthcare provider), and the highest (SA5) and lowest (SA6) estimates of all aspects of care across respondents (most costly and least costly scenario).

4.3 Results

Patients

From December 2007 to March 2011, 139 patients were randomised. In the usual care group, 75% of patients completed the entire trial. In the hospital-at-home group, 90% remained in the trial until the end of the follow-up period. Due to early drop-out or failure to complete questionnaires, no effectiveness data were available for 1% of patients and no cost data for 12%. The characteristics of all randomised patients are presented in table 4.2.

Table 4.2 Baseline characteristics

	Usual hospital care (N=69)	Early assisted discharge (N=70)
Age in years (SD)	67.80 (11.30)	68.31 (10.34)
Male	55.1%	68.9%
Current smoker	39.1%	32.9%
Pack years (SD)	44.52 (31.04)	46.97 (27.27)
Body Mass Index (SD)	25.57 (4.33)	24.97 (5.14)
Receiving homecare before admission	23.2%	24.3%
Charlson comorbidity score (SD)	1.68 (1.10)	1.74 (1.10)
Proportion with score > 1	39.1%	45.7%
CCQ (SD)	2.22 (0.97)	2.63 (1.03)
EQ-5D (SD)	0.71 (0.22)	0.66 (0.26)

Costs of inpatient hospital days

Table 4.3 shows that the first day of the hospital admission was the most costly. After that, the intensity of care by physicians and nurses decreased, which is reflected in lower costs per day. The total inpatient hospital costs during the 7-day treatment phase were €1430 for the usual hospital care group and €976 for the early assisted discharge group.

Table 4.3. Costs per inpatient hospital day

Day	Usual hospital care	Early assisted discharge	Difference
0 (with admission after noon)	€319	€319	
1 (with admission before noon)	€323	€323	
1 (for patients admitted on day 0)	€195	€195	
2	€192	€192	
3	€178	€178	
4	€162	€188	
5	€157	-	
6	€156	-	
7	€167	-	
Total costs for admission*	€**1430**	€**976**	€454
SA2	€3045	€1305	€1740
SA3	€1721	€1122	€599
SA4	€1228	€858	€370
SA5	€2312	€1534	€778
SA6	€952	€655	€297

* Totals are based on the assumption that 50% of patients are admitted on day 0 and 50% on day 1. This does not affect the difference between the treatment arms, because the same assumption is made for both groups.

Abbreviations: SA2 (sensitivity analysis 2): standard costs per inpatient hospital day instead of costs from microcosting study; SA3/4 cost estimates from most and least costly healthcare provider in microcosting study; SA5/6: highest and lowest estimates of care costs across respondents.

Effects

The mean improvement in CCQ scores between days 3 and 7 was larger in the hospital group than in the early assisted discharge group (-0.303 versus -0.013), but this difference was of only borderline significance (see table 4.4). Both groups showed an almost equal improvement in CCQ score between day 3 and three months.

There was a statistically significant difference between the groups in the probability of having a clinically relevant improvement in CCQ score between days 3 and 7 (51.3% in the usual hospital care group versus 31.7% in the early discharge group). It was not significant between day 3 and three months (39.9% versus 35.8%, respectively).

The difference in QALYs was very small and not statistically significant.

Table 4.4 Cost-effectiveness of early assisted discharge versus usual inpatient hospital care

	Usual hospital care	Early assisted discharge	Difference
Effects			
Mean change in CCQ score, day 7	-0.303	-0.013	0.290 (-0.03; 0.61)
Mean change in CCQ score, end of follow-up	0.024	0.065	0.041 (-0.41; 0.48)
Probability of improved CCQ score, day 7	51.3%	32.7%	-19.41% (-36.25%; -0.50%)
Probability of improved CCQ score, end of follow-up	39.9%	35.8%	-4.17% (-21.94%; 15.27%)
QALYs	0.175	0.170	-0.005 (-0.021; 0.0095)
Healthcare perspective	**Usual hospital care**	**Early assisted discharge**	**Difference**
Costs of initial episode	€1463	€1219	-€244 (-€315; -€168)
Costs of initial episode plus follow-up	€4297	€4129	-€168 (-€1253; €922)
*Incremental cost-effectiveness ratios**			
Point deterioration in mean CCQ score, day 7		€842	
Point deterioration in mean CCQ score, end of follow-up		€4098	
Additional patient without improved CCQ score, day 7		€1257	
Additional patient without improved CCQ score, end of follow-up		€4000	
Incremental QALY lost		€31,111	
Societal perspective	**Usual hospital care**	**Early assisted discharge**	**Difference**
Costs of initial episode	€1463	€1398	-€65 (-€152; €25)
Costs of initial episode plus follow-up	€5395	€6304	€880 (-€580; €2268)
*Incremental cost-effectiveness ratios**			
Point deterioration in mean CCQ score, day 7		€224	
Point deterioration in mean CCQ score, end of follow-up	Usual hospital care is dominant		
Additional patient without improved CCQ score, day 7		€335	
Additional patient without improved CCQ score, end of follow-up	Usual hospital care is dominant		
Incremental QALY lost	Usual hospital care is dominant		

* Savings per unit of health lost.

Resource use and costs

Resource use is presented in table 4.5. Table 4.6 shows that the costs for the first hospital admission were, obviously, lower in the early assisted discharge group than in the usual hospital care group. Hospital costs were reduced by €462 per patient. These savings were partly offset by the costs of community nursing care, which were €211, resulting in a net cost reduction of €244. During the follow-up phase, the early discharge group had somewhat higher costs than the usual hospital care group. In total, from a healthcare perspective, early assisted discharge led to mean costs savings of €168 (95% confidence interval (CI): -€1253; +€922) per patient.

From a societal perspective, savings in hospital costs during the 7-day treatment phase were not only offset by the costs of community nursing but also by the costs of informal care and production losses. From this perspective the initial treatment phase was only €65

Table 4.5. Resource use

	Usual hospital care	Early assisted discharge	Difference
Initial treatment phase			
Community nursing, hours	-	3.25	3.25
Informal care, hours	-	13.03	13.03
GP, home visits	-	0.015	0.015
Follow-up period			
GP, consultations	0.76	0.86	0.11
GP, home visits	0.45	0.81	0.36
GP, phone calls	0.44	0.82	0.38
Pulmonologist, consultations	1.34	1.69	0.35
Other specialist, consultations	1.27	1.24	-0.03
Paramedic care, consultations	3.88	8.88	5.01
Emergency room	0.35	0.32	-0.04
Ambulance rides	0.16	0.05	-0.11
Community nursing, hours	9.60	9.65	0.05
Domestic community care, hours	13.93	12.05	-1.87
Readmissions	0.39	0.39	0
Informal care, hours	78.50	118.97	40.47
Production losses, hours (patient)	2.38	15.56	13.18

less costly in the early discharge group. Including the costs during the follow-up phase, which were €945 higher in the early discharge group led to a total estimated cost increase of €880 (-€580; +€2268) per patient in the early discharge group, from a societal perspective. This is primarily due to the higher costs of informal care and the greater productivity loss (table 4.6).

Cost-effectiveness and cost-utility

From a healthcare perspective, all point-estimates of cost and effects pointed towards lower costs but somewhat less effects for early assisted discharge. Therefore, the ICERs represent cost savings per unit of health forgone. After seven days, the savings costs per unit of deterioration in CCQ were €842, at three months this ratio was €4098 (see table 4.4). The savings per additional patient without a clinically relevant improvement in CCQ score were €1257 after 7 days and €4000 at three months. The savings per QALY lost were €31,111. The probability that early assisted discharge was cost-saving from a healthcare perspective was 61.2%.

From the societal perspective, no ICERs were calculated for the outcomes after the follow-up period because the point-estimates of costs and effects pointed towards dominance of the usual hospital care group. The probability that early assisted discharge was cost-saving was 12% from this perspective.

Table 4.6 Treatment costs (in euros, 2009)

	Usual hospital care	Early assisted discharge	Difference
Initial treatment phase			
Inpatient days^	€1463	€1001	-€462
Community nursing	-	€211	€211
Other costs of home treatment (societal perspective)	-	€186	€186
Other costs of home treatment (healthcare perspective)	-	€6	€6
Total (healthcare perspective)	€1463	€1219	-€244 (-315;-168)
Total (societal perspective)	€1463	€1398	-€65 (-152; 25)
Follow-up period			
GP	€46	€71	€25
Pulmonologist	€86	€107	€21
Specialist	€114	€99	-€15
Paramedic care	€191	€314	€123
Emergency room	€52	€48	-€4
Ambulance	€80	€25	-€55
Medication	€346	€396	€50
Community nursing	€971	€932	-€39
Readmissions	€941	€941	€0
Informal care	€973	€1488	€515
Production loss, patient	€71	€466	€395
Total* (healthcare perspective)	€2834	€2910	€76 (-€1005; €1159)
Total* (societal perspective)	€3904	€4848	€945 (-€450; €2375)
Total study period (initial treatment phase plus follow up-period)			
Healthcare perspective*	€4297	€4129	-168 (-€1253; €922)
Societal perspective*	€5366	€6246	880 (-€580; €2268)

*Totals for follow-up period are based on regression analysis; means per cost category are not.
^These costs are higher than those in table 4.3 because the costs of prolonged hospital stay beyond 7 days (usual hospital care group) and the costs of readmission during the initial treatment phase (usual hospital care group) were included.

After seven days, the savings per unit of deterioration in CCQ were €224. The savings per patient without a clinically relevant improvement in CCQ score were €335.

There is considerable uncertainty around incremental costs and effects, as is presented in CE planes, for both perspectives (figure 4.1). From the healthcare perspective, there is a greater probability that early assisted discharge leads to net cost saving than from the societal perspective, as is shown by a greater proportion of combinations of incremental costs and effects below the x-axis. When adopting the healthcare perspective, the largest proportion of all dots was located in the southwest quadrant, with lower costs and less optimal health outcomes for early assisted discharge. From the societal perspective, the majority

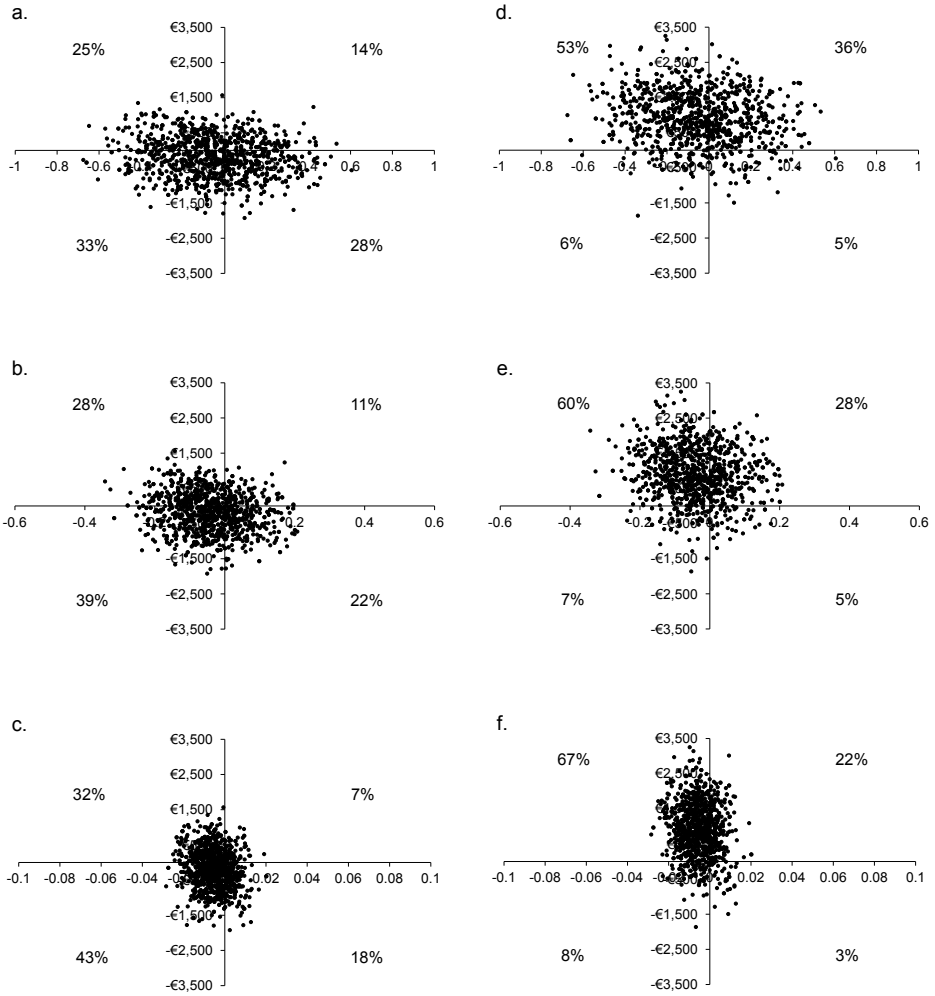

Figure 4.1: Cost-effectiveness planes.

Healthcare perspective, incremental costs set against: a) incremental improvement in CCQ-score, month 3, b) incremental proportion of patients with clinically relevant improvement, month 3; c) incremental QALYs.

Societal perspective, incremental costs set against: d) incremental improvement in CCQ-score, month 3, e) incremental proportion of patients with clinically relevant improvement, month 3; f) incremental QALYs.

of simulated outcomes were found in the northwest quadrant, with higher costs and less optimal health outcomes for early assisted discharge.

The cost-effectiveness acceptability curves in figures 4.2 and 4.3 show that, from a healthcare perspective, early assisted discharge is likely to be cost-effective for thresholds up to €46,000. From a societal perspective, early assisted discharge is unlikely to be considered cost-effective compared to usual hospital care at any threshold of maximum costs per QALY gained. In the base case, this probability is close to 10% for all thresholds.

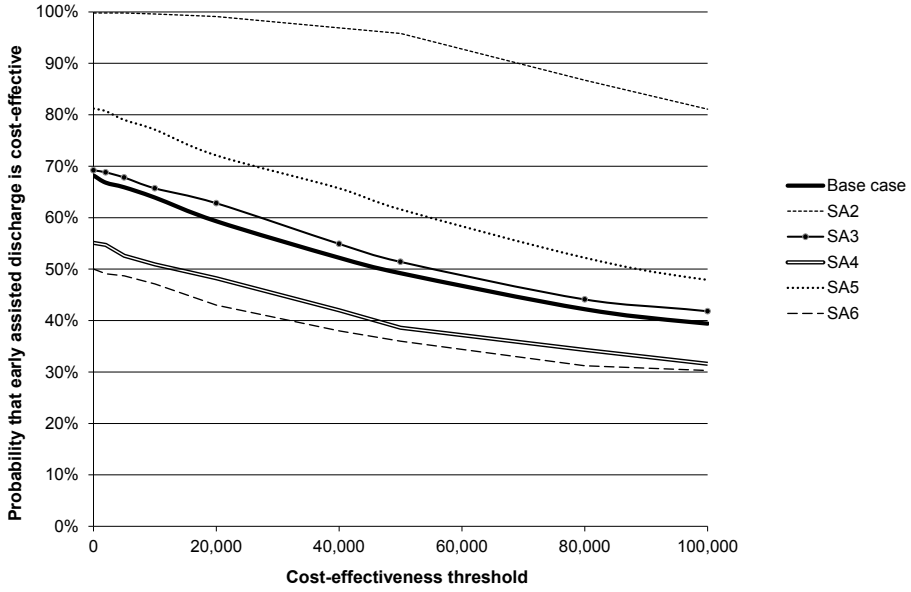

Figure 4.2 Cost-effectiveness acceptability curves for costs per QALY gained, healthcare perspective.
SA2 (sensitivity analysis 2): standard costs per inpatient hospital day instead of costs from microcosting study. SA3/4: cost estimates from most and least costly healthcare provider in microcosting study. SA5/6: highest and lowest estimates of care costs across respondents.

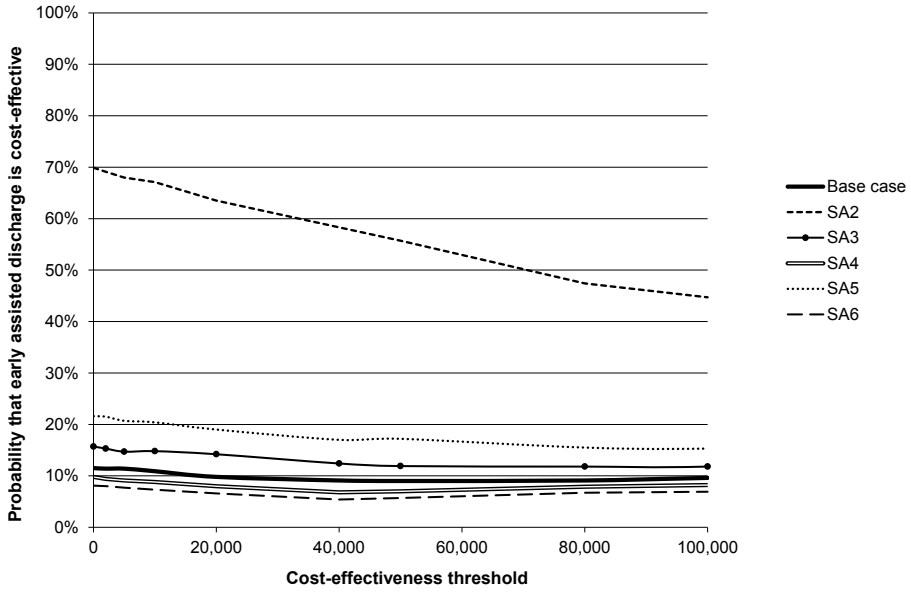

Figure 4.3 Cost-effectiveness acceptability curves, societal perspective.
SA2 (sensitivity analysis 2): standard costs per inpatient hospital day instead of costs from microcosting study. SA3/4: cost estimates from most and least costly healthcare provider in microcosting study. SA5/6: highest and lowest estimates of care costs across respondents.

Sensitivity analyses

Table 4.7 shows the results of the sensitivity analyses. The ICERs for usual hospital care compared to early assisted discharge, were sensitive to changes in the assumptions. In the initial treatment phase, early assisted discharge was almost certain to lead to cost savings from a healthcare perspective, under any of the alternative assumptions about the costs of inpatient hospital days, similar to the base case analysis. From a societal perspective, costs savings were very likely to occur during the initial treatment phase, except if mean costs per hospital day were assumed to be much lower than in the base case analysis (SA4 and SA6).

Over the entire three-month period, cost savings were more likely to occur than cost increases from a healthcare perspective. This likelihood was more or less comparable across sensitivity analyses, except when the standard unit costs for inpatient hospital days were applied (SA2). In this sensitivity analysis, the likelihood that early assisted discharge led to cost savings was 99.8%

From a societal perspective, cost savings were unlikely to occur under all assumptions except when the standard unit costs for inpatient hospital days were used (SA2). This is the only sensitivity analysis in which early assisted discharge was not dominated by usual hospital care.

The cost-effectiveness acceptability curves in figures 4.2 and 4.3 make it clear that assumptions on the costs of inpatients days do not have a strong impact on the probability that early assisted discharge is to be cost-effective.

4.4 Discussion

This study compared the costs and health effects of two treatments for patients who were admitted to the hospital with a COPD exacerbation. Patients stayed in the hospital for seven days, or went home after three days where they were supervised and treated by community nurses. No clear evidence was found to conclude that either treatment was more effective or less costly than the other.

Against the obvious savings in inpatient hospital costs, there were extra costs for community nursing, and, from a societal perspective, informal care. While costs from a societal perspective were higher among patients who were discharged early, this difference was not statistically significant. Cost savings in the healthcare perspective were not significant either. However, early discharge was much more likely to reduce healthcare costs than it was to reduce total societal costs.

At the end of the 7-day treatment phase, all outcomes measures had improved more in the patients in the usual hospital treatment group than in patients in the early discharge group. However, this difference was not statistically significant, except for the probability of having a clinically relevant improvement in CCQ-score on day 7. Patients who underwent

Table 4.7 Sensitivity analyses

	Cost difference (EAD minus HOSP)	ICER*	Probability of cost-savings for early discharge
Healthcare costs, initial episode			
Base case	-€244		100%
SA2	-€1522		100%
SA3	-€389		100%
SA4	-€160		100%
SA5	-€568		100%
SA6	-€86		99.0%
Societal costs, initial episode			
Base case	-€65		93.2%
SA1	-€65		93.2%
SA2	-€1343		100%
SA3	-€210		100%
SA4	€19		33.9%
SA5	-€389		100%
SA6	€93		2.0%
Healthcare costs, 3 months			
Base case	-€168	€31,111	61.2%
SA2	-€1464	€271,111	99.8%
SA3	-€313	€57,963	69.2%
SA4	-€84	€15,556	55.1%
SA5	-€492	€91,111	81.2%
SA6	-€10	€1,852	50.0%
Societal costs, 3 months			
Base case	€880	Dominance	11.5%
SA1	€370	Dominance	25.4%
SA2	-€416	€77,037	69.9%
SA3	€735	Dominance	15.7%
SA4	€964	Dominance	9.8%
SA5	€556	Dominance	21.6%
SA6	€1038	Dominance	8.1%

* Savings per QALY lost.

Abbreviations: EAD, early assisted discharge; HOSP, usual hospital care; ICER, incremental cost-effectiveness ratio; SA1, sensitivity analysis 1: informal care costs during follow-up not included in societal costs; SA2: standard costs per inpatient hospital day instead of costs from microcosting study; SA3/4: cost estimates from most and least costly healthcare provider in microcosting study; SA5/6: highest and lowest estimates of care costs across respondents.

usual hospital care, were more likely to experience an improvement in CCQ score of more than 0.4 points. By the end of the follow-up period, at 3 months, the difference had disappeared. In a publication of the clinical results of this study, it was reported that there was no difference in readmissions and mortality, while treatment failures were somewhat more frequent in the usual hospital care group [12].

From a societal perspective, no incremental cost-effectiveness ratios were calculated for outcomes after the follow-up period because early assisted discharge led to higher mean costs as well as less optimal health outcomes: it was dominated by usual hospital treatment. This was illustrated by the large proportion of bootstrap samples in the northwest quadrant of the CE plane and by the low acceptability curves. The verdict of dominance is often fatal for the conclusion on the treatment to which it is applied. In this case, however, it might be given less weight, since the dominance is based on a very small difference in effects. Analogously, it could be argued that the position of the majority of bootstrap replications on the CE plane should not be described as the northwestern quadrant of the CE plane, but rather as the proximity of the Y-axis and the origin.

ICERs could be calculated for outcomes after seven days from a societal perspective and for all outcome measures from a healthcare perspective. The savings per QALY lost were €31,111. This is close to the threshold values below which an ICER would generally be considered cost-effective if the new treatment were more effective than the comparator. However, in this case the threshold must represent the costs savings that would be required to make a health loss acceptable (willingness to accept). There are indications that this threshold is much higher than the threshold for the amount of incremental costs that society would be willing to pay for health gains [31]. In this light, the interpretation of the acceptability curves could also shift somewhat in favour of the comparator arm [32].

This is the first study to include the costs of informal care in the costs of the early assisted discharge scheme. The impact of this was considerable. In the 7-day treatment phase, the cost savings for early assisted discharge decreased from €244 to €65 per patient. For the full treatment period, cost savings turned into cost increases.

The costs of informal care during the follow-up phase were much higher in the early assisted discharge group. We have no good explanation for this finding. Although it might be a true difference, it is also possible that informal caregivers in the early assisted discharge group were more primed to record their activities as informal care, due to the attention it may have got during the initial treatment phase at home. For this reason, we performed a sensitivity analysis in which informal care costs during the follow-up period were excluded from the calculations. In this analysis, total costs for the early discharge group were still higher, mostly because of the higher number of patient work days lost. The difference was smaller than in the base case, as was the probability of a cost increase.

Generally, the amount of informal care can be recorded in two ways: the diary method, in which resource use is recorded on a daily basis, and the recall method, in which a

respondent is asked to provide information on the preceding week [33]. Both methods have advantages and disadvantages. Most importantly for the diary method, it may not be feasible to ask to complete it over a longer period of time. On the other hand, the recall method has been shown to have a potential to overestimate informal care time, when respondents do not take into account that they have combined certain activities with providing informal care [33]. The diary method was applied during the treatment phase in our study, while the follow-up period was covered by the recall method.

How to value informal care is still debated. Different estimation methods have led to different estimates [34]. Following Dutch guideline recommendations, we used a shadow price of €12.50 per hour, which was based on the standard tariff for the reimbursement of house cleaning costs for chronic patients. When it is applied to informal care, it reflects the assumption that informal caregivers cannot match the efficiency of professionals, who would require a higher hourly tariff. While our cost estimates are dependent on the assumed hourly unit costs of informal care, the unit costs we have used are in the center of the range of costs (€7 to €17) that were estimated by Koopmanschap et al. using different valuation methods [34].

Like the costs of informal care, the productivity losses in the early assisted discharge group were higher as well. This was mostly due to one patient, who incurred a very high amount of costs.

In our study, the duration of hospital or home treatment was fixed. Whenever possible, patients were discharged or homecare was stopped after seven days. It is conceivable that the threshold for adding another day of treatment may be lower for treatment at home. In other studies, in which no fixed treatment duration was used and physicians were fully free to decide on the duration of treatment, different durations were observed for each treatment group and the total duration of treatment in hospital-at-home was longer than that in usual hospital care. Such an approach may have commingled the effects of the treatment per se with the effects of the length of stay or even with the timing of health measurements. Treating patients for a longer time may lead to better health, but measuring their health at a later time may also lead to seemingly better results. Our design made it possible to make the comparison exclusively on the basis of where and by whom treatment was provided. Four patients in the usual hospital care group remained in the hospital for a longer period of time. One patient who was discharged early needed to be re-admitted within the seven days of initial treatment. The additional costs of these patients were included in the costs of the initial treatment phase. No patient required homecare beyond seven days.

It is possible – although far from certain – that the early assisted discharge treatment would in daily practice be longer than the hospital treatment. This would clearly lead to higher costs than in this trial, whereas the study did not yield indications that it would or would not improve health outcomes.

The patients in the trial can be considered representative for other patients who would be eligible and willing to participate. Almost two thirds of screened patients were too ill, did not have an obvious informal caregiver or did not live in the catchment area of the community nursing organization. This may reduce the potential for the early assisted discharge treatment, but it is still considerable, given the size of the patient population.

This study has shown the potential impact of a detailed unit cost calculation of an inpatient hospital day based on treatment-intensity compared to standard tariffs or references prices. A sensitivity analysis using Dutch reference costs, which represent average costs of a hospital day based on all patients irrespective of their diseases [20], led to much larger savings for early assisted discharge. From the societal perspective, the cost increase due to early assisted discharge disappeared almost entirely. From the healthcare perspective, the finding that early assisted discharge led to cost savings was surrounded by almost no uncertainty.

However, using standard costs of a hospital day would not be opportune in this study because only the least-costly inpatient days were substituted by home care. Furthermore, hospital care for patients with COPD exacerbations that meet the in- and exclusion criteria of our trial are likely to be less intensive than the hospital care for the average admitted patient.

The calculations for the costs per inpatient day were not based on a large sample of patients, whose treatment was actually timed and recorded. This was not feasible in this study, due to the unpredictability of hospital admissions, the relatively small number of patients eligible for the study, and the large number of treatment aspects that would have to be recorded. It would have required researchers permanently present in the hospitals for a long period of time. Instead, we interviewed hospital care providers with much experience in treating this patient group. A standardised questionnaire was used, in which all aspects of care on a particular day were distinguished. Respondents were not asked to estimate the total amount of time they spent on each patient but on an average patient. Tan et al. concluded that this method leads to a good balance between feasibility and reliability [22]. A problem with this method is, however, that it does not yield measures of variability on a patient level. This means that the uncertainty about the costs of treatment in the hospital, which inevitably exists, was not represented in the uncertainty around the total costs of treatment. While this may always be the case when fixed unit costs are used, inpatient hospital days are different. They contain a large number of separate elements – not just capital costs, hotel services and overhead costs, which could be fairly similar for all patients, but also time from several healthcare providers for many different aspects of care. It is conceivable that the price of a general practitioner consultation does not differ much across patients because all more or less take the same amount of time, whereas inpatient hospital days are much more different for different patients. This may not be a problem when hospital costs are merely one a relatively infrequent element in the total costs of care, but in the initial treatment

episode for COPD exacerbations the costs of inpatient hospital days are virtually the only cost driver. Therefore, we performed sensitivity analyses assuming different unit cost prices per inpatient hospital day. These gave an indication about the range of possible costs savings and increases.

Most of the other cost studies of hospital-at-home found larger cost savings than we did, also from a healthcare perspective [3,9-14]. This may be explained by the design of our study – early assisted discharge, not admission avoidance, which brought a reduction of four inpatient hospital days. In some other studies, this reduction was larger. In England, two economic evaluations were performed. Skwarska et al. calculated savings of £876 per patient by eliminating five inpatient days (median) in an admission avoidance program (no statistical testing was done) [10]. In contrast, Shepperd et al. found significant cost increases (difference between medians £1176) for an early assisted discharge scheme, in which five inpatient days were substituted for care at home as well [9]. The cost increases in this study were mostly due to the large proportion of patients who were readmitted to the hospital after having been discharged early compared to usual treatment, which makes it plausible that the health effects of treatment were better in the usual hospital care group. Although differences were not statistically significant in their small sample, almost all health indicators were in favor of usual hospital care. In a Spanish study, significant savings of around €800 were reported, the exact amount depending on the analysis [11,12]. In this scheme an average of 3.8 inpatient hospital days was substituted by 1.7 home visits and 2.3 phone calls per patient. Some patients were discharged early, while others avoided admission completely. In this study, even some patients who were randomised to the usual hospital treatment, did not spend a night in the hospital. In an Australian admission avoidance study, in which community nurses were employed instead of hospital-based staff, the savings were AUS$1696 [3]. Aimonino Ricauda et al. examined an admission avoidance program in Italy [14]. The cost difference of US$215 was not significant. However, the hospital-at-home scheme contained visits by physicians and a transport home by ambulance for all patients, which made the cost difference smaller. In a non-randomised study in the United States – all previously mentioned studies were randomised – Frick et al. found the largest savings, US$2314 per patient [13].

In conclusion, transferring hospital care for a COPD exacerbation to the patient's home is likely to lead to modest savings in healthcare costs in the Netherlands, while there is no evidence that it would be medically impossible or irresponsible for selected patients. When the societal costs of informal care and productivity losses are taken into account, the cost savings decrease considerably or even turn into cost increases.

Since there is no compelling reason – from a medical or economic point of view – to recommend either the early supported discharge treatment or usual hospital care, patients' preferences should play an important role in deciding where (s)he is treated. If homecare is preferred, this study has shown that the wide-spread network of homecare organizations in the Netherlands, which employs community nurses, is able to meet this preference.

4.5 References

1 Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. 2004; 23: 932-946.

2 Dal Negro RW, Tognella S, Tosatto R, et al. Costs of chronic obstructive pulmonary disease (COPD) in Italy: the SIRIO study (social impact of respiratory integrated outcomes). Respir Med. 2008; 102: 92-101.

3 Nicholson C, Bowler S, Jackson C, et al. Cost comparison of hospital- and home-based treatment models for acute chronic obstructive pulmonary disease. Aust Health Rev. 2001; 24: 181-187.

4 Nielsen R, Johannessen A, Benediktsdottir B, et al. Present and future costs of COPD in Iceland and Norway: results from the BOLD study. Eur Respir J. 2009; 34: 850-857.

5 Toy EL, Gallagher KF, Stanley EL, et al. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD. 2010; 7: 214-228.

6 Jordan RE, Hawker JI, Ayres JG, et al. Effect of social factors on winter hospital admission for respiratory disease: a case-control study of older people in the UK. Br J Gen Pract. 2008; 58: 400-402.

7 Shepperd S, Doll H, Angus RM, et al. Admission avoidance hospital at home. Cochrane Database Syst Rev. 2008; (4): CD007491.

8 Shepperd S, Doll H, Broad J, et al. Early discharge hospital at home. Cochrane Database Syst Rev. 2009; (1): CD000356.

9 Shepperd S, Harwood D, Gray A, et al. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. BMJ. 1998; 316: 1791-1796.

10 Skwarska E, Cohen G, Skwarski KM, et al. Randomized controlled trial of supported discharge in patients with exacerbations of chronic obstructive pulmonary disease. Thorax. 2000; 55: 907-912.

11 Hernandez C, Casas A, Escarrabill J, et al. Home hospitalisation of exacerbated chronic obstructive pulmonary disease patients. Eur Respir J. 2003; 21: 58-67.

12 Puig-Junoy J, Casas A, Font-Planells J, et al. The impact of home hospitalization on healthcare costs of exacerbations in COPD patients. Eur J Health Econ. 2007; 8: 325-332.

13 Frick KD, Burton LC, Clark R, et al. Substitutive Hospital at Home for older persons: effects on costs. Am J Manag Care. 2009; 15: 49-56.

14 Aimonino Ricauda N, Tibaldi V, Leff B, et al. Substitutive «hospital at home» versus inpatient care for elderly patients with exacerbations of chronic obstructive pulmonary disease: a prospective randomized, controlled trial. J Am Geriatr Soc. 2008; 56: 493-500.

15 Utens CM, Goossens LMA, Smeenk FWJM, et al. Early assisted discharge with communitey nursing for chronic obstructive pulmonary disease exacerbations: results of a randomised controlled trial. BMJ Open. 2012; 2: e001684. doi:10.1136/bmjopen-2012- 001684.

16 Utens CM, Goossens LM, Smeenk FW, et al. Effectiveness and cost-effectiveness of early assisted discharge for chronic obstructive pulmonary disease exacerbations: the design of a randomised controlled trial. BMC Public Health. 2010; 10: 618.

17 Kocks JW, Tuinenga MG, Uil SM, et al. Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. Respir Res. 2006; 7: 62.

18 Lamers LM, McDonnell J, Stalmeier PF, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006; 15: 1121-1132.

19 College voor zorgverzekeringen (Health Care Insurance Board). Guidelines for pharmacoeconomic research, updated version. Diemen, 2006.

20 Hakkaart-van Roijen L, Tan SS, Bouwmans CAM. Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg (Manual for costing studies. Methods and standard unit costs for economic evaluations in healthcare). 2010.
21 College voor zorgverzekeringen (Health Care Insurance Board). Medicijnkosten (Medication costs). www.medcijnkosten.nl 2011
22 Tan SS, Rutten FF, van Ineveld BM, et al. Comparing methodologies for the cost estimation of hospital services. Eur J Health Econ. 2009; 10: 39-45.
23 Tan SS, Van Gils CW, Franken MG, et al. The unit costs of inpatient hospital days, outpatient visits, and daycare treatments in the fields of oncology and hematology. Value Health. 2010; 13: 712-719.
24 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology (3 ed.). Philadelphia: Lippincott Williams & Wilkins, 2008.
25 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40: 373-83.
26 Molenberghs G, Verbeke G. Models for discrete longitudinal data . New York: Springer, 2007.
27 Verbeke G, Molenberghs G. Linear mixed models for longitudinal data (2nd ed.). New York: Springer, 2000.
28 Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ. 1997; 6: 327-340.
29 Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ. 1998; 7: 723-740.
30 van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ. 1994; 3: 309-319.
31 O'Brien BJ, Gertsen K, Willan AR, Faulkner LA. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Econ. 2002; 11:175-80.
32 Severens JL, Brunenberg DE, Fenwick EA, O'Brien B, Joore MA. Cost-effectiveness acceptability curves and a reluctance to lose. Pharmacoeconomics. 2005; 23:1207-14.
33 van den Berg B, Spauwen P. Measurement of informal care: an empirical study into the valid measurement of time spent on informal caregiving. Health Econ. 2006; 15: 447-460.
34 Koopmanschap MA, van Exel JN, van den Berg B, Brouwer WB. An overview of methods and applications to value informal care in economic evaluations of healthcare. Pharmacoeconomics. 2008; 26: 269-280.

Should I stay or should I go home?

A latent class analysis of a discrete choice experiment on hospital-at-home

5

Background This study aimed (1) to quantify patient preferences for different aspects of hospital-at-home in the Netherlands for patients who were admitted with a COPD exacerbation (2) to provide an application of latent class modelling of discrete choice data. This technique is rarely used in health economics.

Methods In a discrete-choice experiment, respondents were asked to make multiple choices between hospital treatment as usual (7 days) and two combinations of hospital admission (3 days) followed by treatment at home. The latter was described by a set of attributes: training of homecare nurses, number of different nurses involved in home visits, number of daily visits day, co-payments, readmission risk, whom to contact in case of worsening disease, and informal caregiver burden. Hospital treatment was constant across choice sets. Respondents were COPD patients in a randomised controlled trial investigating the cost-effectiveness of early assisted discharge, and their informal caregivers. The data were analysed in latent-class conditional logit regression, which allowed for heterogeneous preferences across groups.

Results 25% of respondents opted for hospital treatment regardless of the description of the hospital-at-home program, 46% never chose for the hospital. The best model contained four latent classes of respondents, defined by different preferences for the hospital and for the caregiver burden. Preferences for the other attributes were the same across classes. Attributes with the strongest influence on choices were the burden on informal caregivers and co-payments. Except for the number of visits, all attributes had a significant effect on choices in the expected direction.

Conclusion Considerable segments of respondents had fixed preferences for either treatment option. Applying latent class analysis was essential in quantifying preferences for attributes of hospital-at-home.

Lucas M.A. Goossens, Cecile M.A. Utens, Frank W.J.M. Smeenk, Bas Donkers, Onno C.P. van Schayck, Maureen P.M.H. Rutten-van Mölken

Submitted for publication.

5.1 Introduction

Many patients with chronic obstructive pulmonary disease (COPD) are more or less frequently admitted to the hospital for an exacerbation of their disease. The average annual frequencies have been estimated to vary from 0.11 for patients with mild COPD (GOLD stage I, as defined by lung function [1]) and 0.16 for moderate disease (GOLD II), to 0.22 and 0.28 for severe and very severe COPD (GOLD III and IV)[2]. Nevertheless, the extent to which patients are prone to exacerbations varies substantially within GOLD stages [3].

Hospitalisations for exacerbations are the main drivers of COPD treatment [4-9]. They put pressure on scarce hospital beds of respiratory wards, especially during winter months [10]. However, COPD patients are vulnerable to infections in a hospital environment. They may prefer to be in the hospital for as short a period as possible for reasons of privacy and comfort. It may therefore be attractive to treat suitable patients at home instead of in the hospital, if this is medically possible. This approach is often called hospital-at-home. It can either substitute the entire hospital admission for home treatment (admission avoidance) or the last days of the admission (early assisted discharge) [11,12].

The GO AHEAD trial, which compared early assisted discharge with a conventional hospital admission, did not lead to the conclusion that either treatment was clearly preferable from a medical or economic point of view [13,14]. No clear and significant differences were found in health outcomes or costs, although early assisted was more likely to be the less costly alternative from the healthcare perspective. This lack of clear superiority of either treatment increases the importance of the preferences of patients and their informal caregivers. Adapting a treatment program to their preferences may enhance its acceptability.

The first objective of this chapter was to quantify patients' and informal caregivers' preferences for different characteristics of an early assisted discharge scheme in the Netherlands and to determine when these characteristics make the new scheme more attractive than usual hospital care.

A commonly used technique for eliciting preferences is the discrete choice experiment (DCE), in which respondents are asked to choose between alternatives, which are described by a number of attributes [15]. Statistical analysis is then used to quantify the weight of each attribute in the choices of the respondents. In health economics, one of the most widely applied methods to analyse data from such experiments is McFadden's conditional logit, otherwise known as multinomial logit [16-18]. However, one of the asssumptions of this technique is homogeneity of preference across respondents [19]. When this assumption is violated – in other words, when some respondents have consistently different preferences than others – the model may lead to biased results. The most popular methods to take preference heterogeneity into account are based on random parameters: the random effect probit and logit models for binary choices and the mixed logit model for choices with more than two alternatives [16].

An alternative method, which is particularly suited for segmented samples of respondents, is latent class analysis. Unfortunately, this technique is rarely used in health economics. In a review of DCE methods in this field, Bekker-de Grob et al. found that latent class analysis, was applied only once in the period from 1990 to 2008 [16], in a study on appointments with general practitioners [20]. To our knowledge, the only more recent example of latent class analysis in health economics is a study on preventive treatment of latent tuberculosis [21].

Latent class analysis groups respondents in a pre-specified number of latent classes with distinct preferences. This allows for the estimation of class-specific preference parameters and of the probability of class membership. One of the developers of DCE methodology, Louviere, has argued for a more frequent use of latent class models [22] because they fit the data at least as well as random parameter models while estimation and interpretation are easier.

An additional objective of this chapter was therefore to provide a new example of the application of latent class modelling of discrete choice data by investigating whether there were subgroups with clearly different preferences.

5.2 Methods

Selection of attributes

A literature search led to a selection of characteristics of hospital-at-home treatments for COPD. These were considered potential attributes for the DCE. The attributes had to describe the process, not the outcomes of treatment. The provisional attributes were discussed with physicians connected to the trial and with COPD patients who were admitted to the hospital. They were invited to mention additional attributes and levels. Non-nominal attribute levels were chosen in order to reflect a wide range of possibilities and being able to have an impact on choices, without becoming unrealistic or unimaginable to respondents.

The final questionnaires contained the following attributes for hospital-at-home treatment: (1) specialisation of the community nurse, (2) number of home visits, (3) number of different nurses involved in the treatment, (4) co-payment, (5) whom to contact in case of worsening disease, (6) burden on informal caregivers, (7) risk of readmission to the hospital before the scheduled end of home treatment. Table 5.1 shows the levels of each attribute.

Design of DCE questionnaire

Choice sets consisted of three labelled alternatives: two early-assisted-discharge treatments and the usual hospital treatment (see figure 5.1 for an example). Since many characteristics of hospital-at-home are not applicable to usual hospital treatment and vice versa, only the hospital-at-home treatments were described by attributes. Because all respondents were

Table 5.1 Attributes and levels for early assisted discharge options

Treatment attribute	Levels
Specialisation of community nurse	Generic Pulmonary
Number of home visits	1 per day 2 per day 3 per day
Number of nurses involved in treatment	1 or 2 More than 2
Co-payment	€0 €50 €100
Contact	GP Pulmonary ward, hospital
Burden on informal caregivers	1 hour per day 3 hours per day 5 hours per day
Risk of readmission	1% 5% 10%

hospitalised they were assumed to be familiar with hospital treatment, which was constant over all choice sets.

In order to extract as much choice information as possible, respondents who preferred the hospital option in a certain choice set were subsequently asked which of the hospital-at-home options they preferred.

No opt-out was presented, since all COPD patients who are admitted to the hospital for an exacerbation cannot be left untreated. Respondents were asked to assume that all treatments were equally effective in medical terms, i.e. after seven days, a patient's health state would be the same under all treatment options.

SAS 9.1 software was used to generate a d-efficient fractional factorial design for the questionnaire, which consisted of 36 choice sets divided into three versions. Each questionnaire contained 12 choice sets, to which we added two fixed choice sets with a dominant alternative, i.e. an alternative that is better on all attributes, in order to test the respondents' comprehension of the task. Choice sets were presented in random order.

Respondents

The questionnaires were presented to all COPD patients and their informal caregivers who participated in the GO AHEAD trial, which was carried out in five hospitals in the Netherlands from November 2007 to March 2011. In the early assisted discharge arm of this randomised trial, patients spent three days in the hospital, after which they were treated in their own homes by community nurses for four more days. Patients in the control group remained in the hospital for seven days. Participants had diagnosed COPD, were ≥40 years

	Early assisted discharge A	Early assisted discharge B	Hospital
Nurse specialisation	Generic nurse	Pulmonary nurse	
Number of nurse visits	3 per day	1 per day	
Co-payment	50 euros	100 euros	
Re-admission risk	1 in 10	1 in 20	
Whom to contact in case of worsening disease	Hospital, pulmonary ward	General practitioner	
Informal care burden	3 hours per day	1 hour per day	
Number of different nurses	1 or 2	More than 2	
Which treatment would you choose? (Tick 1 box.)	**A** ☐	**B** ☐	**Hospital** ☐
Which treatment would you choose if you can only opt for early assisted discharge? (Tick 1 box.)	**A** ☐	**B** ☐	

Figure 5.1 Example of choice set from questionnaire.

old, had no major uncontrolled co-morbidities, and had no indication for admission to an intensive care unit or for non-invasive ventilation. After three days in the hospital, they had to be clinically stable in order to be randomised.

Since trial-participants were more likely to have a preference for hospital-at-home than the general patient population additional respondents were recruited among patients who were ineligible for inclusion or who did not consent.

Each respondent was asked to fill out the questionnaire during an outpatient visit to the hospital one month after the initial admission. If patients did not appear at the appointment, questionnaires were sent to home addresses. Ethics approval was obtained from the ethics board of Catharina Hospital in Eindhoven, the Netherlands.

Statistical analysis

First, we investigated how many respondents were principally willing to consider the hospital alternative as well a hospital-at-home alternative or whether they had a fixed preference, irrespective of the attribute levels of the hospital-at-home alternatives. This was done by examining the initial answers of all choice sets for each respondent.

Next, all choices – the initial answer and the possible second answer – were analysed in a multinomial logit model (MNL), otherwise known as conditional logit. All attribute levels

were dummy-coded. Constants were added for the hospital alternative, in order to detect a preference or aversion to this treatment relative to hospital-at-home, and for one of the hospital-at-home-options, in order to test a possible ordering effect (i.e. whether respondents were more likely to choose the option that was presented first or second).

The MNL model assumes that all choices are made independently, which is not realistic when respondents have systematically different preferences than others. To express this potential preference heterogeneity, a latent class model (LC) was developed. A LC model fits the best possible model with a pre-determined number of classes. For each class, different coefficients (or discrete random effects) are estimated for one or more attributes.

The optimal number of classes was determined in an iterative procedure, by making comparisons of models with different numbers of classes, based on the Akaike Information Criterion 3 (AIC) and supported by the AIC. The AIC3 is more critical towards models with more parameters than the AIC. It imposes a penalty of 3 instead of 2 points per model parameter. According to Andrews and Currim, the AIC3 is the best performing criterion in determining the optimal number of classes in logit models [23].

First, a preliminary number of classes was determined by comparing models with the hospital constant as the sole random effect. In the second step, the found number of classes was used in models with different random effects, in addition to the hospital constant. This was done by examining the p-values for the discrete random effects, which were entered into the model one by one. In these analyses, the random effects were expressed as deviations from the mean value of the coefficient for the attribute level over all classes. In the third step, the selected random effects were applied in models with different numbers of classes. Finally, the LC model was estimated without mean coefficients for attributes with discrete random effects, i.e. each random effect was expressed as a deviation from zero.

The association of class membership with treatment group and COPD severity stage defined by GOLD [1] was explored in cross tables and tested using chi-squared tests. Respondents were assumed to belong to the class with the highest predicted probability.

All analyses were performed in Stata 12.0 (Statacorp, College Station, TX), using the GLLAMM procedure [24,25].

5.3 Results

From the GO AHEAD trial, 107 patients and 83 informal caregivers completed the questionnaire. Additionally, 7 patients and 6 informal caregivers returned the questionnaire. The response rates among GO AHEAD participants were 77% for patients and 64% for informal caregivers.

Table 5.2 shows that a quarter of respondents chose the hospital care option in all 14 choice sets, whereas approximately 45% always chose the hospital-at-home option.

Table 5.2 Patterns of chosen alternatives

Choice pattern	Patients (n=113)	Informal caregivers (n=89)
Always usual hospital care	28 (25%)	23 (26%)
Both	33 (29%)	26 (29%)
Always early assisted discharge	52 (46%)	40 (45%)

The preference of the remaining 29% of respondents depended on the description of the hospital-at-home treatment.

Multinomial logit

Table 5.3 shows the results of the multinomial logit model. Results for patients closely resembled those for informal caregivers. The ordering of the alternatives played no role, which

Table 5.3. Results of multinomial logit analyses

	Patients			Informal caregivers		
	Coefficient	P-value	Rank	Coefficient	P-value	Rank
Constant early assisted discharge A	Reference category			Reference category		
Constant early assisted discharge B	0.021	0.724	13	-0.085	0.206	12
Constant hospital	-0.681	<0.0005	1	-0.638	<0.0005	1
Generic nurse	Reference category			Reference category		
Pulmonary nurse	0.371	<0.0005	5	0.374	<0.0005	5
1 or 2 nurses	Reference category			Reference category		
More nurses	-0.351	<0.0005	7	-0.508	<0.0005	3
1 visit per day	Reference category			Reference category		
2 visits per day	0.032	0.751	12	-0.222	0.055	8
3 visits per day	-0.155	0.120	10	-0.258	0.017	7
Co-payment €0	Reference category			Reference category		
Co-payment €50	-0.311	<0.0005	8	-0.134	0.137	11
Co-payment €100	-0.697	<0.0005	2	-0.485	<0.0005	4
Readmission risk 1%	Reference category			Reference category		
Readmission risk 5%	-0.077	0.352	11	-0.046	0.624	13
Readmission risk 10%	-0.189	0.018	9	-0.159	0.081	10
Contact: pulmonary ward	Reference category			Reference category		
Contact: general practitioner	-0.375	<0.0005	4	-0.301	<0.0005	6
Informal carer burden : 1 hour per day	Reference category			Reference category		
Informal carer burden : 3 hours per day	-0.366	<0.0005	6	-0.192	0.024	9
Informal carer burden : 5 hours per day	-0.456	<0.0005	3	-0.508	<0.0005	2
AIC	3894.824			2991.409		
AIC3	3907.824			3004.409		

can be concluded from the small and non-significant coefficient for alternative B. The negative coefficient for the hospital option pointed to a preference for hospital-at-home, given the baseline levels of the other attributes. All attributes had a significant impact on patients' and informal caregivers' choices in the expected direction, except for the number of home visits per day, among patients, and the readmission risk, among informal caregivers, which were not statistically significant.

The rankings show which attribute levels had the strongest impact on choices. For both groups of respondents, the hospital coefficient was the largest coefficient.

Development of latent class multinomial logit models

For patients, the model with four classes was the most appropriate of all models with one random effect (hospital constant), based on the AIC3. The AIC supported this conclusion. This is shown in table 5.4. For informal caregivers, the models with three and four classes appeared to be equally appropriate.

In the four class model, only random effects for hospital and the two levels of informal caregiver burden significantly deviated from the mean effect. When these random effects were applied, the AIC3 pointed to models with four classes for patients as well as caregivers. According to AIC, a five-class model for patients had a slightly better fit. Based on both, models with the three random effects were more appropriate than models with one random effect.

Table 5.4 Goodness of fit for latent class models

Patients			**Informal caregivers**	
	AIC*	AIC3*	AIC*	AIC3*
Models with 1 discrete random effect (hospital)				
1 class	3894.824	3907.824	2991.409	3004.409
2 classes	2674.043	2689.043	2171.029	2186.029
3 classes	2574.699	2591.699	2003.239	2020.239
4 classes	2549.744	2568.744	2001.269	2020.269
5 classes	2553.744	2574.744	2005.269	2026.259
Models with 3 discrete random effects (hospital, informal carer burden 3h/5h)				
2 classes	2677.899	2694.899	2168.506	2185.506
3 classes	2575.479	2596.479	1999.387	2020.387
4 classes	2536.432	2561.432	1991.352	2016.352
5 classes	2533.291	2562.291	Did not converge.	

*Lower value indicates a better fit.

Final latent class model

The final models for both patients and informal caregivers contained four classes with three discrete random effects: hospital, burden on informal caregivers 3 hours and 5 hours per day. The results from these models are shown in table 5.5.

The classes for patients have somewhat different characteristics than the classes for informal caregivers. For patients, class 1 consisted of respondents with a strong preference for the hospital and with a moderate aversion to the higher levels of burden on the informal caregiver. Class 2 contained respondents with a moderate aversion to the hospital and moderate aversion to the highest level of burden to the informal caregiver. In class 3, respondents had a strong aversion to the hospital and to the higher levels of burden on the informal caregiver. Class 4 was formed by patients with the strongest aversion to the hospital, who were indifferent about the burden on the informal caregiver. The informal caregivers in class 1 had a strong preference for the hospital and were indifferent about the burden on informal caregivers. Respondents in group 2 had no significant preferences for the hospital or the level of burden. Class 3 consisted of respondents with a moderate aversion to the hospital and to higher levels of burden. Class 4 was formed by respondents with the strongest aversion to the hospital and a moderate aversion to the highest level of informal caregiver burden.

For the other attributes, strengths of preferences were equal across the classes. For both groups of respondents, all attributes had a significant effect on choices, in the expected di-

Table 5.5 Results of final latent class conditional logit analyses

	Patients			**Informal caregivers**		
	Coefficient	P-value	Rank	Coefficient	P-value	Rank
Outcomes per class						
Class 1						
Hospital	3.30	<0.0005		4.642	<0.0005	
Informal carer burden : 1 hour per day	Reference category			Reference category		
Informal carer burden : 3 hours per day	-0.324	0.033		-0.047	0.790	
Informal carer burden : 5 hours per day	-0.610	<0.0005		-0.216	0.264	
Class 2						
Hospital	-0.984	<0.0005		0.902	0.083	
Informal carer burden : 1 hour per day	Reference category			Reference category		
Informal carer burden : 3 hours per day	0.070	0.732		0.306	0.485	
Informal carer burden : 5 hours per day	-0.478	<0.043		0.648	0.225	

Table 5.5 Results of final latent class conditional logit analyses (continued)

	Patients			**Informal caregivers**		
	Coefficient	P-value	Rank	Coefficient	P-value	Rank
Class 3						
Hospital	-4.749	<0.0005		-1.275	<0.0005	
Informal carer burden : 1 hour per day	Reference category			Reference category		
Informal carer burden : 3 hours per day	-1.740	<0.0005		-0.691	0.002	
Informal carer burden : 5 hours per day	-3.513	<0.0005		-1.651	<0.0005	
Class 4						
Hospital	-5.621	<0.0005		-5.448	<0.0005	
Informal carer burden : 1 hour per day	Reference category			Reference category		
Informal carer burden : 3 hours per day	-0.089	0.466		-0.176	0.159	
Informal carer burden : 5 hours per day	-0.106	0.476		-0.472	0.001	
Shared outcomes for all classes						
Early assisted discharge A	Reference category			Reference category		
Early assisted discharge B	-0.00028	1	10	-0.082	0.239	10
Generic nurse	Reference category			Reference category		
Pulmonary nurse	0.516	<0.0005	3	0.485	<0.0005	4
1 or 2 nurses	Reference category			Reference category		
More nurses	-0.433	<0.0005	5	-0.536	<0.0005	2
1 visit per day	Reference category			Reference category		
2 visits per day	-0.016	0.902	9	-0.168	0.245	8
3 visits per day	-0.062	0.621	8	-0.107	0.420	9
Co-payment €0	Reference category			Reference category		
Co-payment €50	-0.478	<0.0005	4	-0.202	0.042	7
Co-payment €100	-1.107	<0.0005	1	-0.722	<0.0005	1
Readmission risk 1%	Reference category			Reference category		
Readmission risk 5%	-0.241	0.015	7	-0.207	0.057	6
Readmission risk 10%	-0.409	<0.0005	6	-0.307	0.003	5
Contact: pulmonary ward	Reference category			Reference category		
Contact: general practitioner	-0.631	<0.0005	2	-0.510	<0.0005	3
AIC	2536.432			1991.352		
AIC3	2561.432			2016.352		

rection, except for the number of visits by homecare nurses per day. The highest co-payment had the largest coefficient for patients as well as informal caregivers. Also important was the question whom to contact in case of an emergency – with a preference for the pulmonary ward in the hospital – the pulmonary specialization of the nurses and a small number of different nurses who were involved in treatment at home.

Latent class membership

Predicted probabilities of the designated classes were all over 96%, which means that there is hardly any uncertainty with regards to the classes that respondents belonged to. The great majority of patients as well as informal caregivers were in the extreme classes, with the strongest aversion to or preference for the hospital treatment (see table 5.6). For patients, class membership was significantly associated with the treatment they had actually received in the clinical trial (chi-square, p=0.039). This association was not significant for informal caregivers (p=0.648). Respondents with experience with home treatment less likely to have a preference for the hospital. Table 5.6 also shows that informal caregivers in both treatment groups were somewhat more likely to be in a class with a preference for the hospital than patients were.

Table 5.7 presents the distribution of patients in different COPD severity stages, and their informal caregivers, over the classes. Patients in stage IV were relatively likely to be in class 1, with the strongest preference for the hospital and less likely to be in class 4. This association was not statistically significant (p=0.340). In contrast, informal caregivers of patients in GOLD stage IV were relatively more likely to have a preference for treatment at home (p=0.160).

Table 5.6 Distribution of patients and informal caregivers from each treatment group over latent classes

Class	Attitude towards hospital/caregiver burden	Treatment group	
		Hospital	Early assisted discharge
Patients			
1	Strong preference/moderate aversion	38.8%	19.3%
2	Moderate aversion/mild aversion	10.6%	22.8%
3	Strong aversion/strong aversion	12.8%	5.2%
4	Very strong aversion/neutral	38.8%	52.6%
	Total	100%	100%
Informal caregivers			
1	Strong preference/neutral	31.6%	23.4%
2	Moderate preference/neutral	5.26%	6.4%
3	Moderate aversion/moderate aversion	18.4%	12.8%
4	Very strong aversion/mild aversion	44.8%	57.5%
	Total	100%	100%

Table 5.7 Associaton of GOLD stage and DCE class

	Patients			Informal caregivers		
	GOLD 2	GOLD 3	GOLD 4	GOLD 2	GOLD 3	GOLD 4
Class 1*	36.1%	17.1%	42.1%	39.4%	16.1%	26.7%
Class 2*	11.1%	24.4%	15.8%	3.0%	9.7%	6.7%
Class 3	8.3%	12.2%	5.3%	9.1%	29.0%	6.7%
Class 4	44.4%	46.3%	36.8%	48.5%	45.2%	60.0%
Totaal	100%	100%	100%	100%	100%	100%

chi^2 test patients: p=0.340. Informal caregivers: p=0.161.

* For patients, class 1 represents a preference for the hospital. For in informal caregivers, this is the case for classes 1 and 2.

5.4 Discussion

This study used latent class conditional logit models to quantify the preferences of patients and informal caregivers for aspects of early assisted discharge after a hospital admission for a COPD exacerbation.

The results have shown that the average patient and the average informal caregiver do not exist. For both groups, four distinct classes were distinguished, which had different attitudes towards being treated at home or in the hospital. Large proportions of respondents had a preference for either treatment option that could not be influenced by proposing realistic changes in the characteristics of the early assisted discharge treatment. When choosing between two home options – with the hospital option removed from the choice set – copayments and the burden on the informal caregiver had the strongest impact on choices. The number of visits per day did not play a role.

Results from this study could be used in the design of hospital-at-home programs for this category of patients. It did not appear to be required that nurses visit patients more than once a day. The frequency could depend on medical need, unless a higher frequency could lessen the burden on the informal caregiver, which did play a role in respondents' choices.

The most attractive hospital-at-home program would be operated by pulmonary nurses and have no co-payments. Patients would not be visited by more than two different nurses and the pulmonary ward of the hospital would be reachable 24-hours a day in case of sudden worsening of the disease. Patients and informal caregivers appeared to be in agreement on this.

The attractiveness of this combination appears to be consistent across the different classes of respondents. However, this sample is not representative of the population of patients in similar health states and their caregivers. Almost half of the respondents always opted for early assisted discharge. The proportion of people in the overall population is likely to be smaller, since most of the respondents participated in the GO AHEAD trial, in which early assisted discharge was compared to regular hospital treatment. It is obvious that most

of them did not have strong reservations against being treated at home. Patients who wanted to be treated in the hospital, could achieve this by not participating in the trial.

Because of the selection of respondents, this study could not estimate the sizes of different classes in the population. However, their existence was shown. Furthermore, because we took this preference heterogeneity into account, we also showed that there was no substantial heterogeneity across classes with regard to preferences for the other attributes. This makes it plausible that the estimates of the importance of these attributes can be generalized. There is no reason to assume that patients and informal caregiver outside of this study would have different preferences.

The existence of classes with different preferences for either treatment option is an important finding. Apparently, the choice between home and hospital of many respondents cannot be influenced much by adjusting the treatment at home. If this treatment were to become the standard, it would be against the wishes of a large proportion of patients and caregivers. Vice versa, many respondents value early assisted discharge and would not like to be confined to the hospital treatment. Although patients and caregivers who experienced home treatment were more likely to prefer it, the experience did not lead to enthusiasm in everyone. The results of this study, combined with the effectiveness and cost outcomes of the GO AHEAD trial, argue for giving patients a choice between treatment options.

The use of latent class conditional logit models resulted in more valid estimates and a better fit than the conventional multinomial logit model. Application of the AIC3 criterion showed that the best fitting models for patients as well as informal caregivers contained four different classes. Respondents in different classes had clearly different preferences for the hospital treatment. This preference heterogeneity could not have been identified by the multinomial logit model. Using the latter model, it might appear possible to construct early-discharge programs with a higher or lower average utility than the hospital treatment for the average patient. However, this would have been misleading since large proportions of patients preferred one of the treatment options irrespective of the description of the early assisted discharge program. The results of the latent class conditional logit models are consistent with this. The difference between results of the two models is mainly in the estimates for the general preference for the hospital treatment and the weight of the burden on informal caregivers, both of which are discrete random effects in the latent class model. The relative importance of the other attributes is similar, but not equal, for both models.

An alternative method for dealing with preference heterogeneity is the mixed logit model, which estimates a continuous distribution instead of a set of discrete points for the random effects. This model has gained popularity in recent years, also in the field of health economics [16]. Mixed logit and latent class models both have their merits. The latent class model has been described as a semi-parametric variant of the mixed logit. On the one hand, it requires the pre-specification of a number of classes. On the other hand, it frees the analyst from making – possibly incorrect – assumptions on the distribution of

parameters across respondents [26] and the results are more readily interpretable [22]. In our case, the advantages of the latent class model prevailed. It was intuitively obvious that at least three groups should be distinguished based on the choice patterns. Furthermore, using distinct classes was conceptually more consistent with the fixed preference of many patients than using a distribution. The choice of four classes was validated by comparing the AIC3 of each model.

In conclusion, different classes of patients and informal caregivers have different fixed preferences for the hospital or early assisted discharge treatment. These preferences are not changed by alterations in the early assisted discharge program. When choosing between two home options, respondents put the largest weight on copayments and the burden on the informal caregiver. The number of visits per day did not play a role.

5.5 References

1 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of COPD. Available at: http://www.goldcopd.com. 2009.

2 Hoogendoorn M, Feenstra TL, Hoogenveen RT, et al. Association between lung function and exacerbation frequency in patients with COPD. Int J Chron Obstruct Pulmon Dis. 2010; 5: 435-444.

3 Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med. 2010; 363: 1128-1138.

4 Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. 2004; 23: 932-946.

5 Dal Negro RW, Tognella S, Tosatto R, et al. Costs of chronic obstructive pulmonary disease (COPD) in Italy: the SIRIO study (social impact of respiratory integrated outcomes). Respir Med. 2008; 102: 92-101.

6 Miravitlles M, Murio C, Guerrero T, Gisbert R. Costs of chronic bronchitis and COPD: a 1-year follow-up study. Chest. 2003; 123: 784-791.

7 Nielsen R, Johannessen A, Benediktsdottir B, et al. Present and future costs of COPD in Iceland and Norway: results from the BOLD study. Eur Respir J. 2009; 34: 850-857.

8 Oostenbrink JB, Rutten-van Molken MP. Resource use and risk factors in high-cost exacerbations of COPD. Respir Med. 2004; 98: 883-891.

9 Toy EL, Gallagher KF, Stanley EL, et al. The economic impact of exacerbations of chronic obstructive pulmonary disease and exacerbation definition: a review. COPD. 2010; 7: 214-228.

10 Jordan RE, Hawker JI, Ayres JG, et al. Effect of social factors on winter hospital admission for respiratory disease: a case-control study of older people in the UK. Br J Gen Pract. 2008; 58: 400-402.

11 Shepperd S, Doll H, Angus RM, et al. Admission avoidance hospital at home. Cochrane Database Syst Rev. 2008; (4): CD007491.

12 Shepperd S, Doll H, Broad J, et al. Early discharge hospital at home. Cochrane Database Syst Rev. 2009; (1): CD000356.

13 Utens CM, Goossens LMA, Smeenk FWJM, et al. Early assisted discharge with communitey nursing for chronic obstructive pulmonary disease exacerbations: results of a randomised controlled trial. BMJ Open. Accepted for publication;

14 Goossens LMA, Utens CMA, Smeenk FWJM, et al. Cost-effectiveness of early assisted discharge for COPD exacerbations in the Netherlands. Submitted for publication;
15 Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ. 2000; 320: 1530-1533.
16 de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012; 21: 145-172.
17 McFadden D. Economic Choices. American Economic Review. 2001; 91: 351-378.
18 McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior,in P. Zarembka (ed.), Frontiers of Econometrics, Academic Press. In Zarembka P, ed. , Frontiers of Econometrics. New York: Academic Press, 1974.
19 Train KE. Discrete Choice Methods with Simulation (2nd ed.). New York: Cambridge University Press, 2009.
20 Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. J Health Econ. 2008; 27: 1078-1094.
21 Guo N, Marra CA, FitzGerald JM, et al. Patient preference for latent tuberculosis infection preventive treatment: a discrete choice experiment. Value Health. 2011; 14: 937-943.
22 Louviere JL. What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments. Environ Resour Econ. 2006; 34: 173-188.
23 Andrews RL, Currim IS. A Comparison of Segment Retention Criteria for Finite Mixture Logit Models. J Marketing Res. 2003; 40: 235-243.
24 Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models . Boca Raton: Chapman & Hall/CRC, 2004.
25 Skrondal A, Rabe-Hesketh S. Multivlevel logistic regression for polytomous data and rankings. Psychometrika. 2003; 68: 267-287.
26 Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transport Res B-Meth. 2003; 37: 681-698.

6

Is the EQ-5D responsive to recovery from a moderate COPD exacerbation?

Background The aim of the study was to correctly estimate the cost-effectiveness of treatments that reduce COPD exacerbations, the utility gains from preventing exacerbations need to be measured. This requires utility measurement during exacerbations. The aim of this study was to assess the ability of the EQ-5D to detect the recovery from moderate COPD exacerbations.

Methods In the US, 65 COPD and/or chronic bronchitis patients (≥40 years old smokers or ex-smokers with a history of 10 pack-years) were enrolled within 48 h of symptom onset of the exacerbation. Patients completed the EQ-5D at enrolment and after 7, 14 and 42 days. Symptoms and medication use were recorded in diaries. Change over time and loss of quality-adjusted life years (QALYs) due to the exacerbation was estimated. Using standardised response mean (SRM) as the metric of responsiveness, we compared the responsiveness of the EQ-5D to the responsiveness of morning peak expiratory flow rate, rescue medication use and symptom scores. SRMs were also used to assess whether patients with greater improvements in peak expiratory flow rate, rescue medication use, symptom scores, clinician global impression of change, and patient global impression of change had a greater improvement in EQ-5D than patients with smaller improvement.

Results Mean utility index scores (standard deviation) using the US value set were 0.683 (0.209), 0.726 (0.216), 0.768 (0.169) and 0.760 (0.181) at days 1, 7, 14 and 42, respectively. The mean of each patient's lowest index score, either at visit 1 or visit 2, was 0.651 (0.213). Over the course of 6 weeks there was a highly significant improvement in mean utility. The greatest improvement was seen between day 7 and day 14. Patients lost on average 0.00896 QALY (0.0086) or 3.27 (3.13) quality-adjusted life days during the exacerbation. The EQ-5D (SRM: 0.653) was more responsive to change than peak expiratory flow (0.269), rescue medication use (0.343) and sputum symptom scores (0.322) and equally responsive as cough (0.587) and dyspnoea (0.638) symptom scores.

Conclusion The EQ-5D is responsive to the recovery from a moderate COPD exacerbation.

Lucas M.A. Goossens, Michael C. Nivens, Paul Sachs, Brigitta U. Monz, Maureen P.M.H. Rutten-van Mölken

This study was published in Respiratory Medicine, 2011; 105(8):1195-1202.

6.1 Introduction

Many patients with chronic obstructive pulmonary disease (COPD) experience recurrent exacerbations with increased breathlessness and/or wheeze, often accompanied by greater volume of sputum and increased cough. Exacerbations greatly contribute to a decline in health-related quality of life [1-4]. A main objective of treatment is to reduce the frequency and severity of exacerbations. To fully appreciate the effectiveness of treatments, the full health-related quality of life gains that result from this reduction should be included. In clinical studies this is often not done, because health-related quality of life is measured only at regularly scheduled intervals, but not during exacerbations. Specific studies that measure quality of life during an exacerbation are necessary.

Health-related quality of life can be estimated with disease-specific or generic instruments. Disease-specific instruments were designed to detect even small changes in the patient's health-related quality of life. However, health economists prefer the use of generic instruments, because they enable the comparison of health states as well as the benefits of medical interventions across diseases. To express these benefits in terms of quality-adjusted life years (QALYs), utility weights are assigned to health states [5]. Perfect health, by definition, has a utility of 1, while death has a utility of 0. Health states may also be considered worse than death and have negative utility. A year spent in perfect health equals one QALY.

Among the most widely used instruments to measure utilities is the European quality of life scale (EQ-5D).[6]

The EQ-5D has been validated and applied in the stable phase of COPD [7-11]. In addition, O'Reilly et al. and Menn et al. applied the instrument in patients with severe exacerbations, requiring hospital admission [12,13]. A paper by Paterson et al. contains limited information on the use of the EQ-5D at the onset and after treatment of type-1-exacerbations (Anthonisen classification) of chronic bronchitis [14].

However, in order to capture the full health-related quality of life loss during an exacerbation, the entire course of recovery must be described. In the current chapter, we evaluated the ability of the EQ-5D to detect the recovery from a moderate COPD exacerbation, requiring antibiotic or systemic steroid therapy but no hospital admission [15] over six weeks: at the onset of the exacerbation, during treatment, and after recovery.

We tested the changes in the EQ-5D utility index scores over time and compared the effect sizes to those of several symptom scores, peak flow and the use of rescue medication. The change in shortness of breath symptom score as well as medication use were expected to be most closely related to the EQ-5D score, since they have the strongest effect on quality of life. We also, tentatively, estimated the health loss due to a moderate exacerbation.

6.2 Methods

The EQ-5D data were collected alongside an exacerbation study that was primarily conducted to assess the responsiveness of a previously developed cough and sputum assessment questionnaire, the CASA-Q [16].

Patients and setting

After institutional review board approval and written informed consent, 65 male and female COPD and/or chronic bronchitis patients were enrolled at 7 study sites in the United States. They had to be 40 years or older and had to be current or former smokers with a history of at least 10 pack-years. Patients were enrolled when they visited the outpatient clinic within 48 hours of symptom onset. An exacerbation was defined as the increase or new onset of at least two lower respiratory symptoms related to COPD, with at least one symptom lasting three or more days and requiring a change in treatment. An exacerbation was defined as moderate if the change in treatment included the prescription of antibiotics and/or systemic steroids. Exacerbations that required hospital admission were considered severe and were excluded from this study. Also excluded were patients with significant other diseases which could influence the results of the study or the subject's ability to participate in the study. Other exclusion criteria were a history of asthma, cystic fibrosis, bronchiectasis, active pneumonia or tuberculosis.

Design

The study was designed as a prospective cohort study. The change in EQ-5D scores over a period of 6 weeks was observed. In addition to the visit at enrolment (visit 1), patients were evaluated at day 7 (visit 2), day 14 (visit 3) and day 42 (visit 4). The interval between visits was allowed to vary by up to 3 days. The patients were treated for their exacerbation at the discretion of the investigator according to the standards of care.

Measurements

At each visit participants completed the EQ-5D, in which health status is described by ticking off one of three levels of functioning ("no problems", "some problems" and "extreme problems") on five dimensions: mobility, selfcare, usual activities, pain/discomfort and anxiety/depression. Using a set of weights (value set) for each level of functioning in each dimension, the descriptive information can be converted into a single utility index.

Since this study was conducted in the United States, the US value set was used to calculate utilities [17]. Patients were also asked to rate their health on the EQ-5D Visual Analog Scale (VAS), which ranges from 0 (worst imaginable health state) to 100 (best imaginable health state).

In the daily diary, patients recorded how short of breath they were, how often they coughed and how often they brought up sputum using a 5-point response scale ranging from "not at all" to "a lot", or "never" to "always". The diary also contained measurements of morning peak flow (before taking medication) and use of rescue medication. Pre- and post-bronchodilator spirometry was performed at visit 4. Patient's and Clinician's Global Impression of Change (PGI-C, CGI-C) were measured at visit 4: they rated the change in cough and sputum symptoms on a seven-point scale ranging from "very much worse" to "very much better".

Statistical analysis

Changes in the proportions of patients that report either "no problems" or "some problems"/"extreme problems" on the EQ-5D between baseline and visit 4 were analysed using McNemar's test. Baseline EQ-5D VAS and utility index was defined as the score at either visit 1 or 2, whichever was lower. This baseline value captured the point in time when the impact of the exacerbation was most severe.

Mean (standard deviation, SD) scores for EQ-5D VAS and utility index scores were calculated at all visits and displayed graphically.

Responsiveness was expressed as the Standardised Response Mean (SRM) – defined as the average change between EQ-5D index scores at baseline and visit 4 divided by the standard deviation of that change. SRM of the EQ-5D was compared to the corresponding SRM of the various symptom scores derived from the diary items. We used the interpretation of Cohen, where 0.2 is indicative of a small effect, 0.5 of a medium and 0.8 of a large effect [18].

Using a change in morning peak expiratory flow (between the first week and the last week) above the median as an external criterion to define a greater improvement, we assessed whether EQ-5D scores improved more in those with a greater improvement in peak flow than in those with a smaller improvement in peak flow, i.e. whether the SRM of the EQ-5D index score was greater in those with a greater improvement in peak flow. Similarly, SRM for EQ-5D index scores were calculated for the cohort divided into two groups by 1) patient global impression of change ((very) much better versus the remaining response options), 2) clinician global impression of change ((very) much better versus the remaining response options), 3) symptom score (change in mean daily symptom score between the first week and the last week below and above the median), and 4) rescue medication use (change in mean daily number of puffs between the first week and the last week below and above median). The standardised difference between the above mentioned pairs was calculated as the difference between the mean change from baseline at visit 4 divided by the pooled standard deviation of these change scores. Differences in change from baseline in EQ-5D index score between the pairs were tested by t-tests.

Repeated measures analysis was performed to analyse the change from baseline to visit 3 and 4 in EQ-5D index scores using the SAS procedure PROC MIXED with covariance

modelled as "unstructured". The model included time (visits), EQ-5D score at baseline, age, gender, smoking status, lung function at visit 4 (as an approximation of lung function without exacerbation), diagnosis (chronic bronchitis/ COPD) and co-morbidity (either the Charlson Comorbidity Index score [19] or the number of co-morbidities), as well as the interaction of baseline EQ-5D score and time.

For each patient, the QALY loss due to the exacerbation was calculated by subtracting the QALYs during the exacerbation from the QALYs the patient would have had if the COPD had remained stable during the same period. The latter was approximated by taking the highest score at any of the visits. The number of QALYs was calculated by summing the days under observation weighted by their utilities (using linear interpolation).

6.3 Results

Patients In total 59 of the 65 subjects completed the study and were included in the analysis (see table 6.1). Three patients (hospital admissions (n = 2) and concomitant lung cancer (n = 1) proved to be ineligible after being included in the study. Three others discontinued their cooperation after the first visit. One patient had one missing value, but his other three values were included in the analysis.

Table 6.1 Baseline characteristics.

N	59
Age: mean (SD)	61.1 (10.4)
Female: n (%)	40 (67.8)
Current smoker: n (%)	32 (54.2)
Pack-years: mean (SD)	60.14 (30.12)
Number of co-morbidities: mean (SD)	6.69 (3.34)
Charlson co-morbidity index score: mean (SD)	0.66 (0.921)
Diagnosis COPD: n (%)	54 (91.5)
Diagnosis chronic bronchitis only: n (%)	5 (8.5)
GOLD classification at visit 4[a]	
Not obstructed: n (%)	19 (33.3)
Stage 1 (Mild): n (%)	2 (3.5)
Stage 2 (Moderate): n (%)	14 (24.6)
Stage 3 (Severe): n (%)	13 (22.8)
Stage 4 (Very severe): n (%)	9 (15.8)

SD: standard deviation.

[a] Two patients with missing lung function measurements.

EQ-5D dimensions

The proportion of patients with "no problems" significantly increased over time on all dimensions (figure 6.1). P-values for differences in proportions of patients reporting no, some or extreme problems between baseline and visit 4 were 0.008 for the mobility dimension, 0.007 for self-care, 0.021 for usual activities, 0.012 for pain and 0.006 for anxiety and depression.

At any visit the majority of patients had problems performing their usual activities and felt pain and anxiety. A minority experienced problems when performing selfcare. No problems on any dimension were reported by 4 respondents (6.8%) at baseline, by 9 respondents (15.3%) at visit 3 and by 10 respondents (16.9%) at visit 4.

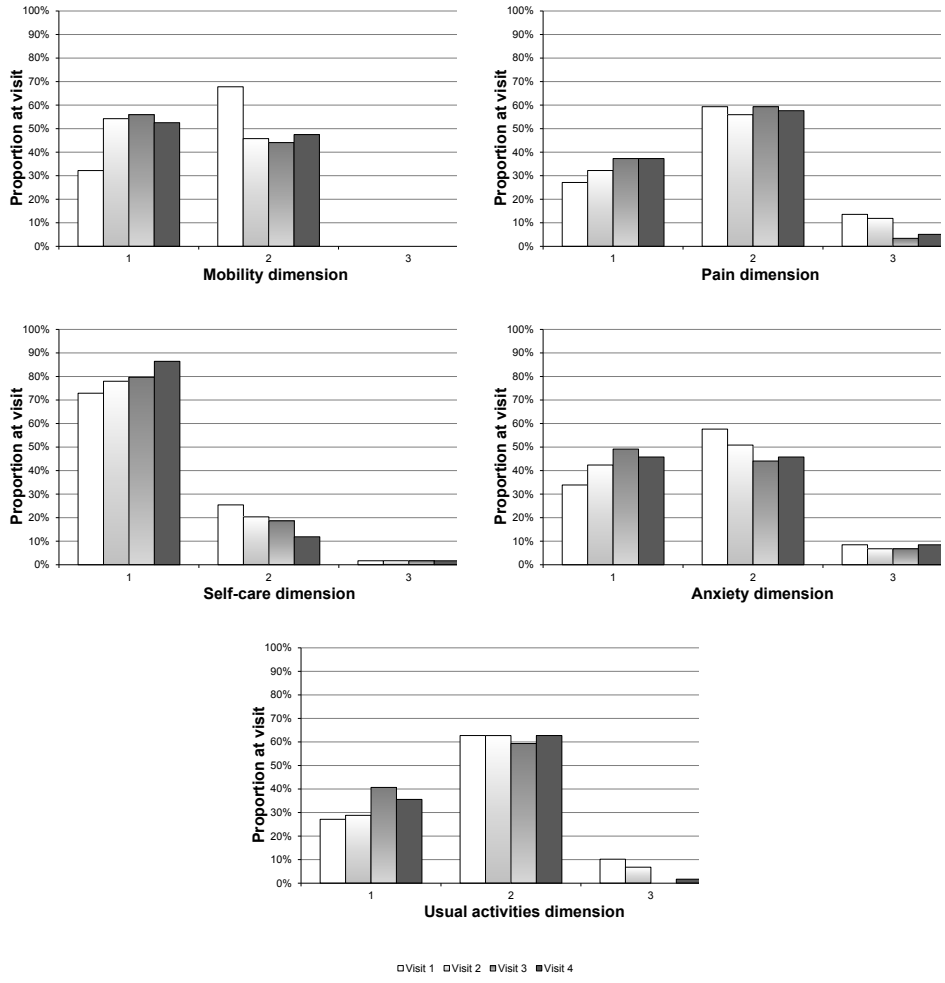

Figure 6.1 Proportion of patients reporting no problems (1), some problems (2) or extreme problems (3) on each EQ-5D dimension, by visit.

EQ-5D VAS scores

For 17% of patients, VAS scores decreased from visit 1 to visit 2. However, most of the mean improvement in VAS scores occurred between these measurements. Table 6.2 shows that the highest VAS scores were reported for visit 4. The differences between baseline and visits 3 and 4 were highly significant.

Table 6.2 EQ-5D VAS and utility index scores per visit.

EQ-5D VAS scores			
Visit	Mean	SD	Range
Lowest[a]	34.75	25.244	1-85
1	36.68	25.244	1-85
2	48.03	32.787	3-100
3	48.19	32.336	5-100
4	50.25	31.840	3-100
Highest[b]	55.81	34.190	5-100
EQ-5D VAS, differences in scores between visits			
	Difference (SD)	P-value	
Visit 2 - Visit 1	11.356 (20.678)	<0.0005	
Visit 3 - Visit 2	0.153 (14.539)	0.936	
Visit 4 - Visit 3	2.068 (14.027)	0.262	
Visit 3 - Lowest[a]	13.441 (21.477)	<0.00002	
Visit 4 - Lowest[b]	15.508 (19.754)	<0.000001	
EQ-5D utility index score (US tariff)			
Visit	Mean	SD	Range
Lowest[a]	0.651	0.213	0.165-1.00
1	0.683	0.209	0.165-1.00
2	0.726	0.216	0.165-1.00
3	0.768	0.169	0.202-1.00
4	0.760	0.181	0.202-1.00
Highest[b]	0.828	0.148	0.202-1.00
EQ-5D utility index score, differences in scores between visits			
	Difference (SD)	P-value	
Visit 2 - Visit 1	0.0381 (0.1560)	0.068	
Visit 3 - Visit 2	0.0413 (0.1523)	0.044	
Visit 4 - Visit 3	-0.0073(0.1574)	0.722	
Visit 3 - Lowest[a]	0.1167 (0.1527)	<0.0000002	
Visit 4 - Lowest[a]	0.1094 (0.1676)	<0.000001	

[a] Lowest = value at visit 1 or 2, whichever is lower.

[b] Highest = highest value at any visit.

EQ-5D index scores

There was a highly significant improvement in mean utility index scores (table 6.2). Patients improved most between visits 2 and 3. The utility index of 76% of patients improved between baseline and visit 3, while it deteriorated in 8.5%. Table 6.2 shows that the mean of each patient's lowest index score was lower than the mean at visit 1, following the fact that quality of life deteriorated in 25% of patients after treatment had started. Overall, VAS scores showed improvement before the utility index scores started to do so. Figure 6.2 presents the course of moderate exacerbations over time. The repeated measures analysis showed that the improvement of the EQ-5D utility index score at visits 3 and 4 since baseline was predicted by its baseline value, the patient's gender and smoking status (goodness of fit of the model: –2 res log likelihood = –135.9). Patients with a higher baseline utility index score showed less improvement. Women did not recover as well as men (difference: 0.07242, p=0.0121). Current smokers covered a larger distance between baseline and recovery than former non-smokers (difference 0.05170 p =0.0602). As before, no significant difference was found between visit 3 and 4. Age (p=0.9694), post-bronchodilator FEV_1 (in % predicted) at visit 4 (p=0.268) and co-morbidity (0.6229) were dropped from the final model because they were not significantly associated with recovery of utility values.

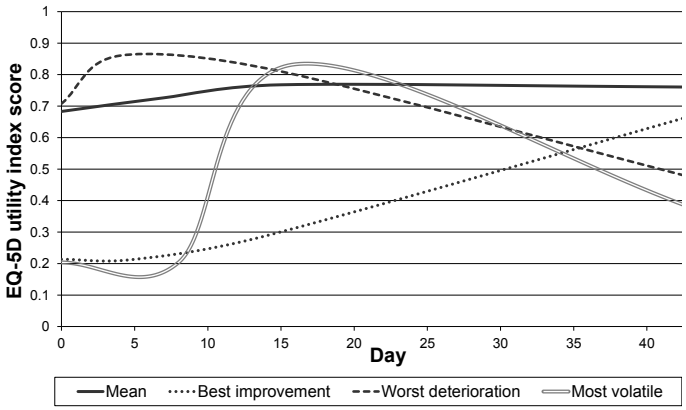

Figure 6.2 EQ-5D utility index scores over the course of a moderate exacerbation: mean over all patients, patient with the best improvement and patient with the worst deterioration, the most volatile patient.

Health loss

Participants lost on average 0.00896 QALY (SD 0.0086) or 3.27 (SD 3.13) quality-adjusted life days during the exacerbation. The largest individual loss was 0.0364 QALY or 13.29 quality-adjusted life days. Two patients incurred no loss of QALYs compared to the situation with stable COPD, because they had the same score throughout the study.

Responsiveness

A medium to large SRM for the change from baseline in utilities was observed (table 6.3). The SRM was comparable for the symptom scores on cough and shortness of breath but smaller for sputum, rescue medication use and expiratory peak flow.

Table 6.3 Standardised response mean (SRM): comparison between the EQ-5D, symptoms scores, rescue medication use and morning peak expiratory flow.

	Mean change	SD of change	SRM
Utility index score EQ-5D[a]	0.109	0.168	0.653
Symptom scores			
Sputum[b]	0.196	0.609	0.322
Cough[b]	0.395	0.673	0.587
Shortness of breath[b]	0.588	0.207	0.638
Rescue medication	0.708	2.068	0.343
Expiratory peak flow[b]	14.144	52.673	0.269

[a] Change between lowest score at visit 1 or 2 and score at visit 4.

[b] Change between first and last week of study.

The largest standardised difference in EQ-5D utility change was found in the patients stratified by their rescue medication use (Figure 6.3). The improvement in EQ-5D utility was significantly greater in patients who had a greater reduction in rescue medication use than in those who had a smaller reduction in rescue medication use (p=0.018). Significance was almost reached for the improvement in EQ-5D utility stratified by improvement of shortness of breath (p=0.051) and peak flow (p=0.058).

The standardised differences were negative for participants with a good or very favourable patient's or clinician's impression of change compared to the rest of the sample. However, the differences in change in EQ-5D index scores between these groups were not significant (p=0.128 and p=0.657 respectively). Neither were those between groups with improvements in cough and sputum scores above and below the median (p=0.144 and p=0.594).

6.4 Discussion

This study has shown that the EQ-5D index score is responsive to recovery from a moderate COPD exacerbation. When the exacerbation was at its worst, the average utility score was 0.651. It increased to 0.768 on day 14, after which it remained largely stable until the final visit at 6 weeks. Three-quarters of patients experienced an improvement in health-related quality of life after the worst day of the exacerbation. The mean total improvement was statistically significant and can be considered medium-sized to large according to Cohen's classification.

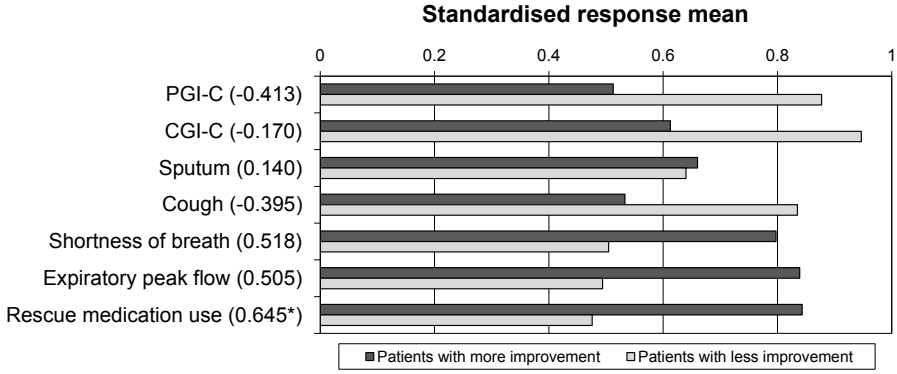

Figure 6.3 Standardised response means of EQ-5D utility index scores, by subgroups of patients (with standardised differences between subgroups in brackets).

*Difference is statistically significant, t-test $p < 0.05$.

Black bar: subgroup of patients with more improvement, where more improvement is defined as patients whose global impression of change is "(very) much better" or who have a change in mean daily symptom score, mean daily number of puffs of rescue medication, or mean daily morning peak flow above the median. Gray bar: subgroup of patients with less improvement (i.e. the remaining patients). The standardised difference between subgroups is given between brackets.

Abbreviations: CGI-C, Clinician's Global Impression of Change; COPD, Chronic Obstructive Pulmonary Disease; EQ-5D, European Quality of Life Scale (5 dimensions); PGI-C, Patient's Global Impression of Change; SD, Standard Deviation; SRM, Standardised Response Mean; QALY, (health-related) Quality-adjusted Life Year; VAS, (health-related) Visual Analog Scale.

As the EQ-5D improved, so did the VAS and symptom scores, peak flow and rescue medication use. However, the SRM for the change in EQ-5D index score was generally larger than for peak flow, rescue medication and phlegm symptom scores. This was probably caused by the index covering more aspects of health-related quality of life.

The responsiveness of the EQ-5D was also demonstrated by greater improvements in EQ-5D scores among the patients with a greater reduction in the use of rescue medication, a greater improvement in morning peak expiratory flow and a greater improvement in shortness of breath symptoms. In contrast, change in EQ-5D scores was not better for patients with better scores on the patient's global impression of change or the clinician's global impression of change.

All five dimensions of the EQ-5D contributed significantly to the mean improvement in the EQ-5D index score. The mobility dimension showed the largest improvement. Most participants had problems with usual activities and experienced pain and anxiety even at the last visit. This may reflect their underlying impairment due to their chronic respiratory disease.

The EQ-5D showed a disutility at baseline in almost all patients.

COPD exacerbations are not necessarily at their most severe at the moment when patients first consult a physician. Indeed, our results showed a considerable proportion of patients whose EQ-5D scores worsened during the first week. When studying an instruments ability to respond to the recovery from an exacerbation, the most informative is the path

from the lowest point of the exacerbation, whenever this occurs, to its resolution weeks later. This is what we did in our repeated measures analysis and SRM calculations, where baseline was defined as the score at either visit 1 or 2, whichever was lower. However, differences between visit 1 and visits 3 and 4 were highly significant as well.

In a tentative calculation, it was derived that the average QALY loss during the exacerbation was 0.00896, or 3.27 quality-adjusted life days. This calculation is sensitive to the estimate of the utility value during the stable phase of the disease. Since we did not have actual data on quality of life during the stable phase, we chose to compare the health-related quality of life that each patient actually experienced, with the highest individual score during the study e as an approximation of the quality of life during stable disease. The results of the calculation must be considered a preliminary estimate.

It is possible that patients had not completely recovered at the end of the study to their stable health-related quality of life levels, although improvement appeared to reach a plateau after two weeks. Furthermore, an exacerbation may cause permanent damage to the patient's health. This would mean that our estimate of quality of life loss would have been too low. To capture the entire loss, patients would have to fill out the EQ-5D questionnaire during the stable phase, at times when no exacerbation is expected. Another approach would be to collect values for COPD health profiles that contain a description of the severity of the COPD during the stable phase and a description of the exacerbation profile, as was recently reported by Rutten-van Mölken et al.[20] The deduction in utility value due to a non-severe exacerbation in this study was 0.010 (Dutch value set); compared to 0.009 (US value set) in the current study.

As expected, the utility scores in our study were much higher than those reported by O'Reilly et al. for patients experiencing a severe exacerbation in Britain. At the onset, the mean EQ-5D utility value for their group was negative (-0.077, according to the UK tariff) and nearly two-thirds of respondents indicated that they felt "worse than death". [12] At discharge the mean utility had improved to 0.576 with only 5% of patients indicating a health-related quality of life "worse than death". However, within the three months after discharge quality of life deteriorated markedly again. We did not find indications of a deteriorating health after an initial improvement in our study, but our observation period was limited to six weeks. Another study in patients with a severe exacerbation was performed in Germany by Menn et al., who also concluded that the EQ-5D is suitable for measuring health-related quality in this patient group. [13] However, they found much higher utility scores than O'Reilly. The authors suggested that patients in Germany might be admitted to the hospital relatively early, which would make average inpatient exacerbations milder than in the UK. Paterson et al. assessed the difference between utility scores in chronic bronchitis patients before and shortly after treatment for a type-1-exacerbation, [14] which could in this case apparently be classified as "moderate". Unfortunately, they did not present utility scores during the exacerbation. They did, however, report the change between the start and

the end of treatment – as 0.17 according to the UK value set e which is comparable to our estimate if it had been based on the same value set.

Respondents completed the CASA-Q instrument [16] before the EQ-5D. Concentrating on respiratory symptoms in the CASA-Q might have made respondents more sensitive to possible changes on the EQ-5D dimensions, which would increase the responsiveness of the latter questionnaire [21,22]. On the other hand, examples have been shown of an effect in the opposite direction, by inducing respondents to exclude considerations about respiratory problems from their answers to the EQ-5D because the impact of symptoms had already been covered by the other questionnaire [22,23]. However, studies on ordering effects in generic and specific quality of life questionnaires did not find any significant or relevant effects in a number of diseases [24-28]. We did not investigate the presence and magnitude of the question order effect.

We had two reasons for not assembling a control group of patients without exacerbations to compare to the patients with exacerbations. Randomisation into groups with and without exacerbation was impossible and we could not ensure that the control group would be sufficiently comparable. It would have been possible to enrol patients and wait until an exacerbation developed. However, this lengthy and costly procedure was not considered worthwhile since utility index score values in patients outside of exacerbations are known to be stable [8]. Therefore, any significant change in mean utility during the exacerbation can be interpreted as a consequence of (the recovery from) the exacerbation.

As a generic instrument for health-related quality of life measurement, the EQ-5D was developed for the comparison of quality of life and of the effects of healthcare interventions across diseases. We do not recommend its use in routine clinical practice. The EQ-5D was not designed for tracking an individual patient's response to treatment or as a measure on which treatment adjustments could be based. Our study does, however, have implications for conducting future cost-effectiveness studies. In contrast to current practice these studies can and should include the utility gains resulting from a reduction in exacerbations. It has been argued that the EQ-5D is not responsive to small changes in health of COPD patients [8,9], but our study has shown that it was responsive over a moderate exacerbation.

In conclusion, the EQ-5D was found to be responsive to the recovery from a moderate COPD exacerbation. The greatest improvement in utility scores was reached within two weeks after the onset of the exacerbation. The EQ-5D was more responsive than expiratory peak flow, rescue medication and sputum symptom scores and equally responsive as cough and dyspnoea symptoms scores.

6.5 References

1 Doll H, Miravitlles M. Health-related QOL in acute exacerbations of chronic bronchitis and chronic obstructive pulmonary disease: a review of the literature. Pharmacoeconomics. 2005; 23: 345-63.
2 Miravitlles M, Ferrer M, Pont A, et al. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: a 2 year follow up study. Thorax. 2004; 59: 387-95.
3 Schmier JK, Halpern MT, Higashi MK, Bakst A. The quality of life impact of acute exacerbations of chronic bronchitis (AECB): a literature review. Qual Life Res. 2005; 14: 329-47.
4 Seemungal TA, Donaldson GC, Paul EA, et al. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998; 157: 1418-22.
5 Drummond MF, Sculpher MJ, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.
6 Brooks R, Rabin R, De Charro F. The measurement and valuation of health status using EQ-5D: A European perspective. Evidence from the EuroQol BIOMED Research Programme . Dordrecht: Kluwer Academic Publishers; 2003., 2003.
7 Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004; 13: 873-84.
8 Harper R, Brazier JE, Waterhouse JC, et al. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. Thorax. 1997; 52: 879-87.
9 Stavem K. Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. Qual Life Res. 1999; 8: 45-54.
10 Rutten-van Mölken MPMH, Oostenbrink JB, Tashkin DP, et al. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? Chest. 2006; 130: 1117-28.
11 Pickard AS, Wilke C, Jung E, et al. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. Respir Med. 2008; 102: 519-36.
12 O'Reilly JF, Williams AE, Rice L. Health status impairment and costs associated with COPD exacerbation managed in hospital. Int J Clin Pract. 2007; 61: 1112-20.
13 Menn P, Weber N, Holle R. Health-related quality of life in patients with severe COPD hospitalized for exacerbations – comparing EQ-5D, SF-12 and SGRQ. Health Qual Life Outcomes. 2010; 8: 39.
14 Paterson C, Langan CE, McKaig GA, et al. Assessing patient outcomes in acute exacerbations of chronic bronchitis: the measure your medical outcome profile (MYMOP), medical outcomes study 6-item general health survey (MOS-6A) and EuroQol (EQ-5D). Qual Life Res. 2000; 9: 521-7.
15 Cazzola M, MacNee W, Martinez FJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. Eur Respir J. 2008; 31: 416-469.
16 Monz BU, Sachs P, McDonald J, et al. Responsiveness of the cough and sputum assessment questionnaire in exacerbations of COPD and chronic bronchitis. Respir Med. 2010; 104: 534-541.
17 Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 2005; 43: 203-20.
18 Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
19 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40: 373-83.
20 Rutten-van Molken MP, Hoogendoorn M, Lamers LM. Holistic preferences for 1-year health profiles describing fluctuations in health: the case of chronic obstructive pulmonary disease. Pharmacoeconomics. 2009; 27: 465-477.

21 Lee S, Grant D. The effect of question order on self-rated general health status in a multilingual survey context. Am J Epidemiol. 2009; 169: 1525-1530.
22 Schwarz N, Strack F, Mai H. Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. Pub Opin Q. 1991; 55: 3-23.
23 Tourangeau R, Rasinski KA, Bradburn N. Measuring happiness in surveys: a test of the substraction hypothesis. Pub Opin Q. 1991; 55: 255-266.
24 Rat AC, Baumann C, Klein S, et al. Effect of order of presentation of a generic and a specific health-related quality of life instrument in knee and hip osteoarthritis: a randomized study. Osteoarthritis Cartilage. 2008; 16: 429-435.
25 Barry MJ, Walker-Corkery E, Chang Y, et al. Measurement of overall and disease-specific health status: does the order of questionnaires make a difference? J Health Serv Res Policy. 1996; 1: 20-27.
26 Childs AL, Submacular Surgery Trials Research Group. Effect of order of administration of health-related quality of life interview instruments on responses. Qual Life Res. 2005; 14: 493-500.
27 McColl E, Eccles MP, Rousseau NS, et al. From the generic to the condition-specific?: Instrument order effects in Quality of Life Assessment. Med Care. 2003; 41: 777-790.
28 Wu AW, Jacobson DL, Berzon RA, et al. The effect of mode of administration on medical outcomes study health ratings and EuroQol scores in AIDS. Qual Life Res. 1997; 6: 3-10.

Collapsibility and censoring

What's the bias in estimating time-to-event?

7

Background Treatment effects in time-to-event analysis are often expressed as hazard ratios. Parametric survival models can also be used to estimate projected mean time-to-event. Two problems in survival analysis are often misunderstood or ignored: non-collapsibility and omitted-variable bias due to non-random censoring. Non-collapsibility exists when the treatment effect changes as prognostic covariates are added to the regression model, even when confounding is absent (e.g. in a randomised controlled trial (RCT)). Earlier studies have sometimes confused these phenomena and misrepresented the problems that non-collapsibility and censoring can induce. The objectives of this study were to disentangle their effects and assess their impact on estimates of mean time-to-event.

Methods Survival, treatment and five normally distributed prognostic covariates were simulated in RCT-like datasets with and without censoring. Weibull regression models with an increasing number of covariates were used to calculate the hazard ratios and mean time-to-event.

Results With uncensored data, hazard ratios decreased as additional covariates were included, while time-to-event remained constant. With censored data, time-to-event estimates increased sharply and improved as covariates were added.

Conclusion Analysis of synthesized data makes it possible to distinguish between the impact of non-collapsibility and censoring. While hazard ratios from Weibull models are non-collapsible, mean time-to-event is collapsible. Censoring threatens the validity of estimates when prognostic factors are omitted from the regression model.

Lucas M.A. Goossens, Chantal W.M. van Gils, W. Ken Redekop

Submitted for publication.

7.1 Introduction

In time-to-event (or survival) analyses, the treatment effect is often expressed in a hazard ratio, the relative probability of instantly experiencing the event of interest compared to a reference group. This hazard ratio is a purely statistical measure and lacks a clear and intuitive interpretation about the size of the actual treatment effect. Fortunately, parametric time-to-event models can be used to estimate the mean time-to-event (or life expectancy) [1]. Nevertheless, time-to-event analysis has two challenging characteristics: *non-collapsibility*, which complicates interpretation of hazard ratios, and bias caused by non-random *censoring*. It is important that these concepts be distinguished.

Non-collapsibility refers to a situation where the estimated treatment effect changes when prognostic covariates are added to the analysis, even when these covariates are not confounders [2,3]. Because of non-collapsibility, several conditional hazard ratios can be estimated for different combinations of covariates in the regression model. As long as confounding has been addressed successfully, these may all be correct and yet different from each other [4,5].

Non-collapsibility is not equal to confounding and does not lead to bias. *Confounding* is traditionally defined as a source of bias arising from causes of the outcome of interest that are associated with treatment, but not affected by it[6]. Estimates are unconfounded if prognostic variables are jointly balanced across treatment groups, which they are asymptotically expected to be after randomisation.

In contrast to non-collapsibility, censoring can lead to bias even when no confounding is present. Estimates from time-to-event analysis are only valid if the probability of being censored is not associated with the risk of the outcome of interest (e.g. dying) within strata of baseline characteristics and treatment [6]. In many medical studies, this assumption of independent censoring is not necessarily valid if all surviving patients are censored at the end of the trial's follow-up period. For example, one could imagine patients with a good prognosis, e.g. in a less severe disease stage, who are less likely to die before the trial ends. Within disease stages, patients receiving the same treatment are equally likely to be censored, i.e. to survive the trial period. Any analysis that adjusts for disease stage will ensure that there is no association between prognosis and probability of censoring. Conversely, if this adjustment is not applied, the probability of being censored is associated with the risk of dying within each treatment group. In that case, the assumption of independence is violated and an omitted-variable bias is introduced.

Several studies in the 1980s and 1990s have shown the potential impact of the combination of non-collapsibility and censoring on hazard ratios from Cox proportional hazards and exponential models [7-12]. However, the distinction between the two issues was often somewhat obscured and misinterpreted.

Furthermore, there seems to be only limited recognition of the censoring problem in the literature. While important prognostic factors tend to be included in Weibull models

on observational data, see for instance [13-15], many examples can be found of studies using data from randomised controlled trials (RCTs) with treatment as the sole covariate, for instance [16-23].

Confounding, non-collapsibility and non-independent censoring may all occur at the same time and cause differences in estimates across models. However, this does not necessarily happen in all of the products of an analysis. While it is well known that odds ratios are non-collapsible, it does not occur in risk ratios, risk differences and mean predicted probabilities, which can all be derived from the same logit model. The collapsibility of projected mean survival in Weibull models was proven by Lancaster, but he did not consider censoring [24].

The objectives of this chapter were to disentangle the effects of non-collapsibility and non-dependent censoring variable bias and to assess their impact on the prediction of survival in Weibull models. We set out to demonstrate the consequences of omitting prognostic variables from a model in a dataset with non-random censoring and to investigate the collapsibility of the predicted mean time-to-event.

An example of non-collapsibility

The concept of non-collapsibility has been explained extensively by Greenland et al.[2]. This phenomenon may appear in odds ratios, hazard ratios and rate ratios, and possibly additional statistical outcome measures. Whether these measures are collapsible depends on the value of the treatment effect and sometimes on the statistical model [11].

Non-collapsibility is often shown in the context of odds ratios. An instructive example from an imaginary RCT is presented in table 7.1. Here we see that while sex is a prognostic factor, it is not a confounder since sex is not associated with probability of treatment. Confounding has been actively avoided, because treatment has been successfully balanced across the sexes by means of randomisation. Hauck has suggested calling variables of this type 'mavericks' [25].

The treatment increases the probability (or 'risk') of being cured in both sexes. The risk difference is constant across these strata and in the full sample. While the risk ratios are different for men and women, the ratio for all patients is simply a weighted average of the two sex-specific ratios. In contrast, the odds ratio shows something counterintuitive: it is constant across sexes but different from the overall odds ratio. In other words, the average treatment effect (or marginal effect) of 2.25 seen in the study population is not the same as the average of the treatment effects seen in individuals in the populations (i.e., the conditional effect of 2.67) [26,27].

A mathematical explanation for this is provided by Greenland [2].

Table 7.1 Risk differences, risk ratios and odds ratios

	Men		Women		Total	
	Treated	Untreated	Treated	Untreated	Treated	Untreated
Cured	80	60	40	20	120	80
Not cured	20	40	60	80	80	120
Total	100	100	100	100	200	200
Risk	0.8	0.6	0.4	0.2	0.6	0.4
Risk differences	0.2		0.2		0.2	
Risk ratios	1.33		2		1.5	
Odds ratios	2.67		2.67		2.25	

7.2 Methods

An RCT-like dataset was synthesized and analysed using several marginal and conditional Weibull time-to-event models. The treatment effect was expressed in hazard ratios and mean differences in life expectancy. The results of all models were compared both with each other and with the true values (i.e., conditional hazard ratio and marginal mean difference in life expectancy). This was repeated after applying censoring on the data at 6 months (or 183 days).

Synthesis of a dataset

We created 100,000 patients, each with a complete set of five covariates (A, B, C, D and E), which were determined independently of each other and randomly chosen from normal distributions with mean 100 and standard deviation 10. The chance of receiving treatment T was determined randomly in order to isolate the collapsibility and censoring effects from any effects of confounding,.

The prognosis of all patients was determined using the following equations for a Weibull function.

$$S_i(t) = \exp(-e^{X\beta} * tp) \tag{1}$$

$$X\beta = \beta_0 + \beta_1 * A + \beta_2 * B + \beta_3 * C + \beta_4 * D + \beta_5 * E + \beta_6 * T \tag{2}$$

With:

$\beta_0 = -1$

$\beta_1 = \beta_2 = \beta_3 = -0.04$

$\beta_4 = \beta_5 = 0.04$

$\beta_6 = -1$ (corresponding with HR=0.368)

$p = 1.3$

$S_i(t)$ denotes the individual probability of being alive at time t. Equation (2) contains the covariates and their parameters. Covariates A, B and C and treatment T had a positive impact on life expectancy, while covariates D and E had a negative impact. The auxiliary parameter p – the shape parameter – is larger than one, which means that the hazard increased with time.

A complete survival curve was constructed for each patient. The survival time actually applied to that patient was established by randomly drawing a value from a uniform distribution (0-1) and determining the associated survival time t_i associated with that value.

For simplicity, all patients were assumed to have started in the trial on the same day. Censoring was determined to take place after half a year (or 183 days). For patients who died before censoring, the censored survival time was equal to the uncensored survival time. For others, it was set at 183 days, with a flag indicating that the event had not taken place.

Analysis

We analysed both uncensored and censored data using six different Weibull regression models. In each case, the number of covariates (besides treatment) was increased one at a time, starting with a model containing no covariates (treatment T only) and ending with a model containing both treatment T and all five covariates (A, B, C, D and E). Hazard ratios from the different models were compared to illustrate the impact of non-collapsibility. Life expectancy was calculated per patient using the following equation:

$$E(Y) = \exp(\tfrac{-1}{p} X\beta)\Gamma(1 + \tfrac{1}{p}) \quad (3)$$

This was repeated for the situations with and without treatment. The mean life expectancy per treatment was the average of the individual life expectancies under that treatment.

7.3. Results

Synthesized outcomes

The mean time-to-event in the synthesized dataset was 257 days with treatment and 118 days without treatment. The difference in mean time-to-event time between the treatment groups was 138 days and the marginal hazard ratio was 0.484. In the censored dataset, 68% of the total sample died within the imaginary trial period of 183 days, while 44% of the treated and 17% of the untreated patients were censored.

Hazard ratios

The non-collapsibility of the hazard ratio is visible in figure 7.1. Here we can see that the ratios decreased as the number of covariates in the regression model increased. In the

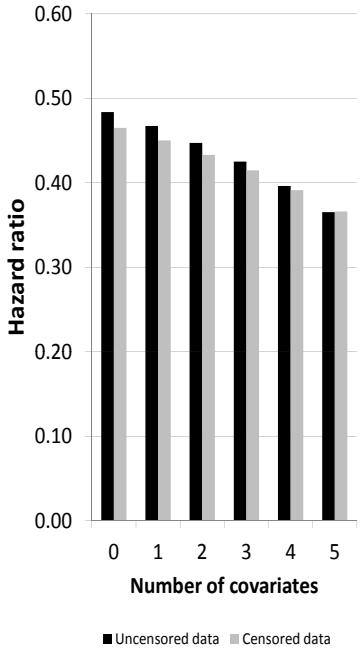

Figure 7.1 Hazard ratios for models with increasing numbers of covariates.

uncensored data, the hazard ratio was 0.48 when treatment was the sole covariate. The most extensive model yielded a hazard ratio of 0.37, which equals the exponent of the coefficient (–1) that was used in the synthesizing process.

When the models were estimated using censored data, the hazard ratios were consistently lower, except for the model that contained all covariates. In that model, the hazard ratios for censored and uncensored data were equal; the simpler models with fewer covariates overestimated the treatment effect.

Predicted mean time-to-event

Figures 7.2a, 7.2b and 7.2c present the life expectancy (mean projected time-to-event) for treated and untreated patients as well as the treatment effect (i.e., the difference in survival between the treated and untreated patients). In contrast to the hazard ratios, the results are quite stable for the uncensored data. All models led to estimates that were very close to the mean uncensored values in the dataset. The estimates for the treatment effect were equal to the marginal effect.

The analysis of the censored data led to very different results. The estimated life expectancy increased sharply for both treatment groups when more covariates were added to the regression model. It was underestimated when no or too few covariates were included in the model. The estimated increase in life expectancy from treatment was correct only when the most extensive model was used.

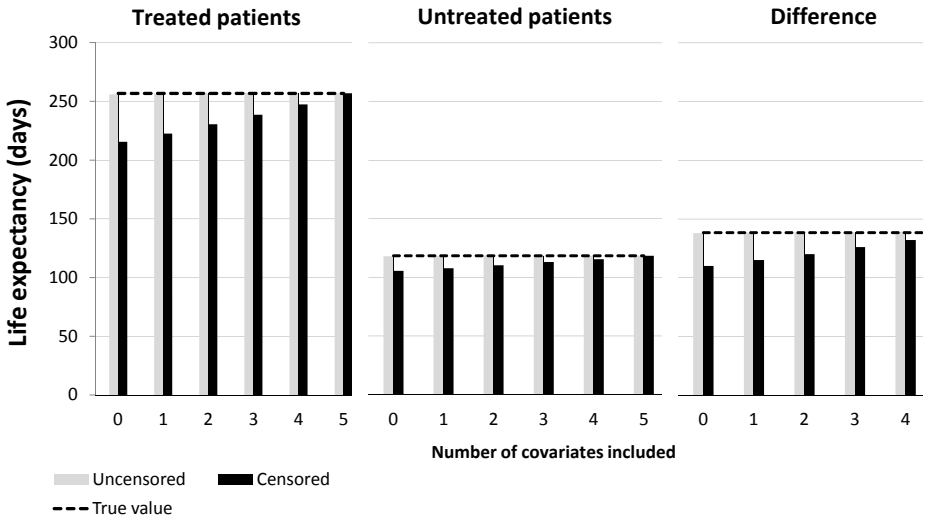

Figure 7.2 Estimated life expectancy for models with increasing numbers of covariates.

7.4 Discussion

The consequences of non-collapsibility and censoring are sometimes counterintuitive and often misunderstood. The analysis of the synthesized data in this chapter makes it possible to disentangle the impact of non-collapsibility and censoring on hazard ratios and predicted survival.

In the uncensored data, the hazard ratio decreased as covariates were added to the regression model. More importantly, the estimated hazard ratio equaled the correct ratio (i.e., the one used in the synthesizing process) only when all covariates were included. The differences in estimates from the different models can be completely ascribed to non-collapsibility since there was no confounding (since the baseline characteristics were perfectly balanced) and no censoring. In contrast to the hazard ratio results, the predicted mean survival and the corresponding treatment effect were stable and therefore collapsible.

If the one step in this study was to illustrate the role of non-collapsibility, another was to examine the impact of censoring in isolation. In the censored data, the estimates of the mean time-to-event increased – and thereby improved – as more covariates were added to the model. Only the fully specified model led to a correct estimate of the marginal treatment effect. This cannot be attributed to non-collapsibility, which was shown in the uncensored data only to affect hazard ratios. Since confounding can also be ruled out because of the balanced baseline characteristics, the differences in predicted survival across models must be caused by omitted-variable bias due to censoring. This bias cannot be seen directly in

the hazard ratios, but it becomes apparent when they are compared to the hazard ratios from the uncensored data. It could also be concluded from the results for projected survival, which showed that the model leads to biased estimates.

Paradoxically, the censoring of data led to an overestimate and an underestimate of the treatment effects in the models with omitted variables. Hazard ratios for similarly parameterized models were lower than those based on uncensored data – thereby overestimating the treatment effect – whereas the difference in projected mean time-to-event survival was smaller. This apparent inconsistency can be explained by the multiplicative character of the Weibull survival model. Life expectancy was underestimated for both treatment groups on the uncensored data. Treatment appeared to prolong life by a larger proportion in the censored data – as reflected in the lower hazard ratio. However, this proportion applied to a shorter period of time – projected mean survival.

The objective of this chapter was to clarify the impact of non-collapsibility and censoring on different outcomes of time-to-event models. While these issues have been investigated in several studies in recent decades, some confusion remains about their dangers. In the 1980s some authors warned readers about 'bias' from 'misspecification' in models [7,10], even when the only issue was non-collapsibility of the hazard ratio. As Hernán remarked in the context of odds ratios, they saw the difference between marginal and conditional results as 'bias' even when no bias existed, in the sense that it represents an association between treatment and prognostic factors that leads to incorrect estimates [28].

In stark contrast, some papers have been rather sanguine and overlooked the bias in the inconsistencies between estimates in the presence of censoring. According to these studies, the differences in results only meant that they were answers to different questions, i.e. about the hazard ratio given the values of prognostic factors (conditional) versus the average ratio over all strata combined (marginal) [11,29].

Most of these studies focused on semi-parametric Cox proportional hazards models. Gail et al. and Schmoor et al., who assessed the results of analysis on uncensored and independently censored data, separately [7,10]. Both found that the hazard ratios from differently specified models were inconsistent and considered this potentially problematic, whether the data were censored or not.

Possibly because incomplete follow-up is ubiquitous in trials with time-to-event as an outcome measure, three other studies only investigated the consistency of the hazard ratio when censoring was present [8,9,12]. Unfortunately, this made it impossible to distinguish between the impact of censoring and the impact of non-collapsibility.

Exponential models were studied by Gail et al., Chastang et al. and Ford et al. [7,11,30], who found that the hazard ratios were constant across models when the data were uncensored. This observation could be interpreted to mean that non-collapsibility does not play a role in exponential time-to-event models. Two of the studies also found, as one might expect, that the ratios changed across models when censoring was applied [7,11].

With regard to Weibull models, Lancaster demonstrated mathematically that hazard ratios based on non-censored data are non-collapsible and ascribed this to 'misspecification' [24]. He also showed that the estimates of mean time-to-event in unconfounded data did not depend on the number of covariates in the model. However, censoring was not considered in this study.

Like earlier studies, our study is based on datasets containing no confounding. However, the conclusions also have a bearing on data from observational studies, where treatment assignment does not occur at random. The non-collapsibility of the hazard ratio makes this measure unfit for estimating a marginal effect in unbalanced data: only a conditional model can address confounding. However, marginal effects can be estimated when mean time-to event is used as an outcome measure. When the results of a fitted conditional model are used to calculate the difference in mean projected outcome between the treatments for all patients, this yieds a marginal treatment effect. Similarly, under the assumption that the hazard ratio is constant across subgroups, the conditional model results can be used to calculate conditional, stratum-specific treatment effects.

Another argument for including covariates was provided by Lagakos and Schoenfeld, who pointed out that inclusion of strongly prognostic factors leads to increased power and prevents possible violation of the proportional hazards assumption [31].

In conclusion, non-collapsibility does not have to be a problem in time-to-event models, as long as confounding is properly addressed and follow-up is complete or censoring is random. Although inconsistent hazard ratios have often been blamed on bias and model misspecification, they can also reflect the choice between different specifications of strata in which the ratios are applicable. Different models can be simultaneously valid. When mean time-to-event is used as the outcome measure, the estimates are consistent across models. However, researchers should beware that the use of a parsimonious model will lead to biased estimates of the hazard ratio and mean time-to-event when censoring is present. This omitted-variable bias can even occur when analyzing RCT data.

7.5 References

1 Tappenden P, Chilcott J, Ward S, et al. Methodological issues in the economic analysis of cancer treatments. Eur J Cancer. 2006; 42: 2867-2875.

2 Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. Stat Sci. 1999; 14: 29-46.

3 Cummings P. The relative merits of risk ratios and odds ratios. Arch Pediatr Adolesc Med. 2009; 163: 438-445.

4 Martens EP, de Boer A, Pestman WR, et al. Comparing treatment effects after adjustment with multivariable Cox proportional hazards regression and propensity score methods. Pharmacoepidemiol Drug Saf. 2008; 17: 1-8.

5 Martens EP, Pestman WR, Klungel OH. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study (p n/a) by Peter C. Austin, Paul Grootendorst, Sharon-Lise T. Normand, Geoffrey M. Anderson, Statistics in Medicine, Published Online: 16 June 2006. DOI: 10.1002/sim.2618. Stat Med. 2007; 26: 3208-10.
6 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology (3 ed.). Philadelphia: Lippincott Williams & Wilkins, 2008.
7 Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1984; 71: 431-444.
8 Bretagnolle J, Huber-Carol C. Effects of omitting covariates in Cox's model for survival data. Scan J Statist. 1988; 15: 125-138.
9 Schumacher M, Olschewski M, Schmoor C. The impact of heterogeneity on the comparison of survival times. Stat Med. 1987; 6: 773-784.
10 Schmoor C, Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. Stat Med. 1997; 16: 225-237.
11 Chastang C, Byar D, Piantadosi S. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. Stat Med. 1988; 7: 1243-1255.
12 Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. Biometrika. 1986; 73: 363-369.
13 Castelli C, Combescure C, Foucher Y, Daures JP. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. Stat Med. 2007; 26: 5557-5571.
14 Cella D, Bushmakin AG, Cappelleri JC, et al. Baseline quality of life as a prognostic survival tool in patients receiving sunitinib for metastatic renal cell carcinoma. Br J Cancer. 2012; 106: 646-650.
15 Edidin AA, Ong KL, Lau E, Kurtz SM. Life expectancy following diagnosis of a vertebral compression fracture. Osteoporos Int. 2012;
16 Aballea S, Chancellor JV, Raikou M, et al. Cost-effectiveness analysis of oxaliplatin compared with 5-fluorouracil/leucovorin in adjuvant treatment of stage III colon cancer in the US. Cancer. 2007; 109: 1082-1089.
17 Heitjan DF, Kim CY, Li H. Bayesian estimation of cost-effectiveness from censored data. Stat Med. 2004; 23: 1297-1309.
18 Thomas KC, Nosyk B, Fisher CG, et al. Cost-effectiveness of surgery plus radiotherapy versus radiotherapy alone for metastatic epidural spinal cord compression. Int J Radiat Oncol Biol Phys. 2006; 66: 1212-1218.
19 Dewilde S, Bruggenjurgen B, Nienaber C, et al. Cost-effectiveness of adjunctive eptifibatide in patients undergoing coronary stenting in Germany. Eur J Health Econ. 2011;
20 Shiroiwa T, Fukuda T, Shimozuma K. Cost-effectiveness analysis of trastuzumab to treat HER2-positive advanced gastric cancer based on the randomised ToGA trial. Br J Cancer. 2011; 105: 1273-1278.
21 Soini EJ, Martikainen JA, Nousiainen T. Treatment of follicular non-Hodgkin's lymphoma with or without rituximab: cost-effectiveness and value of information based on a 5-year follow-up. Ann Oncol. 2011; 22: 1189-1197.
22 Lindgren P, Jonsson B, Duchane J. The cost-effectiveness of early cabergoline treatment compared to levodopa in Sweden. Eur J Health Econ. 2003; 4: 37-42.
23 Brennan A, Kharroubi SA. Expected value of sample information for Weibull survival data. Health Econ. 2007; 16: 1205-1225.
24 Lancaster T. Generalised residuals and heterogeneous duration models. With apllications to the Weibull Model. J Econometrics. 1985; 28: 155-169.

25 Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. J Clin Epidemiol. 1991; 44: 77-81.
26 Groenwold RH, Moons KG, Peelen LM, et al. Reporting of treatment effects from randomized trials: a plea for multivariable risk ratios. Contemp Clin Trials. 2011; 32: 399-402.
27 Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. Stat Med. 2002; 21: 2899-2908.
28 Hernan MA, Clayton D, Keiding N. The Simpson's paradox unraveled. Int J Epidemiol. 2011; 40: 780-785.
29 Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? Control Clin Trials. 1998; 19: 249-256.
30 Ford I, Norrie J, Ahmadi S. Model inconsistency, illustrated by the Cox proportional hazards model. Stat Med. 1995; 14: 735-746.
31 Lagakos SW, Schoenfeld DA. Properties of proportional-hazards score tests under misspecified regression models. Biometrics. 1984; 40: 1037-1048.

8

A propensity to get it right

A Monte Carlo simulation study comparing statistical methods to obtain correct cost-effectiveness estimates in observational studies

Objective Estimates of real-world cost-effectiveness are mostly based on observational data with non-random treatment assignments. Several methods exist to address the resulting confounding-by-indication, including regression and methods based on propensity scores (PS). This study examined the performance of these methods in the context of cost-effectiveness analysis. The PS-methods were: PS matching (kernel and 1-to-1), covariate adjustment using PS, inverse probability-of-treatment weighting (IPTW) and double robustness, each with several specifications.

Methods Thirty-eight adjustment approaches were compared using Monte Carlo simulations. In each simulation, four differently confounded samples were drawn from a synthesized population. Incremental survival time and costs were calculated using the results of Weibull survival and generalized linear regression. These regressions – with treatment as sole covariate or fully specified with all confounders – were performed directly or after applying a PS-method. Each approach was assessed on bias (systematic deviation from the true effect) and accuracy (proportion of simulated results within acceptable distance from the true values).

Results In estimates of the average treatment effect in the treated (ATT), kernel and 1-to-1 PS matching had similar bias and accuracy results. Regarding average treatment effects for the sample as a whole (ATE), double robustness and IPTW had the least bias and the best accuracy. Combining PS methods with fully specified regression models was most likely to lead to good results. PS covariate adjustment and regression without a PS method scored worst.

Conclusions PS methods are preferable to conventional regression for use in observational cost-effectiveness studies. Combining a PS method with fully specified regression should be considered for the analysis. Since no method is always superior, it is advised that sensitivity analyses with different techniques be performed.

Lucas M.A. Goossens, Chantal W.M. van Gils, Maiwenn J. Al,
Carin A. Uyl-de Groot, Miriam Koopman, W. Ken Redekop

Submitted for publication.

8.1 Introduction

Randomised controlled trials (RCTs) and cost-effectiveness analyses piggy-backed onto RCTs have long been viewed as the gold standard for estimating clinical treatment effects and cost-effectiveness of healthcare interventions. However, patients enrolled in clinical trials may not be representative of the patients seen in daily practice, to whom the therapy will eventually be applied. As an alternative to RCTs, the use of observational data allows investigators to estimate the (cost)-effectiveness of treatment in day-to-day practice, which should result in evidence that is more relevant to policy makers. Yet this approach has its own challenges. A key concern in observational studies is that treatment is not assigned randomly. This is likely to lead to systematic differences between treatment groups. In epidemiological terms, the association of treatment assignment and prognostic factors is known as confounding. If this confounding is not removed, estimates of the treatment are biased.

Traditionally, regression methods have been used for this purpose. The estimate of the treatment effect is adjusted by taking the effect of additional baseline covariates into account. More recently, methods based on PS have become increasingly popular. The propensity score (PS) is defined as a patients' probability of receiving a specific treatment assignment, based on certain observed characteristics of that patient. This score, which is usually derived from a logit or probit regression model, can be used in several ways to address confounding. The most popular PS method is covariate adjustment [1-3][1,3,4], in which the PS replaces the original baseline covariates in a regression model.

Other applications of the PS are aimed at removing the association between treatment and prognostic factors and to create an RCT-like design. These methods include matching, weighting and stratification on the PS. After the PS method has been applied successfully, an unbiased treatment effect can be estimated without additional adjustment.

Nevertheless, it has been argued that additional adjustment is useful and that PS methods should be used as a pre-processing stage before applying an endpoint regression model with the original baseline covariates [5,6]. Another approach in which PS and a fully specified regression model are combined is called double robustness [7].

All PS based methods are aimed at balancing baseline covariates across treatment groups. However, they may lead to conceptually different estimates of the treatment effect. The first of these effects focuses on the patients who were actually treated, the Average Treatment effect for Treated patients (ATT). This is the difference between the realised outcomes and the outcomes if they had not been treated. An ATT can result from PS matching methods, which adjust the composition of the control group to make it similar to the treatment group.

The second treatment effect is the Average Treatment Effect (ATE), which represents the difference in average outcomes between the hypothetical situations in which all patients are treated and in which none are treated. This effect is estimated when covariate adjustment,

inverse probability-of-treatment weighting, double robustness or conventional regression methods are used.

In a successfully randomised trial, the ATT and ATE are expected to be equal, since there should be no baseline differences between the treated and untreated patients. In an observational study, in contrast, this may not be the case. The ATT is different from the ATE when certain characteristics associated with a better treatment effect occur more frequently in one of the treatment groups.

PS methods have mostly been used in epidemiology and medicine, with a focus on clinical treatment effects. Most of the literature on the relative performance of different PS methods was produced in the same context [1,8,9]. Although examples of the application of PS methods also exist in cost-effectiveness analysis [10-18], they are still relatively scarce. Two methods, IPTW and double robustness, have not been used in observational cost-effectiveness studies. Furthermore, the properties of PS methods in this field have not been investigated. More knowledge about the value of these methods in cost-effectiveness analyses would be useful given certain important differences between health economic studies and effectiveness studies that have consequences for the application of PS methods.

Firstly, health economic studies need different effectiveness measures than the primary outcomes found in most effectiveness studies. Cost-effectiveness analysis requires natural units of health gain instead of purely statistical measures. If the outcome is dichotomous and analysed in a logistic regression model, health economists are interested in the number of events, not in the odds ratio, rate ratio or risk difference. In a survival model, the health economic outcome of interest is the increase in survival time (e.g. life-years gained) incremental number of day of survival; not the hazard ratio.

Secondly, cost data have different properties than data on clinical effectiveness. They are typically skewed and exhibit large individual variation.

Thirdly, health economic studies examine two outcomes simultaneously, incremental costs and incremental effects, and combine them into incremental cost-effectiveness ratios (ICERs). This intrinsic link between two outcomes may have consequences for the specification of adjustment or propensity-score models. Brookhart et al. and Austin et al. investigated the optimal specifications of PS models and concluded that a model should contain all variables that are prognostic for the outcome of interest, not merely confounders [19,20]. They also recommended that variables associated with treatment assignment but not with the outcome, should be omitted. However, the optimal model for the effect estimate is not necessarily equal to the optimal model for the cost estimate. Nevertheless, one model must be used for both outcomes; otherwise costs would be investigated in a different patient group than effects.

The objective of the current study was to evaluate the performance of several PS methods with varying specifications and conventional regression in the context of cost-effectiveness analysis. In order to be able to assess the difference between the estimates and the true

treatment effects (ATT and ATE), a source population was synthesized. Incremental effects, incremental costs and incremental cost-effectiveness ratios were calculated by using conventional regression, 1-to-1 PS matching, kernel PS matching, inverse probability-of-treatment weighting and double robustness.

8.2 Methods

All methods described in this section are summarized in a flowchart (figure 8.1).

Source population with potential outcomes

We performed a simulation study on a synthesized source population of 20,000 patients. In order to obtain reality-like data, this source population dataset reflected the variable distributions and covariance structures of two Dutch empirical studies on combination therapy (from now on: treated) versus sequential therapy (from now on: untreated) in stage-IV colorectal carcinoma [21].

The synthesizing process was based on the Neyman-Rubin-Holland causal model [22-24], which assumes the existence of potential (or counterfactual) outcomes for each treatment option for all patients. Only one of these potential outcomes materializes: the one for the treatment that is actually received. The other potential outcome remains unobserved, or counterfactual. All patients in the synthesized dataset had potential outcomes on health (survival time) and healthcare expenditure for both treatment options. Which outcomes were to be observable, depended on the treatment assignment.

The 'true' individual treatment effect was defined as the difference between the potential outcomes for each treatment. The ATE was the mean of the individual treatments effects in a sample. The ATT was the mean individual treatment effect across patients who were assigned to the new treatment.

Synthesis of survival time and healthcare expenditure

A detailed description of the synthesizing process can be found in Appendix A. In short, survival time was assigned based on a Weibull survival regression of the empirical trial data. The regression results were combined with the synthesized covariates to construct individual survival functions per treatment option. Random drawings from a uniform distribution (0-1) determined the point on the curves for each treatment option until which the patient survived. Patients who responded relatively well (compared to other patients in this treatment group) to the new treatment, also responded relatively well to the conventional treatment. The treatment effect was varied across patients by performing drawings from a distribution around the treatment coefficient from the regression analysis. For approximately 10% of patients, the effect was negative.

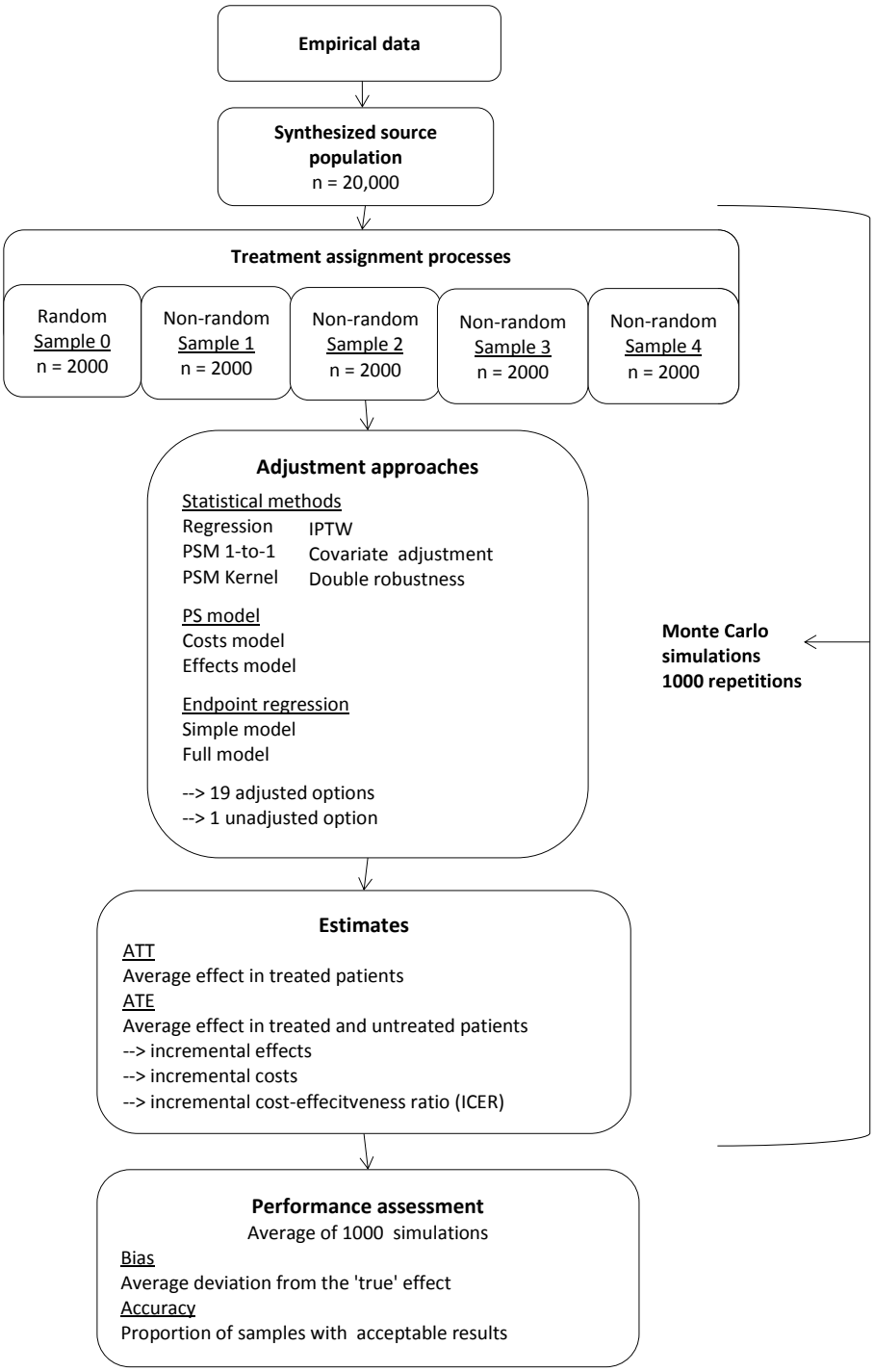

Figure 8.1. Flowchart of the method steps .

A generalized linear model with a power link was fit on the cost data in order to estimate a model with which data could be synthesized. Using the coefficients from this model, predicted mean costs and gamma distributions around this mean were calculated for each patient, from which the individual patient's costs for each treatment were drawn.

Treatment assignment process

This source population was used to draw samples, in which treatment assignment could be associated with certain baseline characteristics. Treatment was assigned in four processes, besides randomisation. A detailed description of these processes can be found in Appendix B.

The mean distribution of the baseline characteristics and treatment effects for samples resulting from these processes are summarized in table 8.1. The first treatment assignment process was based on covariates that were only in the synthesizing model for costs, not for effects: older and female patients were less likely to receive the new treatment. The second process assigned the new treatment disproportionately to patients with a relatively good prognosis and low projected costs. In the third process, the new treatment was assigned mostly to patients with an unfavourable prognosis. The fourth process was based on combination of favourable and unfavourable risk factors.

Table 8.1 Drawing of samples with non-random treatment assignment

	Impact on probability to receive treatment	Non-random treatment assignment process[1]			
		1	2	3	4
Covariates with negative impact on survival					
Number of metastasis > 1	Positive			●	●
Unresected tumour	Positive			●	●
Abnormal LDH[2]	Positive			●	●
Performance score > 0	Negative		●		●
Abnormal AF[3]	Negative		●		●
Higher WBC[4]	Negative		●		●
Covariates associated with higher costs					
Higher Age	Negative	●	●		●
Female sex	Negative	●			●
Female sex	Positive		●		

[1] Sample 0 is not included since this treatment assignment process was random.

[2] LDH = Lactate dehydrogenase level in blood

[3] AF = Alcalic phosphatase level in blood

[4] White blood cell count

Monte Carlo simulations

For each of the four treatment assignment processes as well as for randomisation, 1000 samples were drawn from the source population. Each set contained 1000 patients who received the new intervention and 1000 patients who were treated according to the conventional treatment.

Statistical methods

Six methods were compared.

(1) **Conventional regression**. A fully specified Weibull regression model was used to estimate the survival gains. The covariates were the prognostic factors that were used for synthesizing survival time. Costs were analysed in a generalized linear model with a log link and Gaussian variance function, with days of survival time (linear and squared) and the interaction terms of survival time (linear and squared) with treatment, age and sex. These covariates were also used for synthesizing costs data.

(2) **PSM 1-to-1**: PS matching (with replacement and common support requirement). Propensity scores were calculated by fitting a probit model. The closest matching untreated patient was selected for each treated patient, based on the differences between their respective PS. Untreated patients could be matched to more than one treated patient. Matching only took place for treated patients with common support: those with a PS in the range of the scores of the untreated patients. After matching, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(3) **PSM kernel**: PS matching with kernel smoothing. Each treated patient was matched to all untreated patients, but the latter were weighted according to their similarity to the particular treated patient. The combined weights of the matches equalled 1. Weighting was based on the distance between the PS scores and the Epanechnikov kernel function [25,26]. After the matching procedure, analysis was similar to the previous method.

(4) **PS covariate adjustment** regression analysis. Propensity scores were used as covariates in the Weibull model and GLM for survival and costs.

(5) **IPTW**: inverse probability of treatment weighting. [27,28] Treated patients were weighted by the inverse of their PS (the probability of being treated), controls were weighted by the inverse of 1 minus the PS (or the propensity of being untreated). After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(6) **Double robustness**. This technique combines regression and weighting by the PS in one equation [7]. Results should be unbiased if either the regression model or the propensity-score model is correct.

Model specification

All methods using PS (methods 2-6) were based on a model that contained only prognostic factors for survival (the effects PS model) or, additionally, covariates that predicted costs (the costs PS model).

Furthermore, the Weibull regression models under methods 2 through 5 were either fully specified (full endpoint model) or contained only a variable for treatment (simple endpoint model, which for PS covariate adjustment also contained the PS). The GLMs contained all covariates from the synthesizing model (full model) or only survival time (linear and squared), treatment and their interactions (simple model).

Calculation of incremental effects and costs

The mean survival for each treatment option was calculated as follows. For each patient, the expected survival time was projected for each option, based on their baseline covariates, according to this equation:

$$E(Y) = \exp(\tfrac{-1}{p} X\beta)\Gamma(1 + \tfrac{1}{p})$$

in which p is the shape parameter from the Weibull regression and $X\beta$ denotes the combination of regression coefficients and corresponding baseline characteristics, including the treatment. The projected costs were the fitted values of the GLM for costs, based on the projected survival.

Also unadjusted results were calculated via the Weibull and GLM models, without taking patient characteristics into account.

To calculate ATEs, the average predicted effects and costs over all patients were calculated for each treatment option. Incremental effects, costs and cost-effectiveness ratios were derived from these averages. When calculating the ATT, this process was applied to treated patients only.

Assessment of performance

Two criteria were used to assess the performance of each methods and specification: bias and accuracy. The bias was defined as the average positive or negative deviation of the estimate from the 'true' effect. This true effect was calculated for each of the 1000 datasets. Degree of bias was expressed as a percentage of the true effect.

Accuracy was defined as the proportion of samples with an effect estimate within an acceptable range around the true effect. These ranges were +/– 20% for effectiveness and costs, and €5000 for incremental cost-effectiveness ratios.

8.3 Results

Baseline characteristics and outcomes

Table 8.2 summarizes the distribution of the baseline characteristics and health, costs and cost-effectiveness outcomes in the synthesized source population and the mean results

Table 8.2 Baseline characteristics, health, costs and cost-effectiveness outcomes

			Treatment assignment process									
	Source population		Random		Non-random							
			0		1		2		3		4	
Untreated (U)/ Treated (T)	U=T		U	T	U	T	U	T	U	T	U	T
Baseline characteristics												
Female	38%		38%	38%	46%	30%	31%	47%	38%	38%	41%	35%
Age, mean	61		61	61	66	56	65	55	61	61	66	56
Performance Score > 0	46%		46%	46%	46%	45%	53%	35%	40%	47%	52%	40%
Abnormal AF	63%		63%	63%	59%	67%	69%	54%	51%	66%	64%	62%
WBC, mean	89		89	89	88	91	91	87	85	90	90	89
Number of metastasis > 1	55%		55%	55%	54%	55%	54%	57%	36%	60%	47%	62%
Unresected primary tumour	37%		37%	37%	37%	37%	38%	35%	23%	40%	30%	42%
Abnormal LDH	45%		44%	45%	41%	48%	46%	43%	20%	51%	35%	52%
Health outcomes												
Mean survival (days)	531	664	532	664	533	665	498	725	634	634	540	656
Unadjusted incremental effect			133		132		227		0		116	
True incremental effect ATT	133		133		133		145		126		131	
True incremental effect ATE	133		133		133		134		142		133	
Cost outcomes (euros)												
Mean total costs	23,016	34,612	23,222	36,111	22,284	36,976	21,973	38,109	25,432	35,337	22,366	36,348
Unadjusted incremental costs			12,889		14,692		16,135		9,905		13,982	
True incremental costs ATT	11,596		11,560		11,071		10,186		11,429		10,743	
True incremental costs ATE	11,596		11,579		11,604		11,381		11,797		11,648	
Cost-effectiveness outcomes (euros)												
Unadjusted mean ICER			36,547		41,993		26,119		-1,223,872		46,290	
Lowest value from iterations			22,242		26,495		19,006		-671,860,412		28,610	
Highest value from iterations			70,345		90,013		36,010		19,988,728		171,586	
True mean ICER ATT	31,845		31,921		30,525		25,768		33,189		30,044	
Lowest value from iterations			24,822		24,133		19,021		25,802		22,642	
Highest value from iterations			41,583		37,564		32,794		44,496		38,614	
True mean ICER ATE	31,845		31,907		31,974		30,983		30,309		32,088	
Lowest value from iterations			27,391		26,546		25,425		24,441		25,593	
Highest value from iterations			36,712		38,418		35,654		35,900		38,075	

Abbreviations: AF, alcalic phosphatase level in blood; WBC, white blood cell count; LDH, lactate dehydrogenase level in blood; ATT, average treatment effect in the treated patients; ATE, average treatment effect; ICER, incremental cost-effectiveness ratio.

over all iterations for the random and non-random treatment assignment processes. In the source population all patients were, counterfactually, treated twice. Treatment groups were not distinguished. This means that there was no difference between ATT and ATE and between true and unadjusted results. The incremental effect was 133 days and the incremental costs were €11,596, which resulted in an ICER of €31,845. As expected, the mean results over 1000 iterations for treatment assignment process 0 were almost equal to the means in the source population.

For treatment assignment process 1, the mean unadjusted estimate of incremental effects was unbiased, but the costs and ICER were overestimated. For process 2, incremental effects and costs were overestimated, while the ICER was underestimated. In the unadjusted analysis of treatment assignment process 3, incremental effects nearly disappeared on average, while costs were underestimated. This led to a very large negative mean ICER, implying that the new treatment was dominated. For process 4, incremental effects were underestimated and incremental costs were overestimated. The mean ICER was an overestimation.

Bias and accuracy

The performance of all methods in all treatment assignment processes is summarized in tables 8.3a to 8.5b. The methods, including their specifications, are ranked for bias and accuracy based on their performance over the four non-random treatment assignment processes combined.

Most methods succeeded in removing most of the bias and improving the accuracy in effectiveness estimates compared to the unadjusted analyses (see tables 8.3a and 8.3b). However, for treatment assignment processes 0 and 1, many adjusted analyses introduced or added bias and reduced accuracy of the estimates.

With regard to costs, most adjustments improved the estimates. The same was seen in cost-effectiveness results, with the exception of treatment assignment process 2, where the unadjusted results were almost unbiased and quite accurate. However, this was due to simultaneous biases on the estimates of costs and effects under this treatment assignment process, which cancelled each other out in the cost-effectiveness ratio.

The proportion of accurate estimates varied greatly across methods and treatment assignment processes: from 37% to 87% for effectiveness, from 14% to 100% for costs and from 26% to 82% for cost-effectiveness.

Endpoint regression model

Fully specified endpoint regression models generally performed somewhat better than simple models in producing unbiased and accurate estimates of effectiveness ATTs, with regards to bias as well as accuracy (table 8.3a). The differences were more pronounced in for costs and for both ATTs and ATEs estimates of cost-effectiveness (tables 8.4a, 8.5a and 8.5b). However, for ATEs of effectiveness there were no clear differences, while, simple models generally performed better with regard to bias on ATEs for costs (tables 8.3b and 8.4b).

Table 8.3a ATT effectiveness results

BIAS								ACCURACY							
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process				
			Random	Non-random							Random	Non-random			
			0	1	2	3	4				0	1	2	3	4
Unadjusted			0%	0%	62%	-100%	-11%	Unadjusted			77%	76%	1%	0%	66%
Regression		Full	2%	0%	1%	1%	0%	Regression		Full	86%	81%	81%	87%	80%
IPTW	Costs	Full	2%	3%	2%	3%	3%	IPTW	Effects	Full	86%	72%	70%	86%	65%
IPTW	Effects	Full	2%	3%	3%	3%	3%	IPTW	Costs	Full	86%	67%	70%	85%	62%
PSM 1-to-1	Costs	Simple	0%	2%	9%	2%	-5%	PSM kernel	Effects	Full	85%	73%	71%	73%	65%
PSM kernel	Costs	Simple	-3%	2%	10%	-1%	-7%	PSM kernel	Costs	Full	86%	68%	70%	72%	64%
PSM kernel	Effects	Full	2%	3%	5%	7%	5%	PSM kernel	Effects	Simple	78%	68%	57%	84%	61%
Double R	Effects	Full	0%	3%	-5%	13%	0%	PSM kernel	Costs	Simple	80%	62%	56%	84%	59%
Cov adjust	Costs	Simple	0%	4%	7%	8%	-3%	Double R	Effects	Full	84%	68%	68%	66%	57%
IPTW	Effects	Simple	0%	3%	-5%	13%	1%	IPTW	Effects	Simple	83%	68%	68%	65%	57%
PSM kernel	Effects	Simple	-4%	-2%	12%	-1%	-8%	IPTW	Costs	Simple	83%	60%	65%	65%	56%
PSM 1-to-1	Effects	Simple	0%	-2%	11%	2%	-9%	Double R	Costs	Full	84%	60%	63%	66%	55%
Double R	Costs	Full	0%	4%	-7%	13%	-1%	PSM 1-to-1	Effects	Full	72%	59%	57%	57%	56%
IPTW	Costs	Simple	0%	4%	-7%	13%	0%	PSM 1-to-1	Effects	Simple	67%	57%	48%	71%	51%
PSM kernel	Costs	Full	2%	6%	5%	7%	7%	PSM 1-to-1	Costs	Full	72%	55%	57%	58%	52%
Cov adjust	Effects	Simple	1%	-2%	9%	10%	-10%	Cov adjust	Costs	Full	72%	55%	56%	58%	51%
PSM 1-to-1	Effects	Full	4%	6%	9%	11%	8%	PSM 1-to-1	Costs	Simple	65%	52%	45%	69%	50%
PSM 1-to-1	Costs	Full	3%	9%	8%	11%	11%	Cov adjust	Costs	Simple	65%	51%	49%	59%	49%
Cov adjust	Costs	Full	4%	10%	8%	11%	12%	Cov adjust	Effects	Full	61%	50%	49%	45%	49%
Cov adjust	Effects	Full	5%	9%	10%	16%	11%	Cov adjust	Effects	Simple	57%	45%	43%	49%	41%

ATE results are shaded.

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Table 8.3b ATE effectiveness results

BIAS								ACCURACY								
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process					
			Random	Non-random							Random	Non-random				
			0	1	2	3	4				0	1	2	3	4	
Unadjusted			0%	0%	70%	-100%	-13%	Unadjusted			70%	70%	0%	0%	61%	
Regression		Full	2%	0%	1%	1%	0%	Regression		Full	83%	78%	78%	77%	74%	
Double R	Costs	Full	0%	4%	1%	0%	-2%	IPTW	Effects	Full	83%	70%	68%	74%	61%	
IPTW	Costs	Simple	1%	4%	1%	1%	-1%	IPTW	Costs	Full	83%	63%	66%	74%	59%	
IPTW	Effects	Simple	1%	4%	3%	1%	-1%	Double R	Effects	Full	77%	65%	62%	72%	54%	
Double R	Effects	Full	0%	3%	3%	0%	-1%	IPTW	Effects	Simple	76%	65%	62%	71%	55%	
IPTW	Costs	Full	2%	3%	2%	3%	3%	PSM kernel	Effects	Full	83%	70%	67%	50%	61%	
IPTW	Effects	Full	2%	3%	3%	3%	3%	PSM kernel	Effects	Simple	72%	62%	48%	80%	56%	
PSM 1-to-1	Costs	Simple	0%	2%	10%	2%	-5%	IPTW	Costs	Simple	76%	58%	60%	72%	54%	
Cov adjust	Costs	Simple	1%	6%	8%	7%	0%	Double R	Costs	Full	77%	58%	59%	72%	54%	
PSM kernel	Costs	Simple	-3%	2%	12%	-1%	-7%	PSM kernel	Costs	Simple	73%	57%	48%	80%	56%	
PSM kernel	Effects	Simple	-3%	-2%	14%	-1%	-9%	PSM kernel	Costs	Full	82%	63%	64%	49%	59%	
Cov adjust	Effects	Simple	1%	1%	9%	10%	-6%	Cov adjust	Costs	Full	71%	52%	53%	50%	48%	
PSM 1-to-1	Effects	Simple	1%	-2%	13%	2%	-9%	PSM 1-to-1	Effects	Full	69%	56%	53%	38%	50%	
PSM kernel	Effects	Full	2%	2%	0%	18%	6%	PSM 1-to-1	Effects	Simple	59%	51%	37%	64%	45%	
PSM kernel	Costs	Full	2%	6%	-1%	18%	7%	PSM 1-to-1	Costs	Simple	62%	48%	39%	65%	45%	
PSM 1-to-1	Effects	Full	5%	6%	3%	23%	9%	PSM 1-to-1	Costs	Full	66%	52%	53%	38%	45%	
Cov adjust	Costs	Full	5%	11%	8%	12%	13%	Cov adjust	Costs	Simple	64%	46%	46%	51%	45%	
PSM 1-to-1	Costs	Full	5%	9%	1%	22%	12%	Cov adjust	Effects	Full	61%	47%	47%	39%	47%	
Cov adjust	Effects	Full	6%	9%	10%	16%	12%	Cov adjust	Effects	Simple	53%	43%	41%	42%	38%	

ATT results are shaded

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Table 8.4a ATT costs results

BIAS								ACCURACY							
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process				
			Random	Non-random							Random	Non-random			
			0	1	2	3	4				0	1	2	3	4
Unadjusted			12%	33%	54%	-13%	30%	Unadjusted			75%	16%	0%	70%	22%
Double R	Costs	Full	0%	3%	11%	4%	4%	Double R	Costs	Full	100%	92%	70%	98%	81%
PSM kernel	Effects	Full	7%	4%	17%	6%	6%	PSM kernel	Costs	Full	90%	90%	59%	92%	85%
PSM kernel	Costs	Full	7%	5%	15%	6%	7%	PSM kernel	Effects	Full	90%	91%	55%	92%	87%
PSM 1-to-1	Effects	Full	7%	4%	17%	6%	6%	PSM 1-to-1	Costs	Full	88%	84%	56%	87%	78%
PSM 1-to-1	Costs	Full	7%	6%	16%	6%	8%	PSM 1-to-1	Effects	Full	85%	84%	52%	86%	80%
Double R	Effects	Full	0%	20%	3%	4%	16%	Double R	Effects	Full	99%	52%	84%	98%	62%
PSM 1-to-1	Costs	Simple	11%	8%	23%	8%	7%	PSM kernel	Costs	Simple	77%	81%	35%	86%	78%
PSM kernel	Costs	Simple	11%	9%	25%	9%	8%	PSM 1-to-1	Costs	Simple	73%	76%	37%	82%	72%
PSM 1-to-1	Effects	Simple	11%	13%	21%	9%	10%	PSM kernel	Effects	Simple	77%	65%	40%	86%	71%
PSM kernel	Effects	Simple	11%	15%	22%	9%	12%	IPTW	Costs	Full	90%	70%	25%	97%	61%
IPTW	Costs	Full	7%	14%	28%	1%	17%	PSM 1-to-1	Effects	Simple	72%	66%	41%	79%	67%
IPTW	Effects	Full	7%	15%	28%	0%	16%	IPTW	Effects	Full	90%	70%	23%	98%	63%
Regression		Full	7%	17%	27%	3%	18%	Regression		Full	90%	63%	22%	97%	60%
Cov adjust	Costs	Full	8%	17%	26%	8%	18%	Cov adjust	Costs	Full	88%	62%	28%	87%	55%
Cov adjust	Effects	Full	8%	18%	26%	8%	19%	Cov adjust	Costs	Simple	74%	73%	14%	74%	63%
IPTW	Costs	Simple	12%	14%	23%	21%	16%	Cov adjust	Effects	Full	85%	59%	29%	82%	53%
Cov adjust	Costs	Simple	12%	12%	35%	12%	15%	IPTW	Costs	Simple	75%	67%	39%	45%	58%
IPTW	Effects	Simple	12%	23%	18%	21%	23%	IPTW	Effects	Simple	75%	45%	50%	44%	46%
Cov adjust	Effects	Simple	12%	24%	28%	12%	22%	Cov adjust	Effects	Simple	72%	40%	26%	71%	45%

ATE results are shaded

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Table 8.4b ATE costs results

BIAS								ACCURACY								
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process					
			Random	Non-random							Random	Non-random				
			0	1	2	3	4				0	1	2	3	4	
Unadjusted			-1%	17%	22%	-16%	11%	Unadjusted			94%	62%	44%	66%	81%	
Double R	Costs	Full	0%	-1%	1%	0%	-4%	Double R	Costs	Full	100%	93%	90%	100%	85%	
PSM 1-to-1	Effects	Simple	-2%	7%	0%	-2%	4%	Cov adjust	Costs	Simple	95%	82%	89%	89%	85%	
IPTW	Costs	Simple	0%	-4%	-1%	4%	-5%	Regression		Full	91%	88%	90%	74%	91%	
PSM 1-to-1	Costs	Simple	-1%	-5%	6%	-3%	-4%	IPTW	Effects	Full	91%	90%	86%	82%	85%	
IPTW	Effects	Simple	0%	5%	-7%	5%	2%	Cov adjust	Effects	Simple	94%	84%	85%	87%	86%	
Cov adjust	Costs	Simple	-1%	-9%	3%	6%	-6%	IPTW	Effects	Simple	95%	86%	81%	91%	83%	
Cov adjust	Effects	Simple	0%	7%	-6%	7%	5%	Double R	Effects	Full	100%	71%	85%	99%	84%	
IPTW	Effects	Full	7%	4%	6%	11%	5%	IPTW	Costs	Simple	95%	82%	86%	91%	79%	
IPTW	Costs	Full	7%	4%	7%	12%	5%	IPTW	Costs	Full	91%	87%	83%	81%	84%	
Double R	Effects	Full	0%	15%	-7%	1%	7%	PSM kernel	Costs	Full	92%	85%	68%	91%	81%	
Regression		Full	7%	6%	5%	14%	6%	Cov adjust	Costs	Full	76%	92%	76%	76%	79%	
PSM kernel	Costs	Full	7%	7%	14%	7%	9%	Cov adjust	Effects	Full	75%	89%	79%	75%	79%	
PSM 1-to-1	Costs	Full	7%	3%	6%	22%	8%	PSM kernel	Effects	Full	91%	79%	73%	91%	77%	
Cov adjust	Costs	Full	-14%	-4%	-12%	-14%	-11%	PSM 1-to-1	Costs	Simple	88%	77%	73%	88%	76%	
PSM kernel	Effects	Full	7%	11%	12%	7%	11%	PSM 1-to-1	Effects	Simple	90%	73%	75%	90%	77%	
Cov adjust	Effects	Full	-14%	-5%	-11%	-14%	-12%	PSM 1-to-1	Costs	Full	87%	84%	82%	44%	76%	
PSM 1-to-1	Effects	Full	7%	7%	4%	22%	10%	PSM 1-to-1	Effects	Full	87%	82%	83%	42%	76%	
PSM kernel	Costs	Simple	11%	11%	23%	9%	10%	PSM kernel	Costs	Simple	82%	75%	40%	85%	76%	
PSM kernel	Effects	Simple	11%	21%	18%	9%	17%	PSM kernel	Effects	Simple	82%	45%	55%	85%	59%	

ATT results are shaded

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Table 8.5a ATT cost-effectiveness results

BIAS								ACCURACY							
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process				
			Random	Non-random							Random	Non-random			
			0	1	2	3	4				0	1	2	3	4
Unadjusted			15%	38%	1%	-3800%	54%	Unadjusted			59%	17%	93%	0%	6%
PSM 1-to-1	Costs	Full	6%	2%	14%	-1%	1%	PSM kernel	Effects	Full	74%	72%	60%	74%	70%
PSM 1-to-1	Effects	Full	6%	2%	13%	-1%	3%	PSM kernel	Costs	Full	73%	70%	60%	73%	69%
PSM kernel	Costs	Full	7%	3%	14%	1%	3%	PSM 1-to-1	Effects	Full	68%	64%	61%	60%	63%
PSM kernel	Effects	Full	7%	4%	14%	1%	4%	PSM 1-to-1	Costs	Full	65%	59%	63%	61%	59%
Double R	Costs	Full	3%	5%	22%	-6%	12%	Double R	Costs	Full	82%	65%	49%	70%	58%
Cov adjust	Costs	Full	7%	12%	21%	0%	12%	Double R	Effects	Full	82%	49%	71%	71%	50%
Cov adjust	Effects	Full	8%	20%	21%	-2%	13%	Cov adjust	Costs	Full	69%	61%	50%	63%	60%
IPTW	Effects	Full	7%	14%	25%	-1%	18%	IPTW	Effects	Full	74%	58%	34%	75%	57%
IPTW	Costs	Full	7%	15%	27%	-1%	18%	IPTW	Costs	Full	74%	58%	33%	76%	57%
Double R	Effects	Full	3%	21%	10%	-6%	24%	Cov adjust	Effects	Full	65%	58%	54%	54%	55%
PSM kernel	Costs	Simple	18%	11%	20%	11%	25%	PSM kernel	Costs	Simple	51%	61%	51%	60%	49%
Regression		Full	7%	20%	26%	4%	20%	PSM 1-to-1	Costs	Simple	51%	53%	52%	58%	46%
PSM 1-to-1	Effects	Simple	17%	25%	22%	10%	-14%	PSM kernel	Effects	Simple	51%	47%	61%	59%	42%
PSM kernel	Effects	Simple	18%	23%	15%	12%	31%	Regression		Full	73%	48%	31%	75%	48%
IPTW	Costs	Simple	14%	15%	34%	9%	25%	PSM 1-to-1	Effects	Simple	54%	46%	55%	55%	40%
PSM 1-to-1	Costs	Simple	17%	20%	23%	10%	30%	IPTW	Costs	Simple	59%	54%	28%	65%	46%
IPTW	Effects	Simple	14%	23%	26%	9%	31%	Cov adjust	Costs	Simple	53%	53%	36%	56%	42%
Cov adjust	Costs	Simple	17%	21%	34%	10%	51%	IPTW	Effects	Simple	59%	44%	38%	63%	38%
Cov adjust	Effects	Simple	24%	42%	25%	10%	78%	Cov adjust	Effects	Simple	51%	36%	48%	51%	29%

ATE results are shaded

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Table 8.5b ATE cost-effectiveness results

BIAS								ACCURACY								
Method	PS model	Endpoint model	Treatment assignment process					Method	PS model	Endpoint model	Treatment assignment process					
			Random	Non-random							Random	Non-random				
			0	1	2	3	4				0	1	2	3	4	
Unadjusted			2%	21%	-27%	-4157%	34%	Unadjusted			57%	46%	10%	0%	32%	
PSM kernel	Effects	Full	6%	0%	4%	-2%	-4%	PSM kernel	Effects	Full	72%	68%	68%	74%	59%	
Double R	Costs	Full	2%	0%	4%	2%	5%	IPTW	Effects	Full	71%	65%	66%	70%	61%	
PSM kernel	Costs	Full	6%	-2%	4%	-2%	-4%	PSM kernel	Costs	Full	72%	63%	66%	74%	58%	
IPTW	Costs	Simple	1%	-3%	4%	6%	4%	Regression		Full	72%	65%	69%	60%	66%	
PSM 1-to-1	Costs	Simple	4%	1%	-7%	7%	8%	Double R	Costs	Full	76%	59%	65%	77%	53%	
IPTW	Effects	Full	6%	4%	6%	10%	6%	IPTW	Costs	Full	71%	62%	64%	68%	59%	
PSM 1-to-1	Effects	Full	4%	-7%	-5%	10%	-7%	Double R	Effects	Full	76%	58%	60%	75%	56%	
IPTW	Effects	Simple	1%	6%	-5%	7%	11%	IPTW	Effects	Simple	64%	54%	51%	63%	48%	
PSM 1-to-1	Costs	Full	5%	-7%	-4%	10%	-8%	PSM 1-to-1	Effects	Full	66%	48%	54%	62%	49%	
IPTW	Costs	Full	6%	5%	8%	11%	7%	PSM kernel	Costs	Simple	54%	58%	59%	44%	50%	
Cov adjust	Costs	Simple	3%	-2%	4%	6%	23%	PSM 1-to-1	Costs	Full	67%	47%	52%	66%	44%	
Regression		Full	6%	9%	6%	15%	8%	IPTW	Costs	Simple	63%	47%	51%	64%	47%	
Double R	Effects	Full	2%	15%	-7%	3%	16%	PSM kernel	Effects	Simple	52%	52%	60%	43%	46%	
PSM kernel	Costs	Simple	17%	6%	3%	22%	17%	Cov adjust	Costs	Simple	53%	35%	44%	55%	37%	
Cov adjust	Effects	Full	-15%	-3%	-12%	-20%	-16%	Cov adjust	Effects	Simple	46%	42%	35%	48%	36%	
Cov adjust	Costs	Full	-16%	-7%	-13%	-19%	-16%	PSM 1-to-1	Effects	Simple	49%	40%	32%	52%	36%	
PSM 1-to-1	Effects	Simple	3%	7%	-9%	8%	-32%	PSM 1-to-1	Costs	Simple	51%	36%	34%	55%	34%	
PSM kernel	Effects	Simple	17%	17%	-2%	22%	23%	Cov adjust	Costs	Full	38%	41%	37%	32%	30%	
Cov adjust	Effects	Simple	7%	22%	-7%	7%	51%	Cov adjust	Effects	Full	36%	36%	32%	26%	29%	

ATT results are shaded

Abbreviations: ATT, average treatment effect in the treated patients; ATE, average treatment effect; PS, propensity score IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching; Double R, double robustness; Cov adjust, propensity score covariate adjustment.

Propensity-score model

For the estimation of a difference in effects, no clear difference between both PS models was seen (tables 8.3a and 8.3b). For costs and cost-effectiveness, the costs model for the PS performed at least as well as the effects model and often better. This was especially the case when the application of a PS method was followed by a simple regression (with only treatment as a covariate). When fully specified endpoint regression models were applied, the specification of the propensity-score model did not play an important role. The findings were similar for ATT and ATE.

Statistical method

For the estimation of ATTs of effectiveness, ATT methods and ATE methods achieved similar results (see table 8.3a). For ATEs of effectiveness, ATE methods performed better than ATT methods (table 8.3b). For costs, ATT methods generally had less bias and more accuracy in the estimation of ATTs (table 8.4a), while ATE methods generated better results on ATEs (8.4b). The same distinction was seen for cost-effectiveness (tables 8.5a and 8.5b).

Conventional regression achieved the best results in the estimations of differences in effects. For costs and cost-effectiveness, all ATE methods performed equally well, except for covariate adjustment by the PS, which often led to biased and inaccurate estimates. The performance of both of the matching techniques, 1-to-1 and kernel matching was similar.

8.4 Discussion

Main findings

The objective of this study was to evaluate the performance of different propensity-score statistical methods and conventional regression in the context of an economic evaluation based on observational time-to-event data. Incremental health effects, incremental costs and incremental cost-effectiveness ratios in a synthesized dataset were estimated using conventional regression, 1-to-1 PS matching, kernel PS matching, inverse probability-of-treatment weighting and double robustness. These methods were compared in their ability to reduce bias and maximise accuracy. Different model specifications were considered: costs or effects model for calculating PS, full or simple regression models for calculating final outcomes.

Our main findings are as follows. First, no method worked best in all circumstances. The optimal choice was not constant for different outcome measures and different treatment assignment processes.

Secondly, all methods failed to achieve accurate estimates in a substantial proportion of samples, even when treatment assignment was random. It was most pronounced for cost-effectiveness ratios. Deviations in either cost or effect estimates are amplified when

the incremental cost effectiveness ratio is calculated, especially when they work in different directions. However, when they work in the same direction, they can cancel each other out somewhat: that is the combination of an overestimated incremental effect with overestimate incremental costs may lead to an unexpectedly accurate ICER.

Third, estimates were mostly better when propensity-score methods were followed by a fully specified regression adjustment. This was most pronounced for estimating costs and cost-effectiveness.

Fourth, propensity-score models containing all covariates that predicted costs (the costs PS model), performed at least as good as the models that only contained prognostic factors for survival (the effects PS model), if not better. Several authors have argued that models to estimate PS should include all covariates that are associated with the outcome – not only confounders – while covariates that are not prognostic should be excluded [19,20].

We have seen that these rules may conflict when performing cost-effectiveness analyses. Some variables would need to be included in the propensity-score model for cost estimates, but excluded from the model for effect estimates. In our example, sex and – to a lesser extent – age had an impact on cost outcomes, but not on health effects. Our findings do not support the recommendation to keep covariates out of the propensity-score model that are unrelated to the outcome. They are, however, consistent with Stuart's advice to `err on the side of including more rather than fewer' [1].

Fifth, this study illustrates the importance of choosing a statistical method that fits with the research question. When an average treatment effect in the treated (ATT) is estimated, an ATT method should be used, while ATE methods performed best when estimating an average treatment effect for the full sample (ATE). This was most pronounced when estimating costs. It should be noted that when estimating ATT using ATE methods, the ATE estimates were applied to the treated population instead of the total population. This could explain the relatively small difference in ATT versus ATE as found in the effectiveness estimates.

Sixth, for estimating ATTs, kernel and 1-to-1 matching led to similar results. For ATEs, double robustness and IPTW had the best performance. Covariate adjustment achieved the lowest rankings. Regression achieved the highest rank in estimating effectiveness but performed much worse in estimating costs and cost-effectiveness.

Regression versus PS methods

To further explain differences in PS methods versus conventional regression we want to emphasize that apart from the observed differences in performance, these methods also differ on a conceptual level. Adjustment for confounding can be approached from two angles. Defined simply, confounding is the combination of two associations – the association of a variable with the outcome (making it a risk factor) and the association of this variable with treatment assignment [29]. The problem can be solved by addressing either of these associations.

Regression focuses on modelling the effect of the prognostic factors on the outcome. The treatment effect that is estimated by regression is the average effect of treatment over all observed values of covariates. Propensity-score methods eliminate the association of the confounder with treatment. This leads to some advantages of PS methods compared to conventional regression. First, it separates the design of the study from the analysis. The researcher can verify whether important covariates are balanced across treatment groups; in other words, whether adjustment has been successful. Furthermore, by comparing the distributions of the PS and matching treated and untreated patients, it is possible to check whether there is sufficient overlap between the treatment groups. This monitoring option is to a lesser extent also available in weighting and covariate adjustment by the PS, but not at all in conventional regression.

The second advantage of PS methods is in the reduction of the dimensionality problem. The possible number of covariates in regression is not limitless. It is restricted by the number of subjects or events. A ratio of at least 10 subjects or events per independent variable has been mentioned [30]. PS, which provide a scalar summary of the covariate information, do not have this limitation. After successful propensity-score adjustment, the only required covariate in the final analysis model would be the treatment variable.

A third advantage, which conflicts with the previous point, is that propensity-score methods can be used for 'pre-processing' the data before the final analysis is performed. Rubin et al. have argued that the consequences of a misspecified regression model are less severe when the covariate structure is more balanced [5,6]. This approach, which does not equal the double robustness method, does require that several covariates are included in the model.

Some aspects of using PS methods have not been discussed here. The practicalities have been discussed extensively in other papers [1,4].

New aspects of this study

Several aspects of this study are new. Propensity-score methods have mostly been assessed on clinical outcomes, much less on costs [31,32] and not on cost-effectiveness. Next to this, to our knowledge, our clinical measure, the difference in mean survival, has never been the outcome of interest. Martens et al. and Austin tested propensity-score methods in Cox proportional hazard models [33], but only assessed the impact of different models on the hazard ratio.

Third, our most important criterion for the assessment of the estimates was accuracy, which we defined as the proportion of samples with an effect estimate within an acceptable range around the true value. Such criterion has never been used before. To a large extent, accuracy is quite similar to the mean squared error, which is equal to the sum of the variance of the estimator and the square of the bias. Therefore, it represents a quantification of the variance-bias trade-off [8]. Accuracy does the same thing. Since high accuracy is achieved when bias and variance are comparatively small. However, a relatively higher bias may, in

this measure, be compensated for by a relatively small variance. In contrast to the mean-squared error, the acceptability depends on the arbitrary definition of the acceptable range. However, this barely influences the ordering of successful and less successful methods and accuracy has the important advantage that it can be interpreted intuitively: the probability of getting it right.

Limitations

This study has a number of limitations. A study based on simulation is not the same as a study on real data. However, without simulation the 'real' effect would not be known and bias would not have been quantified. The synthesised data was based on real data from a randomised controlled trial and an observational study of chemotherapy in patients with colorectal carcinoma. Next to this, an effort was made to prevent an artificially good fit of the propensity-score models by using the same model in synthesizing as well as analysing the data. Instead, we chose the – rarely used – complementary log–log model for treatment assignment in the synthesizing process, while probit models were applied to estimate the PS.

Another limitation is that we did not apply censoring in our synthesized data. Although many real life datasets do contain censored data, this choice helped to isolate the effects of the other model specifications.

We also assumed that there was no unmeasured confounding. When treatments are assigned to patients, something like the intuition of the treating physician may play a role. This cannot be explicitly expressed in a variable for which adjustment can take place. On the other hand, if balance is achieved on the major predictors of the outcome, this may not always be a problem. However, it cannot be ruled out that the physician somehow has more information than the data show. If this information is associated with measured covariates, adjustment approaches that use many adjustment variables may perform better than others.

This study compared only two matching methods, 1-to-1 matching with replacement and kernel matching. Although both are considered good methods [1], others also exist. Baser compared more matching methods for estimating cost differences and found the choice may have a substantial effect on the estimated treatment effect [31]. In that case, kernel matching performed well when it was combined with regression. Without the added regression, Mahalanobis matching had the best results

Lastly, the performance results of the methods and specifications examined in this study should be seen as a minimum estimate. When they are used on real data, it is possible to assess covariate balance applying a propensity-score method. If necessary, the propensity-score model can be adjusted to achieve better balance. That was not possible in our Monte Carlo simulations.

In this chapter we did not examine the reliability under the different methods. In the next chapter, we assess this uncertainty by expressing it in confidence intervals and cost-effectiveness confidence ellipses.

Recommendations

Based on the current study, certain suggestions can be made regarding cost-effectiveness analysis using observational time-to-event data. The researcher should explicitly choose to estimate an ATT or an ATE. If an ATT is required, matching methods are candidates. If the outcome of interest is an ATE, inverse probability weighting and double robustness are most likely to lead to acceptable results.

It is important that the same statistical method is applied to the estimation of both cost and effects. This prevents the estimates from being based on different samples of patients. The adjustment model should contain all baseline variables with an impact on costs or effects. After the application of a PS method, the 'pre-processed' data is best analysed in a fully specified regression model, including all baseline variables.

Since no method and specification has been shown to lead to the best results in all circumstances, it is recommended that several methods are applied and compared in sensitivity analyses. If they lead to similar results, this would strengthen confidence in the conclusions, since they are not sensitive to the choice of the statistical method. Presenting inconsistent results of different methods makes this structural uncertainty transparent.

8.5 References

1 Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2007; 15: 1-21.

2 Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006; 59: 437-447.

3 Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol. 2005; 58: 550-559.

4 Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011; 46: 399-424.

5 Rubin DB, Thomas N. Combining propensity score matching with additional covariate adjustments for prognostic covariates. J Am Stat Assoc. 2000; 573-585.

6 Rubin DB. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics. 1973; 29: 185-203.

7 Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects. Am J Epidemiol. 2011; 173: 761-767.

8 Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biom J. 2009; 51: 171-184.

9 Austin PC. A Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. Multivariate Behav Res. 2011; 46: 119-151.

10 Mitra N, Indurkhya A. A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data. Health Econ. 2005; 14: 805-815.

11 Indurkhya A, Mitra N, Schrag D. Using propensity scores to estimate the cost-effectiveness of medical therapies. Stat Med. 2006; 25: 1561-1576.
12 Sekhon JS, Grieve RD. A matching method for improving covariate balance in cost-effectiveness analyses. Health Econ. 2012; 21: 695-714.
13 Gruschkus SK, Lairson D, Dunn JK, et al. Cost-effectiveness of white blood cell growth factor use among a large nationwide cohort of elderly non-Hodgkin's lymphoma patients treated with chemotherapy. Value Health. 2011; 14: 253-262.
14 Mojtabai R, Zivin JG. Effectiveness and cost-effectiveness of four treatment modalities for substance disorders: a propensity score analysis. Health Serv Res. 2003; 38: 233-259.
15 Weiss JP, Saynina O, McDonald KM, et al. Effectiveness and cost-effectiveness of implantable cardioverter defibrillators in the treatment of ventricular arrhythmias among medicare beneficiaries. Am J Med. 2002; 112: 519-527.
16 Dhainaut JF, Payet S, Vallet B, et al. Cost-effectiveness of activated protein C in real-life clinical practice. Crit Care. 2007; 11: R99.
17 Coleman CI, McKay RG, Boden WE, et al. Effectiveness and cost-effectiveness of facilitated percutaneous coronary intervention compared with primary percutaneous coronary intervention in patients with ST-segment elevation myocardial infarction transferred from community hospitals. Clin Ther. 2006; 28: 1054-1062.
18 Haas L, Stargardt T, Schreyoegg J. Cost-effectiveness of open versus laparoscopic appendectomy: a multilevel approach with propensity score matching. Eur J Health Econ. 2011;
19 Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol. 2006; 163: 1149-1156.
20 Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med. 2007; 26: 734-753.
21 Koopman M, Antonini NF, Douma J, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. Lancet. 2007; 370: 135-142.
22 Holland PW. Statistics and causal inference. J Am Stat Assoc. 81: 945-960.
23 Neyman J. On the application of probability theory to agricultural experiments. Essay on principles. Section 9.". Stat Sci. 1923 [1990]; 5: 464-472.
24 Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974; 66: 688-701.
25 Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator. Rev Econ Stud. 1998; 65: 261-294.
26 Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction (2nd ed.). New York: Springer-Verlag, 2008.
27 Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc. 1987; 82: 387-394.
28 Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol. 2001; 2: 278.
29 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology (3 ed.). Philadelphia: Lippincott Williams & Wilkins, 2008.
30 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995; 48: 1503-1510.

31 Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. Value Health. 2006; 9: 377-385.
32 Basu A, Polsky D, Manning WG. Estimating Treatment Effects on Healthcare Costs Under Exogeneity: is there a ‚Magic Bullet'? Health Serv Outcomes Res Methodol. 2011; 11: 1-26.
33 Martens EP, de Boer A, Pestman WR, et al. Comparing treatment effects after adjustment with multivariable Cox proportional hazards regression and propensity score methods. Pharmacoepidemiol Drug Saf. 2008; 17: 1-8.

8.6 Appendix A

Synthesis of survival time

As a first step, a Weibull survival analysis was conducted in the empirical dataset to analyse the effect of treatment and other covariates on survival. Next, a source population of 20,000 patients was synthesized with covariates that preserved the covariance structure of the original data. The covariates were those with a statistically significant effect on survival in the Weibull survival model (age>70 years, performance score, elevation of levels of alkalic phosphatase (AF), white blood cell count (WBC) and lactate dehydrogenase (LDH), resection of the primary tumour, number of involved metastatic sites) plus two additional ones (gender and age).

These synthesized covariates and the Weibull regression results were then combined to construct individual survival functions per treatment option, in which time *t* is related to the probability of survival, *S(t)*. Random drawings from a uniform distribution (0-1) determined the point on *S(t)* (i.e., the y-axis of the survival curve) until which the patient would survive, which corresponded to a time *t* (the point on the x-axis).

For each individual, one drawing for *S(t)* was performed, which was used for both treatment arms. Patients who responded relatively well (compared to other patients in this treatment group) to combination treatment C, also responded relatively well to sequential treatment.

Instead of a fixed treatment effect, individual variation was introduced by performing random draws from Gaussian distributions of the confidence interval around the estimated treatment effect as well as the coefficients for the other covariates. The distribution for the treatment effect was broadened to achieve that the treatment effect would be negative in approximately 10% of patients. In order to retain the mean and covariance structure of the original data and the regression coefficients, the random draws were combined with the Cholesky decomposition of the variance-covariance matrix from the Weibull model.

As a consequence, survival times under both treatments depended on two elements of change, drawings for *S(t)* and for the treatment effect, which was reflected in the coefficients for other covariates as well.

Synthesizing costs

A generalized linear model was fit on the cost data in order to estimate a model with which data could be synthesized. The best fitting model contained a power link (power=1.29) and a Poisson variance function. The covariates included in the final model were survival days, square of survival days, plus the interactions of these two variables with treatment, age, and sex.

Using the coefficients from this model, predicted mean costs were estimated for each synthesized patient for each treatment arm, as well as a standard deviation. In order to introduce individual variation in individual costs, the predicted mean and standard deviation were used to describe gamma distributions per patient and per treatment arm, from which the individual patient's costs for each treatment were drawn.

8.7 Appendix B

Treatment assignment process

The process of drawing samples with random or non-random treatment assignment consisted of three steps.

This process consisted of three steps. First, a complementary log-log model equation was specified in order to project patient-specific probabilities of receiving the new treatment for all patients in the source population, based on their baseline characteristics. In the second step, these probabilities were applied in Bernoulli distributions, from which the treatment assignment was drawn for each patient. Thirdly, 1000 patients from each treatment were drawn from each treatment group into the sample.

The process was repeated for five combinations of parameters. In the random treatment assignment process, all coefficients in the complementary log-log model were set at 0, which lead to a 50% of receiving new treatment for all synthesized patients. In the first non-random treatment assignment process (see table 8.1), older and female patients were less likely to receive combination treatment. The second process made treatment more likely to be assigned to younger, female patients with a good performance score and normal values for AF and WBC. Under the third process, the probability of receiving combination treatment was positively related to having an abnormal LDH, an unresected tumour and more than one metastatic site. The fourth process contained the variables from the third process plus gender, age, elevated AF and elevated WBC.

9

A propensity to be precise

A Monte Carlo simulation study comparing statistical methods to obtain reliable cost-effectiveness estimates in observational studies

Objective Estimates of real-world cost-effectiveness are mostly based on observational data with non-random treatment assignments. Several methods exist to address the resulting confounding-by-indication, including regression and methods based on propensity scores (PS). This study examined the precision of the estimates from these methods in the context of cost-effectiveness analysis. The PS-methods were: PS matching (kernel and 1-to-1), covariate adjustment using PS, inverse probability-of-treatment weighting (IPTW) and double robustness, each with several specifications.

Methods A simulated source population was used to draw ten samples (n=400 or n=2000) in which patients were randomly or non-randomly assigned to a new or conventional treatment. Mean incremental costs and effects and the incremental cost-effectiveness ratios were calculated using each method. A bootstrapping procedure was used to express the uncertainty around the estimates. The precision of each method in each sample was quantified by 95% confidence intervals and by the area of the confidence ellipse on the cost-effectiveness plane.

Results In estimates of the average treatment effect in the treated (ATT) PS kernel matching led to more precise estimates than PS 1-1 matching. Regarding average treatment effects for the sample as a whole (ATE) conventional regression performed best, while PS covariate adjustment led to the least precise estimates.

Conclusions For the ATT, kernel matching is an attractive option. There appears to be no good reason to use PS covariate adjustment.

Lucas M.A. Goossens, Chantal W.M. van Gils, Maiwenn J. Al.,
Carin A. Uyl-de Groot, Miriam Koopman, W. Ken Redekop

9.1 Introduction

The precision of estimates is an important issue in cost-effectiveness analysis. Especially when the point estimate of the incremental cost-effectiveness ratio (ICER) is close to a decision maker's willingness to pay, a large amount of uncertainty increases the risk of a wrong decision [1,2].

When the estimated ICER is considered acceptable, a new technology could be adopted when its true (unobserved) cost-effectiveness is insufficient. Alternatively, the new intervention could be rejected while it would actually be cost-effective.

The precision of an estimate depends on the sample size and the variation across subjects, but is also affected by the statistical methods that are used in the analysis. The inclusion of prognostic factors in the analysis of randomised controlled trials, when these factors are asymptotically balances across treatment groups [3,4]. This has been shown to enhance the statistical power and precision of the estimates [5]. According to Senn, ignoring balanced prognostic factors in the analysis is 'a grave mistake' [6].

In observational studies, making this mistake is not even an option. Prognostic factors have to be included in the analysis to adjust for possible systematic differences between treatment groups, which would lead to confounding bias. In addition to conventional regression methods, a new class of methods is gaining popularity in this context. These are based on estimated propensity scores (PS), which are defined as a patient's probability of receiving a specific treatment assignment, given certain observed characteristics. Possible applications of the PS include stratification on the PS, matching on the PS, the use of PS to replace original baseline covariates in a regression model (PS covariate adjustment), weighting on the PS (IPTW) and double robustness (DR). These can be applied with different model specifications.

The ability of these methods to successfully address bias has been investigated extensively in recent years [7-10]. In the previous chapter, we evaluated the performance of several PS methods with varying specifications and conventional regression in the context of cost-effectiveness analysis.

However, those studies did not cover the precision of the estimates. Although it has been shown that conventional regression leads to more precise estimates propensity scores methods in general [11], no comparison between different PS methods is available. A criterion in many comparative studies on PS methods was the mean squared error (MSE), which was presented as combined measure of bias and precision [9,12]. It was calculated as the mean squared difference between the real treatment effect and the model estimate over a large number of simulations. Although MSE is a very useful measure – it could be interpreted as the expected squared 'incorrectness' of the estimates – it only captures precision within the sample. In contrast, the term precision is generally used to describe the relationship between the estimate on a sample and the treatment effect in the population from which the sample was taken [13]: if the estimate is more precise, there is less uncertainty about the

value in the population. This interpretation of the term is used in confidence intervals, and in cost-effectiveness analysis, cost-effectiveness planes and acceptability curves.

This current chapter builds on the findings in the previous chapter on propensity score methods in cost-effectiveness studies. Its objective was to examine the precision of cost-effectiveness estimates from different PS-based statistical methods and conventional regression in non-randomised data.

9.2 Methods

The methods can be divided in following steps: 1) synthesizing source population, 2) drawing ten samples, 3) analyzing samples with PS-based methods and conventional regression, 4) assess precision of estimates.

Synthesis of source population

We performed a simulation study on a synthesized source population of 20,000 patients. The synthesizing process has been described in detail in the previous chapter. In short, the variation in characteristics found in the source population reflected the distributions and covariance structures of two Dutch empirical studies comparing a new treatment in metastatic colorectal cancer to the conventional one [14]. The synthesized dataset contained values for baseline characteristics and potential outcomes in terms of health (survival time) and healthcare expenditures for both treatment options.

This source population was used to draw samples in which patients were assigned to the new or the conventional treatment in a random or non-random fashion. Five different treatment assignment processes were used (0, 1, 2, 3, 4) in sample sizes of n = 400 and n = 2000, which resulted in a total of ten samples. In process 0 the new treatment was randomly assigned. In contrast, the treatment assignment in processes 1-4 was associated with certain baseline characteristics. The first treatment assignment process assigned the new treatment disproportionally to patients with higher projected cost. The second process assigned the new treatment disproportionately to patients with a relatively good prognosis and low projected costs. In the third process, the new treatment was assigned mostly to patients with an unfavourable prognosis. The fourth process was based on combination of favourable and unfavourable risk factors.

Drawing ten samples

The sample drawing process described above was repeated ten times, resulting in a total of 100 preliminary samples. Ten final samples, one for each treatment assignment process and sample size were selected from the preliminary samples. This selection was based on the accuracy of the point estimates for the ICER from all methods in this particular case: the

sample for which the estimates were the most accurate was selected. The previous chapter has shown that all methods had a risk of misestimating the treatment effect in certain samples. The selection was made in order to isolate the issue of precision from validity and to make the precision estimates from different methods more comparable. In nearly all cases, the point estimates of the ICER for all methods in the selected samples were within 5000 euros of the true value.

Analysis of samples with PS-based methods and conventional regression

Six methods were compared.

(1) **Conventional regression**. A fully specified Weibull regression model was used to estimate the survival gains. The covariates were the prognostic factors that were used for synthesizing survival time. Costs were analysed in a generalized linear model with a log link and Gaussian variance function, with days of survival time (linear and squared) and the interaction terms of survival time (linear and squared) with treatment, age and sex.

(2) **PSM 1-to-1**: PS matching (with replacement and common support requirement). Propensity scores were calculated by fitting a probit model. The closest matching untreated patient was selected for each treated patient, based on their PS. Untreated patients could be matched to more than one treated patient. Matching only took place for treated patients with common support: those with a PS in the range of the scores of the untreated patients. After matching, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(3) **PSM kernel**: PS matching with kernel smoothing. Treated patients with common support were matched to all untreated patients, but the latter were weighted according to the distance between their PS and the treated patients' in such a way that the combined weights equalled 1 [15]. The Epanechnikov kernel was used as a weighting function. After the matching procedure, analysis was similar to the previous method [16].

(4) **PS covariate adjustment** regression analysis. Propensity scores were used as covariates in the Weibull model and GLM for survival and costs.

(5) **IPTW**: inverse probability of treatment weighting. [17,18] Treated patients were weighted by the inverse of their PS (the probability of being treated), controls were weighted by the inverse of 1 minus the PS (or the propensity of being untreated). After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(6) **Double robustness**. This technique combines regression and weighting by the PS in one equation [19]. Results should be unbiased if either the regression model or the propensity-score model is correct.

Model specification

Propensity scores were based on a model that contained all prognostic factors for survival and the covariates that predicted costs. The Weibull regression models under methods 2

through 5 were either fully specified (full endpoint model) or contained only a variable for treatment (simple endpoint model, which for PS covariate adjustment also contained the PS). The GLMs contained all covariates from the synthesizing model (full model) or only survival time (linear and squared), treatment and their interactions (simple model).

Estimation of incremental costs and effects

For each patient, the expected survival time was projected for each treatment option, based on their baseline covariates and the Weibull regression results. The projected costs were the fitted values of the GLM for costs, based on the projected survival. The ATE was the mean of treatment effects (the differences between the two individual potential treatment outcomes) in the sample. The ATT was the mean individual treatment effect across patients who were assigned to the new treatment.

Assessment of precision

The assessment of the precision of the estimates was based on a bootstrapping procedure with 5000 replications. In a bootstrapping procedure, a large number of random samples with replacement is taken from the original sample, with a size equal to the original size. [20,21]. For each bootstrap sample, incremental costs and effects were calculated. The pairs of incremental costs and effects for all samples can be presented graphically in a cost-effectiveness plane. If the results have a bivariate normal distribution, the scatter plot is shaped as an ellipse.

The precision of each combination of method and sample size was quantified in two ways. First, 95% confidence intervals for incremental costs and effects were determined using the percentile method: the 2.5% highest and the 2.5% lowest ICER estimated were excluded from the intervals. However, these intervals can only be interpreted when all bootstrap effect estimates are positive (or all are negative); in other words, if the scatter plot does not straddle the y-axis. Otherwise, a negative ICER could be very favourable (when health benefits are combined with cost savings) or unfavourable (health losses and cost increases). Alternatively, a positive ICER could be the result from a treatment benefit and cost increase (in which case a lower ICER is preferred) or from health loss and costs savings (in which case a higher ICER is preferred).

The width of the confidence intervals was calculated for incremental costs and effects and – if the interval was interpretable – for cost-effectiveness. Larger confidence intervals represented less precision.

The second measure of precision was the area of the 95% confidence ellipses, which could only be calculated when the scatter plot was ellipse-shaped. The area of an ellipse is given by this equation:

$$Area = \pi ab \quad (1)$$

where a= the semi-major axis, i.e. half of the longest diameter of the ellipse and b = the semi-major axis, i.e. half of the shortest diameter. The major axis is orthogonal to the minor axis.

The ellipse parameters are given by

$$a^2 = \left(\frac{\sigma_x^2 + \sigma_y^2}{2} + \sqrt{\frac{\left(\sigma_x^2 - \sigma_y^2\right)^2}{4} + \sigma_{xy}^2} \right) * \gamma$$

$$b^2 = \left(\frac{\sigma_x^2 + \sigma_y^2}{2} - \sqrt{\frac{\left(\sigma_x^2 - \sigma_y^2\right)^2}{4} + \sigma_{xy}^2} \right) * \gamma$$

where σ_x^2 and σ_y^2 are the variances of the bootstrap estimates of incremental costs and effects and σ_{xy} as the covariance of incremental costs and effects. The factor γ adjusts the ellipse to the desired level of confidence. It is based on the χ^2-distribution with 2 degrees of freedom. For a 95% confidence ellipse:

$$\gamma = \chi^2(2, 0.05) \approx 5.99$$

9.3 Results

The widths of the confidence intervals and the areas of the confidence ellipses are summarized in table 9.1. They could be calculated for all samples with n=2000. Twelve out of the 50 intervals, most of them for sample 4, were not interpretable because they included both negative and positive estimates of the incremental effects. One area could not be calculated because the scatter plot was not ellipse-shaped.

The widths and areas for the samples with 400 patients were larger than those for the corresponding method and samples of 2000 patients. This means that larger samples produced more precise estimates, which was to be expected. It is illustrated by the confidence ellipses for IPTW and PSM kernel, both with full endpoint models from sample 2, in figure 9.1.

The centres of the ellipses, which represent the mean of the estimates for incremental costs and effects, did not overlap perfectly. This is due to the fact that different methods lead to different estimates. Furthermore, although the samples were taken from the same source population, they were drawn separately. Samples with n = 400 were not nested in samples with n = 2000. Nevertheless, the differences in shape and size of the ellipses are apparent.

Table 9.1 ICER confidence interval widths and ellipse areas

		Interval width[a]					Ellipse area[b]				
		Treatment assignment process					Treatment assignment process				
		0	1	2	3	4	0	1	2	3	4
N = 400											
ATT methods											
PSM 1-to-1	Simple	121	73	-[c]	180	-[c]	37	55	68	55	71
PSM 1-to-1	Full	106	33	90	48	-[c]	25	40	49	34	34
PSM kernel	Simple	53	48	74	78	-[c]	21	36	47	32	41
PSM kernel	Full	79	27	90	38	-[c]	15	27	32	23	21
ATE methods											
Regression	Full	85	25	36	41	108	15	19	19	26	15
Covariate adjustment	Simple	118	69	-[c]	134	-[c]	34	46	47	64	47
Covariate adjustment	Full	107	30	85	65	-[c]	26	37	43	-[d]	30
IPTW	Simple	47	76	53	95	-[c]	21	33	39	38	34
IPTW	Full	69	68	61	59	-[c]	15	29	23	27	21
Double robustness	Full	49	-[c]	38	69	-[c]	13	41	31	20	27
N = 2000											
ATT methods											
PSM 1-to-1	Simple	43	61	24	42	35	8	12	12	9	14
PSM 1-to-1	Full	27	28	16	31	19	5	8	8	8	7
PSM kernel	Simple	22	32	17	25	23	4	7	8	4	8
PSM kernel	Full	18	22	13	22	15	3	5	5	4	5
ATE methods											
Regression	Full	19	18	15	19	17	3	5	5	4	4
Covariate adjustment	Simple	41	52	23	44	34	6	10	9	10	13
Covariate adjustment	Full	25	29	16	30	19	5	8	7	7	7
IPTW	Simple	22	32	22	24	24	4	9	7	5	7
IPTW	Full	19	30	16	19	17	3	8	5	4	5
Double robustness	Full	19	23	20	18	19	3	7	4	3	5

[a] Widths were divided by 10^3 for clarity of presentation. [b] Areas were divided by 10^5 for clarity of presentation. [c] Interval could not be interpreted. [d] Scatter plot was not ellipse shaped.

Abbreviations: ICER, incremental cost-effectiveness ratio; ATT, average treatment effect in the treated patients; ATE, average treatment effect; IPTW, inverse probability-of-treatment weighting; PSM, propensity score matching.

Simple versus full endpoint models

In all cases, the confidence ellipse areas for methods with fully specified endpoint models were smaller than those for methods with simple endpoint models. The results for the widths of most all confidence intervals were consistent with this observation. An illustration is presented in figure 9.2 for IPTW and PSM kernel in sample 2 with n=400. The ellipses for the fully specified endpoint models are inside the ellipses of for the simple models.

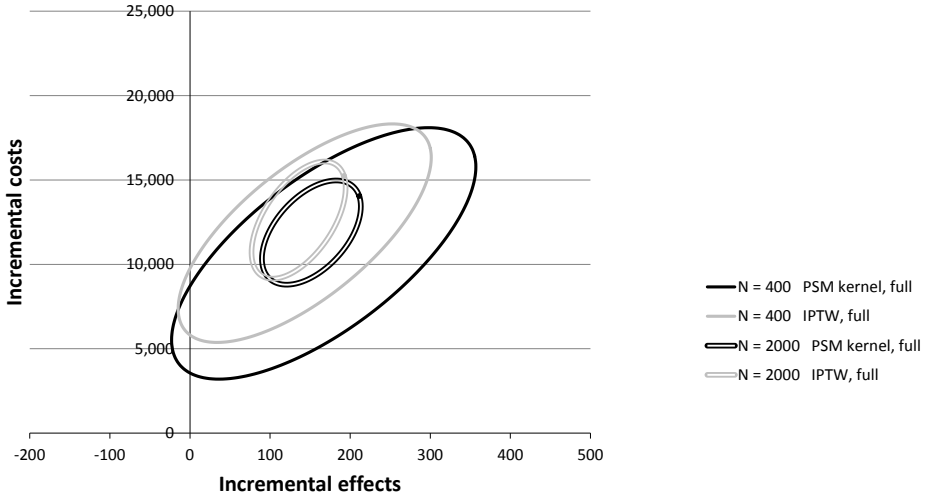

Figure 9.1 95% Confidence ellipses on the CE-plane: 400 versus 2000 (sample2).

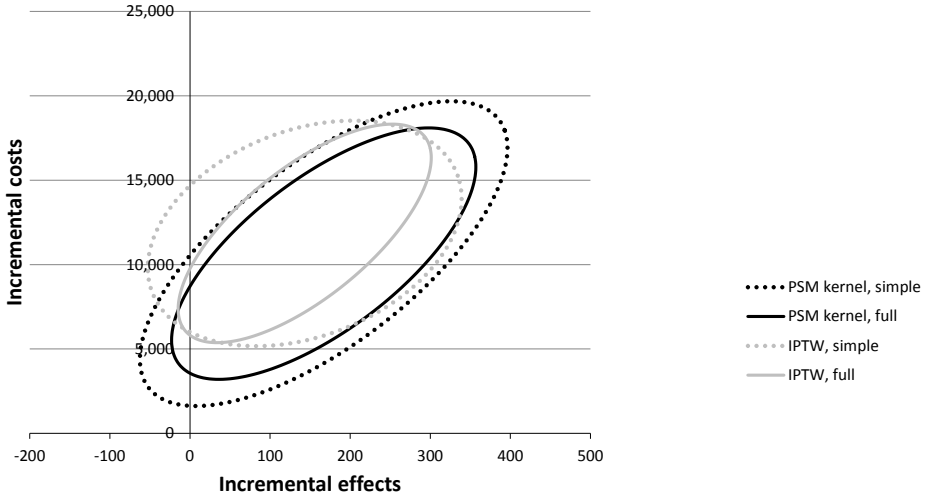

Figure 9.2 95% Confidence ellipses on the CE-plane: full versus simple model (sample 2, n = 400).

ATT methods compared

The PSM kernel method consistently led to smaller ellipse areas and interval widths than the PSM 1-to-1 method. Figure 9.3 illustrates this by showing the ellipses for both methods with fully specified endpoint models in sample 2 with n=400.

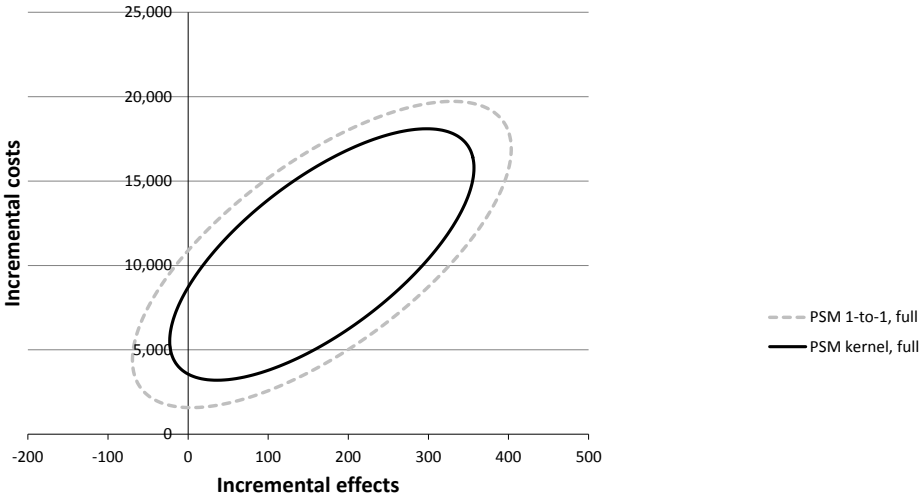

Figure 9.3 95% Confidence ellipses on the CE-plane: ATT methods compared (sample 2, n = 400).

ATE methods compared

The differences between the ATE methods were less clear than those for the ATT methods. Figure 9.4 shows the ellipses for conventional regression, double robustness covariate adjustment (with full model), IPTW (with full endpoint model) for sample 2 with n=400.

When only the methods with fully specified endpoint models were compared, the intervals for covariate adjustment were consistently wider than for the other models in the samples with n=400. The intervals for conventional regression were relatively small. These

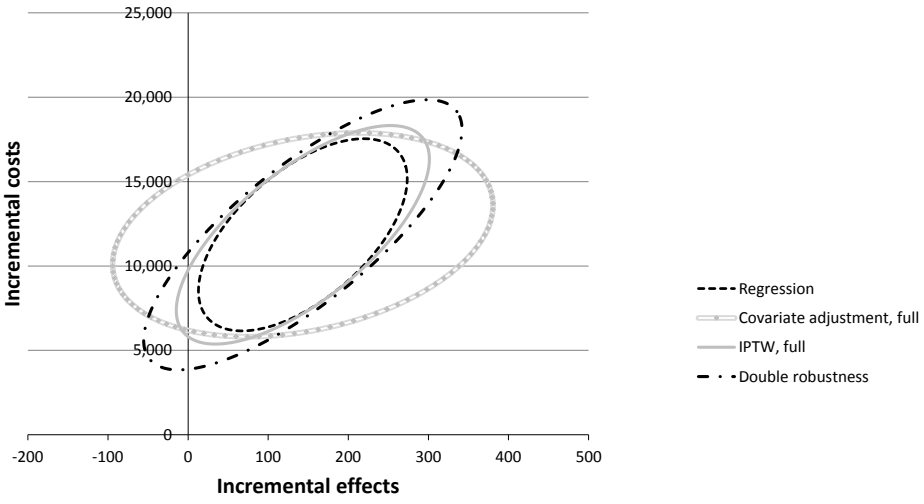

Figure 9.4 95% Confidence ellipses on the CE-plane: ATE methods compared (sample 2, n = 400).

trends were less obvious in the larger samples. With regards to the ellipse areas, however, regression had the best performance in most cases, while covariate adjustment performed worst.

9.4 Discussion

This study examined the precision of the cost-effectiveness estimates from several propensity-score based adjustment methods and conventional regression.

The relevance of this study lies in the increasing interest in observational studies to assess the impact of medical treatments on health, cost and cost-effectiveness outcomes. While randomised controlled trials (RCTs) and economic evaluations piggy-backed onto RCTs have long been viewed as the gold standard for estimating these outcomes, interest nowadays also focuses on estimating cost-effectiveness in 'real-world' settings, outside the tightly controlled confines of an RCT. In observational studies, propensity-score-based methods can be useful tools for addressing confounding bias, especially when an ATT is required.

The following are our main findings. Firstly, in very large samples (n=2000) the differences in precision were small and did not consistently favour one method over another. In contrast, when the sample size was smaller (n=400), the difference between methods and specifications were substantial.

Secondly, in estimates of the average treatment effect in the treated (ATT), kernel matching led to more precision than 1-to-1 matching without replacement. This could be explained by the fact that, under kernel matching, each treated subject is always matched to a great number of controls. This reduces the impact of the presence of each individual control patient in the matched sample.

Thirdly, regarding average treatment effects for the sample as a whole (ATE), covariate adjustment with the PS had the worst precision, and conventional regression led to the most precise estimates. This was most pronounced in the smaller samples. The other methods, double robustness and IPTW had a moderate precision.

Fourthly, including covariates in the endpoint model resulted in more precision than simple models. This confirms earlier findings that for reasons of efficiency, it is preferable to model the health outcome instead of treatment assignment [11].

Some authors have advocated the pre-processing approach in order to enhance the likelihood of achieving an estimate that is close to the true treatment effect [22]. This means that in the pre-processing stage, a propensity-score matching or weighting method is applied in order to achieve balance in the covariates across treatment groups. Next, a full endpoint regression is applied to the pre-processed data. With an eye to obtaining an unbiased point estimate, this approach is attractive because it offers two opportunities to solve the problem

of confounding. If covariate balance is not fully achieved in the first step of the process, this can be repaired by the endpoint model. Or, from the opposite perspective, if the endpoint model is not completely correct, this is less of a problem when the covariates have been balanced to a reasonable extent.

Our study adds an argument in favour of the pre-processing approach, at least compared to simple endpoint models after a PS-based method. Adjusting for prognostic covariates in the endpoint model does not only improve validity, but also the precision of the estimates. It must be noted, however, that applying an outcome regression model without weighting or matching on the propensity score still led to more precise estimates than pre-processing.

We used two measures to evaluate the precision of estimates: the width of the 95% confidence intervals for the ICER and the area of the 95%-confidence ellipse on the CE plane.

When the ICER is the sole interest of the study, the width of the confidence interval appears the most obvious choice. A smaller interval means that the estimate is relatively precise. However, this does not necessarily mean that incremental costs and effects have been estimated precisely. If costs and effects are strongly and positively correlated, a much larger amount of uncertainty surrounding these outcomes may coincide with precision on the ICER. Furthermore, the interval is only interpretable when the bootstrap results do not straddle the x-axis of the CE plane.

The area of 95% confidence, in contrast, can always be interpreted. However, the area can only be calculated when the bootstrap replications on the plane are shaped in an ellipse. The area reflects the combined uncertainty on incremental costs and effects. This does not automatically comprise uncertainty on the ICER. When costs and effects are strongly and negatively correlated, the area is small, while the uncertainty on the ICER may be large.

However, the discrepancies between the two evaluation measures of precision may only occur in extreme circumstances. That is why, in our study, they led to the same conclusions, although costs and effects were correlated positively and substantially. Both are closely linked to the tools that are typically used to assess the uncertainty around the ICER estimate in economic evaluations: CE planes and acceptability curves [23]. Nevertheless, the ellipse and the interval are conceptually different.

This study has a number of limitations. First, a study based on simulation is not the same as a study on real data. However, without simulation the 'real' effect would not be known and we would not have been able to select samples that gave a relatively unbiased ICER results. In addition, eliminating bias gave us the opportunity to examine precision separately from validity. We paid great attention to ensure that the synthesised dataset reflected real-life uncertainty. The data were based on real data from a randomised controlled trial and an observational study of chemotherapy in patients with colorectal carcinoma. Next to this, an effort was made to avoid overestimating the performance of the propensity-score models that would have arisen if we had used the same model in synthesizing as well as analysing the data. Instead, we chose the – rarely used – complementary log–log model

for treatment assignment in the synthesizing process, while probit models were applied to estimate the PS.

Another limitation is that we did not apply censoring in our synthesized data. Although many real life datasets do contain censored data, this choice helped to isolate the effects of the other model specifications.

We also assumed that there was no unmeasured confounding. When treatments are assigned to patients, something like the intuition of the treating physician may play a role. This cannot be explicitly expressed in a variable for which adjustment can take place and could cause bias that no regression or PS adjustment method can adjust for, nor is this taken into account when estimating the uncertainty surrounding the ICER results. It is expected that not taking unmeasured confounding into account led to an underestimation of this uncertainty, but this is not expected to alter our findings would alter since none of the adjustment methods can take the unobserved confounding into account.

McCandless et al. proposed to perform additional sensitivity analysis to more adequately reflect the uncertainty by taking potential unmeasured confounding into account [24]. It is advised to perform such additional uncertainty analysis when using observational data.

The conclusions of this chapter can be combined and compared with those of the earlier study, which focussed on the validity of the estimates of incremental cost-effectiveness from several adjustment methods, in order to convey several recommendations

In previous chapter, we found no difference between the performance of 1-to-1 matching with replacement and kernel matching with regards to the validity of ATT estimates. However, kernel matching led to more precision in the current study, which makes it a more attractive option.

For ATE estimates, both studies concluded that PS-as-covariate adjustment was likely to lead to suboptimal results. The method had a relatively high risk of incorrect estimates, while the uncertainty was large. There appears to be no good reason to apply this approach, although it is the most frequently used PS method [25].

Conventional regression performed moderately with regards to validity, but left relatively little uncertainty. A disadvantage of this method is that it requires a correct specification of the regression model. Double robustness and inverse probability-of-treatment weighting (IPTW) led to somewhat more valid estimates, but at the cost loss of precision. For PS-based ATT as well as ATT methods, the pre-processing approach could considerably reduce uncertainty while improving accuracy.

Nevertheless, in for ATEs a trade-off between precision and accuracy remains. The optimal strategy to solve this problem could be comparing the results from conventional regression and pre-processing (IPTW followed by fully specified regression). If both methods lead to similar results, the superior precision of conventional regression would prevail. If the results are different, the pre-processing results are likely to be the most accurate.

9.5 References

1 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999; 18: 341-364.

2 Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: the consequences of ignoring decision uncertainty. Health Econ. 2011; 20: 212-224.

3 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med. 2002; 21: 2917-2930.

4 Steyerberg EW. Clinical predicition models. A practical approach to development, validation, and updating . New York: Springer, 2009.

5 Hernandez AV, Eijkemans MJ, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? Ann Epidemiol. 2006; 16: 41-48.

6 Senn S. Controversies concerning randomization and additivity in clinical trials. Stat Med. 2004; 23: 3729-3753.

7 Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011; 46: 399-424.

8 Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biom J. 2009; 51: 171-184.

9 Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol. 2006; 163: 1149-1156.

10 Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2007; 15: 1-21.

11 Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics. 1992; 48: 479-495.

12 Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. Stat Med. 2010; 29: 2137-2148.

13 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology (3 ed.). Philadelphia: Lippincott Williams & Wilkins, 2008.

14 Koopman M, Antonini NF, Douma J, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. Lancet. 2007; 370: 135-142.

15 Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator. Rev Econ Stud. 1998; 65: 261-294.

16 Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction (2nd ed.). New York: Springer-Verlag, 2008.

17 Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc. 1987; 82: 387-394.

18 Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol. 2001; 2: 278.

19 Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects. Am J Epidemiol. 2011; 173: 761-767.

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

21 DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statist Sci. 1996; 189-228.

22 Rubin DB, Thomas N. Combining propensity score matching with additional covariate adjustments for prognostic covariates. J Am Stat Assoc. 2000; 573-585.

23 Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ. 1998; 7: 723-740.

24 McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. J Clin Epidemiol. 2008; 61: 247-255.

25 Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol. 2005; 58: 550-559.

Adjusting for COPD severity in database research

10

Developing and validating an algorithm

Background When comparing chronic obstructive lung disease (COPD) interventions in database research, it is important to adjust for severity. Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines grade severity according to lung function. Most databases lack data on lung function. Previous database research has approximated COPD severity using demographics and healthcare utilization. This study aims to derive an algorithm for COPD severity using baseline data from a large respiratory trial (UPLIFT).
Methods Partial proportional odds logit models were developed for probabilities of being in GOLD stages II, III and IV. Concordance between predicted and observed stage was assessed using kappa-statistics. Models were estimated in a random selection of 2/3 of patients and validated in the remainder. The analysis was repeated in a subsample with a balanced distribution across severity stages. Univariate associations of COPD severity with the covariates were tested as well.
Results More severe COPD was associated with being male and younger, having quit smoking, lower BMI, osteoporosis, hospitalizations, using certain medications, and oxygen. After adjusting for these variables, co-morbidities, previous healthcare resource use (e.g., emergency room, hospitalizations) and inhaled corticosteroids, xanthines, or mucolytics were no longer independently associated with COPD severity, although they were in univariate tests. The concordance was poor (kappa = 0.151) and only slightly better in the balanced sample (kappa =0.215).
Conclusion COPD severity cannot be reliably predicted from demographics and healthcare use. This limitation should be considered when interpreting findings from database studies, and additional research should explore other methods to account for COPD severity.

Lucas M.A. Goossens, Christine L. Baker, Brigitta U. Monz, Kelly H. Zou
Maureen P.M.H. Rutten-van Mölken

This study was published in the International Journal of COPD, 2011; 6:669-679.

10.1 Introduction

Treatment effectiveness and healthcare resource use in daily practice may be different from effectiveness and resource use in clinical trials. This is increasingly recognized by researchers, policy makers and decision makers responsible for pricing and reimbursement of healthcare interventions. Hence, the importance of data generated from sources that reflect the use and the associated outcomes in routine practice settings is growing. Suitable sources of such real-life data could be health records kept by physicians (e.g., their routine records or specifically established databases; paper-based or electronic), patient registries that enrol patients with specific diseases or other characteristics of interest (e.g., cancer registries), or administrative claims databases of healthcare insurers and provider organizations set up for the purpose of reimbursement of providers for their expenses.

The inherent problem of analyses conducted on such databases relates to the fact that the data are often collected for other purposes. When treatment is not assigned to patients at random, disease severity and prognosis of patients may differ systematically across treatments. In order to adjust for this, data that reflect disease severity must be available. The currently most widely used severity classification is the spirometric classification of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [1], which is based on lung function in terms of the forced expiratory volume in one second as a percentage of the value predicted for sex, age and height (FEV_1-%predicted). It is used to diagnose COPD, to monitor disease progression and to aid in treatment decisions. Lung function is an important indicator of COPD severity because it is associated with mortality [2], exacerbations [3,4], health care utilization and costs [5]. There is also an association between lung function and quality of life6 although this relationship is not as strong [6,7]. Consequently, the same associations have been shown for the GOLD classification as it is based on FEV_1 [8-12].

It is widely recognized, also by GOLD [1], that the impact of COPD does not only depend on lung function and furthermore that is not only a lung disease but a multi-dimensional disease with many systemic, extra-pulmonary effects. Therefore, composite measures of COPD severity have been proposed. However, almost all of these still accept the importance of airflow limitation and they are partly based on the FEV_1-%predicted [13-17].

However, lung function measures obtained through spirometry are not routinely available in databases and registries. Most routine databases also lack information on the other parameters that constitute the more recent composite measures of COPD severity. Previous retrospective database research has approximated COPD severity using demographic (e.g., age and/or smoking status) and resource utilization data (eg, medications used and/or hospital admission) [18-23].

In this study, we attempted to develop a multivariable predictive algorithm to derive the severity of COPD as classified by GOLD, using variables that are commonly available in routine databases. The analysis was performed on the baseline data of the “Understanding

Potential Long-term Impacts on Function with Tiotropium" (UPLIFT) trial, a large, 4-year trial in COPD patients with rate of decline in FEV_1 as primary endpoint [24]. This trial was especially suitable for this purpose because of its size (with almost 6000 patients randomised), and because among the baseline data collected are many data commonly available in routine databases (e.g., demographics, medications used in the past, hospital admissions in the year preceding enrolment). Moreover, FEV_1 was collected with high quality, thus providing confidence in the GOLD severity assignment of trial patients.

10.2 Methods

Data source

We used baseline data from the UPLIFT trial. The UPLIFT trial was a randomised, double-blind, placebo-controlled 4-year trial, investigating the effect of tiotropium on the yearly rate of decline in FEV_1 in patients with moderate to very severe COPD according to GOLD (stages II to IV). Patients were permitted to use all respiratory medications except inhaled anticholinergics. 5993 patients were randomised. 25 Main inclusion criteria besides a diagnosis of COPD included age of 40 years or over, a smoking history of at least 10 pack-years, a post-bronchodilator FEV_1-%predicted of 70% or less, and an FEV_1 of 70% or less of the forced vital capacity. Key exclusion criteria were a history of asthma, a COPD exacerbation or respiratory infection within 4 weeks before screening, use of supplemental oxygen for more than 12 hours per day. UPLIFT collected information on patients' demographics, co-morbidities and co-medications as well as COPD characteristics and smoking history, along with contacts with health care providers in various settings in the year preceding enrolment.

In order to enhance the homogeneity of treatment patterns, the sample was limited to 3698 patients from Western Europe, the United States, Australia and New Zealand with COPD severity stage II, III or IV. The sample was randomly split into two sets: one for developing the algorithm (2/3 of patients) and the other for validation (1/3 of patients).

Selection of variables

Potential variables for inclusion in the algorithm to approximate disease severity were those used in previous database research and those used in diagnostic and severity classifications [13-16,25-28]. Variables were then selected if available in UPLIFT and likely to be available in routine databases. The final list of variables to be tested consisted of age (continuously in years, squared, and in categories: 40–49, 50–59, 60–69, 70–79, 80–90, 90+). BMI (continuously and categorized: <18.5, 18.5–21.0, 21.0–25.0, 25.0–30.0, >30.0), smoking status (current/former smoker), pack-years of smoking (continuously and categorized: <20, 20–60, >60), sex, type of respiratory maintenance medications (short-acting and long-acting bronchodilators, inhaled corticosteroids (ICS), mucolytics, leukotriene modifiers, xan-

thine), number of medication types (0 to 5), treatment of exacerbations (number of courses of antibiotics or oral steroids, categorized: 0, 1, 2, ≥3 in one year), resource use in one year (number of emergency room visits with/without hospital admission, categorized (0, 1, ≥2), number of scheduled and unscheduled general practitioner (GP) visits (categorized: 0, 1, 2, 3–4, ≥5 and 0, 1, 2, ≥3, respectively), hospital admissions (yes/no), use of oxygen at home (yes/ no), statin use (yes/no), use of other cardiovascular medication (yes/no), number of co-morbidities (categorized: 0, 1, 2, 3–5, 6–9, ≥10), Charlson comorbidity index [29] (categorized: 1, 2, 3, ≥4), presence of selected co-morbidities (arrhythmia, coronary heart disease, vascular disease, hypertension, disorders of nervous system, stroke, diabetes, depression, anaemia, and platelet disorders, osteoporosis and cataract), as well as interaction terms (age*gender, age*BMI, BMI*gender).

Analysis

Several partial proportional odds (PPO) ordered logit models were used to estimate the probabilities of being in GOLD stages II (FEV_1%-predicted >50%), III (FEV_1%- predicted 30%–50%), and IV (FEV_1%- predicted <30%), as COPD severity categories are ordered. The "development dataset" was used for this purpose.

The PPO resembles the standard ordered logit model, which is the best known ordered regression model. In our case, with 3 possible outcomes Y, it estimates the probability that Y>stage II and the probability that Y>stage III. Ordered regression models assume that the observed ordered outcome Y is a function of a continuous and unobservable variable Y*. Two thresholds τ_j are assumed to determine in which stage a patient is classified: $Y > Y_j$ if $Y^* > \tau_j$, where j denotes a specific outcome and t_j is the upper limit of Y^* for this outcome. Y^* is related to the explanatory variables X: $Y_i^* = X_i\beta_j + \varepsilon_j$, where i denotes individual patients. Since the latent variable Y^* does not completely equal the sum of the products of coefficients and variable values, the outcome Y cannot be determined with certainty from the data. Assuming a logit distribution for the random error term makes it possible to model the probability that a patient is in a certain stage or worse ($Y^* > \tau_j$). This probability may be written as:

$$\Pr(Y_i > j) = \frac{\exp(-\tau_J + X_i\beta_J)}{1 + \exp(-\tau_J + X_i\beta_J)}, j = 1, 2, \ldots, M - 1.$$

M is the number of possible outcomes.

In contrast to the PPO model, however, the ordered logit model can only estimate one coefficient for each explanatory variable. This coefficient is assumed to be identical for all dichotomizations of the outcome variable (in this case, stage IV versus II/III and stage III/IV versus II), and thus coefficient $\beta_j = \beta$. This is the proportional odds or parallel regressions assumption. If this assumption is violated, which often happens in practice, estimates are invalid and important differences in the relationships at different thresholds may go un-

noticed. The partial proportional odds model relaxes this assumption for variables where it does not hold [30]. For these variables, a coefficient is estimated for each dichotomization.

The final model was developed by stepwise backward elimination from the full model, which contained all variables. In each step the variable with the largest *P*-value was eliminated and the model was re-estimated. This process was repeated until all variables had at least a value of $P \leq 0.10$.

We performed two sensitivity analyses in order to account for the fact that the proportion of patients with very severe COPD (GOLD IV) in the dataset was smaller than the proportion of patients with moderate or severe COPD. Firstly, the final model was re-estimated in a subsample with a balanced distribution of patients across all three severity stages. This balance was achieved by using all GOLD stage IV patients and random draws from patients in stages II and III. Secondly, the patients in GOLD stage III and IV were combined into one group. The probability of being in 'severe/very severe' was then analysed in a binary logit model with the same variables as in the final PPO model.

All analyses were performed in Stata 11.1, (StataCorp LP, College Station, TX) using the gologit2 command for the PPO model [31]. Univariate tests were analysis of variance (ANOVA) for continuous variables and chi-square for categorical variables. Statistical significance was reached when a two-sided *P*-value was ≤ 0.05.

Predictive performance

The regression results were used to predict the probabilities of being in each GOLD severity stage for each patient in the validation dataset. The predicted stage was defined as the stage with the highest predicted probability. The agreement between the predicted and the observed stage was assessed using kappa statistics [32]. For a kappa statistic, a value of 0 indicates that agreement has occurred by chance, whereas a value of 1 indicates perfect agreement. No generally accepted interpretation of the magnitude of the kappa-statistic exists. Landis and Koch suggested the following labels of agreement for ranges of values, which are often quoted: slight (up to 0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0) [33]. We present kappa statistics unweighted and weighted. In the weighted kappa statistics, patients misclassified in the neighbouring category (i.e., II or IV instead of III, or III instead of II or IV) count as 0.5 agreement and patients misclassified in a non- neighbouring category (II instead of IV or vice versa) as no agreement).

In addition to the agreement per disease stage, the overall agreement was calculated. For the binary logit model, the c-statistic was computed as a measure of its discriminative performance. The c-statistic represents the area under the receiver operating characteristic (ROC) curve, a plot of sensitivity against specificity. Values can range from 0.5 (no predictive ability) to 1 (perfect discrimination).

Lastly, in an analysis of correctly and incorrectly classified patients, the average values of FEV_1%-predicted were compared per severity stage.

10.3 Results

Description of sample

Patient characteristics are summarized in table 10.1a, demonstrating that 47% were in GOLD stage II, 44% in stage III, and 9% in stage IV. Patients in more severe disease states were more likely to have a lower BMI, to have quit smoking and to suffer from osteoporosis. The mean age of patients in stage IV was 2–3 years lower than in the less severe categories. The disease stage was not significantly associated with pack-years, the number of co-morbidities, Charlson's comorbidity index, and various concomitant diseases.

Tables 10.1b and 10.1c show that disease stage was associated significantly with almost every type of medication and resource use. Patients with more severe COPD were more likely to use various types of pulmonary maintenance medication, oxygen at home, exacerbation medication, consult their GP more often (on a scheduled and unscheduled basis), visit the emergency room with resulting admission more often, and were more likely to have

Table 10.1a Patient characteristics, risk factors, and comorbidity per GOLD stage in full sample (development and validation sets combined)

Stage	II	III	IV	P-value*
	n=1720 (47%)	n=1640 (44%)	n=338 (9%)	
Age (SD)	64.9 (8.47)	65.8 (7.98)	63.0 (8.23)	0.000
Male	70.10%	72.80%	75.40%	0.066
BMI<21	9.94%	12.56%	26.63%	0.000
BMI>25	62.15%	56.58%	34.61%	
Underweight (BMI<18.5)	2.44%	3.84%	10.06%	0.000
Low weight (18.5>BMI<21)	7.50%	8.72%	16.57%	
Normal BMI	27.91%	30.85%	38.76%	
Overweight (25<BMI<30)	36.92%	36.45%	23.37%	
Obese (BMI>30)	25.23%	20.12%	11.24%	
Current smoker	33.43%	28.23%	27.51%	0.002
<20 Pack years	6.63%	5.85%	7.69%	0.700
20-60 pack years	66.86%	66.95%	64.79%	
>60 pack years	26.51%	27.20%	27.51%	
Comorbidities: 0	12.67%	12.20%	11.54%	0.752
1	14.59%	15.30%	17.46%	
2	15.52%	14.27%	12.72%	
3-5	27.38%	27.80%	29.29%	
6-9	18.31%	29.29%	19.82%	
Charlson comorbidity index: 1	69.36%	69.51%	74.56%	0.398
2	21.28%	21.16%	15.98%	
3	6.40%	5.91%	5.92%	
>=4	2.97%	3.41%	3.55%	

Table 10.1a (continued)

Stage	II	III	IV	P-value*
	n=1720 (47%)	n=1640 (44%)	n=338 (9%)	
Coronary heart disease	14.13%	13.41%	12.43%	0.657
Vascular disease	45.29%	43.78%	38.46%	0.067
Hypertension	41.05%	40.00%	35.50%	0.163
Nervous system disorders	14.65%	14.45%	15.68%	0.844
Stroke	0.41%	0.39%	0.30%	0.865
Diabetes	9.24%	8.35%	5.92%	0.126
Depression	10.64%	10.24%	11.54%	0.768
Anemia	0.93%	1.10%	0.30%	0.381
Platelet disorders	0.29%	0.18%	0.00%	0.533
Osteoporosis	5.64%	8.54%	8.28%	0.004
Cataract	2.50%	2.32%	3.55%	0.419

* Two-sided p-values from analysis of variance for continuous variables and from the chi-squared test for categorical variables.

Abbreviations: BMI, body mass index; SD, standard deviation; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

Table 10.1b Medication per GOLD stage in full sample (development and validation sets combined)

Stage	II	III	IV	P-value*
	n=1720 (47%)	n=1640 (44%)	n=338 (9%)	
No short-acting bronchodilators	33.78%	20.91%	13.31%	0.000
1 short-acting bronchodilator	36.10%	37.56%	26.04%	
2 short-acting bronchodilators	30.12%	41.52%	60.65%	
Long-acting bronchodilator	62.03%	70.55%	75.74%	0.000
Inhaled corticosteroids	62.21%	69.51%	76.33%	0.000
Other steroids	5.76%	10.00%	17.16%	0.000
Xanthine	9.94%	19.33%	20.12%	0.000
Leukotrin modifier	2.79%	5.67%	8.58%	0.000
Mucolytics	4.19%	5.85%	5.33%	0.084
Home oxygen	1.05%	3.05%	9.17%	0.000
Statins	19.13%	16.52%	13.61%	0.020
Cardiovascular medication	51.16%	54.82%	48.82%	0.020
Courses of oral steroids: 0	70.76%	59.45%	52.66%	0.000
1	17.44%	22.68%	23.08%	
2	6.22%	9.02%	8.84%	
>=3	5.58%	8.84%	11.54%	
Courses of antibiotics: 0	53.31%	44.27%	43.20%	0.000
1	22.97%	24.94%	25.15%	
2	12.91%	16.10%	16.27%	
>=3	10.81%	14.70%	15.38%	

* Two-sided p-values from the chi-squared test for categorical variables

Abbreviation: GOLD, Global Initiative for Chronic Obstructive Lung Disease.

Table 10.1c Health care resource use per GOLD stage in full sample (development and validation sets combined)

Stage	II	III	IV	P-value*
	n=1720 (47%)	n=1640 (44%)	n=338 (9%)	
Scheduled GP visits: 0	20.17%	14.82%	12.43%	0.000
1	18.72%	15.85%	15.09%	
2	23.43%	23.05%	17.75%	
3-4	23.49%	28.78%	31.36%	
>=5	14.19%	17.50%	23.37%	
Unscheduled GP visits: 0	74.71%	70.12%	64.50%	0.003
1	13.14%	16.46%	18.64%	
2	6.40%	6.65%	9.17%	
>=3	5.76%	6.77%	7.69%	
Emergency room (no admission): 0	92.40%	90.47%	89.88%	0.071
1	4.97%	6.02%	8.04%	
>=2	2.63%	3.50%	2.08%	
Emergency room and admission: 0	91.87%	88.19%	82.74%	0.000
1	6.26%	9.04%	11.90%	
>=2	1.87%	2.77%	5.36%	
Direct hospital admissions (yes/no)	3.39%	4.55%	7.14%	0.006

* Two-sided p-values from the chi-squared test for categorical variables

been admitted for any reason to the hospital in the last year. The use of statins and other cardiovascular medication was more frequent in patients with less severe disease. There was no difference between the dataset used for developing the algorithm and the validation dataset.

Regression results

In the final model (table 10.2), the parallel regression assumption was violated for five variables (age, gender, xanthine and oxygen use, and two categories of the BMI variable, overweight and obese). In these cases, a coefficient was estimated for each dichotomization (GOLD III/IV versus GOLD II and GOLD IV versus GOLD II/III). For all other variables, one coefficient was estimated for both dichotomizations.

In the multivariable analysis, a higher risk of more severe COPD was significantly associated with being younger, being male, having a lower BMI, having quit smoking, suffering from osteoporosis, using oxygen, courses of oral steroids, having been hospitalized in the previous year, and certain types of respiratory maintenance medication (long- and short-acting bronchodilators, xanthines, leukotriene modifiers, oral steroids).

The thresholds were not statistically significantly different from 0 (p-values: 0.466 and 0.338, see table 10.2). However, the first threshold was significantly lower than the second (–0.251 versus 0.403, p-value 0.042).

Table 10.2 Regression results, final partial proportional odds ordered logit model

	Parallel regression assumption holds		Parallel regression assumption violated			
	Shared by both dichotomizations		Stages III/IV vs stage II		Stage IV vs stages II/III	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Age^2			-0.00007	0.122	-0.00055	<0.0005
Male			0.423	<0.0005	0.848	<0.0005
Underweight (BMI<18.5)	Reference category					
Low weight (18.5>BMI<21)	-0.374	0.110				
Normal BMI	-0.699	0.001				
Overweight (25<BMI<30)			-0.951	<0.0005	-1.775	<0.0005
Obese (BMI>30)			-1.187	<0.0005	-1.912	<0.0005
Current smoker	-0.155	0.097				
Osteoporosis	0.331	0.042				
Longacting bronchodilator	0.262	0.003				
No short-acting bronchodilator	Reference category					
1 short-acting bronchodilator	0.338	0.001				
2 short-acting bronchodilators	0.854	<0.0005				
Leukotrin modifier	0.661	0.001				
Xanthine			0.615	<0.0005	-0.076	0.706
Oral steroids (maintenance)	0.304	0.045				
No incidental courses of oral steroids	Reference category					
1 course of oral steroids	0.222	0.037				
2 courses of oral steroids	0.292	0.064				
≥3 courses of oral steroids	0.218	0.184				
Oxygen			0.759	<0.0005	1.599	<0.0005
Any hospital admissions in previous year	0.349	0.081				
Threshold			-0.251	0.466	0.403	0.338
N	2423					
Log likelihood	-2057.973					
Wald test for model significance	<0.00005					

Abbreviation: BMI, body mass index.

Table 10.3 Predictive performance: predicted and observed GOLD stage in validation data set

a. PPO ordered logit model estimated in full development data set								
		Predicted GOLD stage						
		II	**III**	**IV**	**Total**		**Unweighted**	**Weighted**
Observed GOLD stage	**II**	368	205	7	580	**Overall agreement**	53%	77%
		63%	*35%*	*1%*	*100%*			
						Kappa	0.151	0.230
	III	260	283	4	547	**95% CI**	0.103 - 0.198	0.196 - 0.263
		46%	*53%*	*1%*	*100%*			
	IV	28	82	10	120			
		24%	*68%*	*8%*	*100%*			
	Total	656	570	21	1,247			
b. PPO ordered logit model estimated in balanced sample (sensitivity analysis)								
		Predicted GOLD stage						
		II	**III**	**IV**	**Total**		**Unweighted**	**Weighted**
Observed GOLD stage	**II**	64	34	23	121	**Overall agreement**	48%	68%
		53%	*28%*	*19%*	*100%*			
						Kappa	0.215	0.290
	III	42	40	38	120	**95% CI**	0.142 - 0.288	0.205 - 0.370
		35%	*33%*	*32%*	*100%*			
	IV	20	32	68	120			
		17%	*27%*	*57%*	*100%*			
	Total	126	106	129	361			
c. Binary logit model estimated in development data set								
		Predicted GOLD stage						
		II	**III/IV**	**Total**				
Observed GOLD stage	**II**	316	264	585		**Overall agreement**	62%	
		54%	46%	100%				
						Kappa	0.228	
	III/IV	212	455	667		**95% CI**	0.173 - 0.284	
		32%	68%	100%				
	Total	528	719	1,247		**c-statistic**	0.614	
						95% CI	0.587 - 0.640	

Abbreviations: CI, confidence interval; Gold, Global Initiative for Chronic Obstructive Lung Disease.

The following variables were not maintained in the final model: inhaled corticosteroids, mucolytics, cardiovascular medication, courses of antibiotics, most co-morbidities (except osteoporosis), and healthcare resource use (emergency room, GP consultations).

Predictive performance

Sixty three percent of patients in stage II, 53% of patients in stage III and 8% of patients in stage IV were correctly identified with the final model (see table 10.3a). The overall agreement was 53% (unweighted) to 77% (weighted), leading to 'slight' to 'fair' kappa statistics of 0.151 (unweighted) and 0.230 (weighted).

The observed values of FEV_1%-predicted were slightly higher for COPD stage II patients who were correctly classified by the model than for patients who were incorrectly classified (59.4% versus 57.7%, respectively). In stage III, the opposite was observed (FEV_1%-predicted was 39.4% in the correctly classified versus 41.2% in the incorrectly classified), and in stage IV, FEV_1 was comparable among the correctly and incorrectly classified groups (24.9% and 25.6%).

Sensitivity analyses

In the balanced sample (regression results not presented), agreement was lower for stages II and III than in the unbalanced sample, but higher for stage IV: 53%, 33% and 57% respectively (table 10.3b). The overall agreement was 48% (unweighted) to 68% (weighted). The kappa statistics were 'fair' with 0.215 and 0.290, respectively.

In the binary logit model (table 10.3c), agreement for patients with moderate and severe/very severe COPD was 54% and 68%. The kappa-statistic was 'fair': 0.228.

10.4 Discussion

Our study has two important findings. Firstly, the variables that were independently related to more severe COPD defined by lung function were a lower age, male gender, a lower BMI, being an ex-smoker, having osteoporosis, having been hospitalized in the previous year, using oxygen and certain types of respiratory maintenance medication (long- and short- acting bronchodilators, xanthines, leukotrien modifiers, oral steroids). Other variables expected to be associated with lung function impairment, such as other resource use variables, were not maintained in the final model. Secondly, the performance of the final model was such that the confidence in using the selected variables to adjust for COPD severity in the absence of lung function parameters was judged to be limited.

Of the variables in the final model, the impact of age may be partly due to a healthy survivor effect, whereas long acting bronchodilators, multiple short-acting bronchodilators, oral steroids and xanthines are clearly indicated for treating more advanced stages of COPD.

Patients with more severe disease are more likely to have quit smoking. A low BMI, often associated with loss of muscle mass is well known to be more frequent in severe COPD and the higher prevalence of osteoporosis might be a side-effect of a long history of corticosteroid use.

We used a partial proportional odds model instead of the standard ordered logit model. This made it possible to estimate different coefficients for the probability of being in stage III/IV over stage II than for the probability of being in stage IV, if this was required. The parallel regression assumption, which states that the coefficients are equal for both dichotomizations, was violated in five instances.

In the final base case model, only 53% of patients were classified in the correct GOLD stage. Especially for patients in stage IV, the predicted stage was unlikely to be correct (8% correct). Results from sensitivity analyses with a more balanced sample or a binary logit model were only slightly better.

We chose to present both unweighted and weighted kappa statistics. In cases with more than two categories, it is customary to weight the kappa-statistics in order to penalize disagreements in terms of their seriousness – i.e., a higher penalty for misclassifying a patient from stage II as a stage IV patient than as a stage III patient – whereas unweighted kappa treats all disagreements equally. However, in our study all disagreements were considered serious and the weighted overall agreement might give an overly favourable impression. In the end, the unweighted and the weighted kappa statistics were quite similar. Based on all values of the kappa statistics, the agreement between predicted and observed GOLD stages can be characterized as slight to fair.

Several explanations for these findings may be considered. Firstly, the cut-off points between GOLD stages are inevitably somewhat arbitrary and artificial, especially as the decline in FEV_1 is a continuous process. A patient with a FEV_1%-predicted of 49% (GOLD stage III) is probably less similar to a patient with a value of 31% (also GOLD stage III) than to someone with a value of 51% (GOLD stage II). If this were an important explanation for our prediction results, misclassified patients in stage IV should have markedly better lung functions – closer to stage III – than the correctly classified patients, while the opposite should be true for misclassified patients in stage II. However, the actual differences in lung function between the correctly and incorrectly classified patients in our analysis were small (1.74%-point for stage II and 0.66%-point for stage IV).

The second explanation concerns the source of our data, which was a clinical trial. It is conceivable that patients with very severe disease and many symptoms were less willing to participate in the trial. This could then obscure some of the associations between disease severity and the predictors. This problem would be expected to occur particularly when trying to predict GOLD stage IV. However, our models did not perform well at distinguishing moderate from severe patients either. In the balanced sample analysis, misclassifications occurred equally often in each GOLD stage. Furthermore, the proportion of very severe patients in the sample does not appear to be low compared to the proportion in the gen-

eral population of COPD patients in Finland [34], The Netherlands [35], Greece [36], the United Kingdom [37], and a combination of European and North-American countries [38]. Moreover, a broad range of patients was allowed to participate. For examples, patients were permitted to use all respiratory medications concomitantly during the trial except inhaled anticholinergic drugs, thus closely resembling routine care. Altogether, this protocol makes the UPLIFT a suitable trial for this study.

The third explanation would be the far from perfect association of resource use, which is often driven by symptoms and exacerbations, with GOLD stage. Patients with a relatively good lung function do not necessarily experience fewer symptoms than patients with worse lung function. Indeed, the UPLIFT sample contains a non-negligible number of patients with very severe COPD who do use little or no maintenance medication, as well as patients with moderate disease who use four or five different types of medications. Overall, we observed an independent association of GOLD stage with several respiratory medications, but not with other types of resource use such as ER visits and physician consults in our multivariable model, which was not expected beforehand. In univariate analyses, patients with more severe COPD were more likely to have higher COPD-related resource use (almost all types of medication use, courses of oral steroids and antibiotics, scheduled and unscheduled GP visits and hospital admissions with and without visits to the emergency room), not to use statins and other cardiovascular medication, to be younger, male, underweight, ex-smoker, and suffer from osteoporosis.

Several database studies have attempted to adjust for possible differences in COPD severity in the absence of lung function data. In an article comparing the assessment of COPD patients in the UK General Practice Research Database with the clinical opinion of the patient's GP, Soriano et al. [18] based their severity classification on medication use. Sin and Tu [19] assessed the effects of ICS use on mortality and applied medication use, ER visits and physician services as proxies of disease severity. Similarly, Suissa [20] adjusted only for age, sex and medication use. Breekveldt-Postma et al. [21] identified determinants of patient's persistence with ICS therapy. They considered hospital admissions and medication use to be proxies for disease severity. In a study relating COPD severity with cardiovascular disease, Curkendall et al. [22] assumed that COPD severity could be defined as the likelihood of being hospitalized, given the relationship with mortality. They concluded that this probability of hospitalization is associated with medication and oxygen use, previous hospitalizations, recent exacerbations, pneumonia and lung emphysema. Griffin et al. [23] assessed the effects of tiotropium compared to combined ipratropium and salbutamol on exacerbations and hospitalizations. They adjusted for a combination of resource use data and risk factors. Based on the current analyses, we conclude that the variables used in these database studies cannot be relied upon to adequately adjust for COPD severity in terms of lung function.

In conclusion, data from a well-controlled trial setting indicated that COPD severity defined by lung function thresholds cannot be reliably predicted from patient characteristics

and their previous healthcare use. This limitation should be considered when interpreting findings from database studies, and additional research should explore other methods allowing accounting for COPD severity.

10.5 References

1 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of COPD. Available at: http://www.goldcopd.com. 2011.

2 Anthonisen NR, Wright EC, Hodgkin JE. Prognosis in chronic obstructive pulmonary disease. Am Rev Respir Dis. 1986; 133: 14-20.

3 Burge PS, Calverley PM, Jones PW, et al. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. BMJ. 2000; 320: 1297-303.

4 Dewan NA, Rafique S, Kanwar B, et al. Acute exacerbation of COPD: factors associated with poor treatment outcome. Chest. 2000; 117: 662-671.

5 Andersson F, Borg S, Jansson SA, et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). Respir Med. 2002; 96: 700-8.

6 Jenkins C, Rodriguez-Roisin R. Quality of life, stage severity and COPD. Eur Respir J. 2009; 33: 953-955.

7 Stahl E, Lindberg A, Jansson SA, et al. Health-related quality of life is related to COPD disease severity. Health Qual Life Outcomes. 2005; 3: 56.

8 Ekberg-Aronsson M, Pehrsson K, Nilsson JA, et al. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. Respir Res. 2005; 6: 98.

9 Ekberg-Aronsson M, Lofdahl K, Nilsson JA, et al. Hospital admission rates among men and women with symptoms of chronic bronchitis and airflow limitation corresponding to the GOLD stages of chronic obstructive pulmonary disease--a population-based study. Respir Med. 2008; 102: 109-120.

10 Antonelli-Incalzi R, Imperiale C, Bellia V, et al. Do GOLD stages of COPD severity really correspond to differences in health status? Eur Respir J. 2003; 22: 444-449.

11 Medinas Amoros M, Mas-Tous C, Renom-Sotorra F, et al. Health-related quality of life is associated with COPD severity: a comparison between the GOLD staging and the BODE index. Chron Respir Dis. 2009; 6: 75-80.

12 Rutten-van Mölken MPMH, Oostenbrink JB, Tashkin DP, et al. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? Chest. 2006; 130: 1117-28.

13 Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med. 2004; 350: 1005-1012.

14 Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. Lancet. 2009; 374: 704-711.

15 Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. Eur Respir J. 2009; 34: 648-654.

16 Briggs A, Spencer M, Wang H, et al. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. Arch Intern Med. 2008; 168: 71-79.

17 Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. Chest. 2005; 127: 1890-1897.
18 Soriano JB, Maier WC, Visick G, Pride NB. Validation of general practitioner-diagnosed COPD in the UK General Practice Research Database. Eur J Epidemiol. 2001; 17: 1075-1080.
19 Sin DD, Tu JV. Inhaled corticosteroids and the risk of mortality and readmission in elderly patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001; 164: 580-584.
20 Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. Am J Respir Crit Care Med. 2003; 168: 49-53.
21 Breekveldt-Postma NS, Gerrits CM, Lammers JW, et al. Persistence with inhaled corticosteroid therapy in daily practice. Respir Med. 2004; 98: 752-759.
22 Curkendall SM, Lanes S, de Luise C, et al. Chronic obstructive pulmonary disease severity and cardiovascular outcomes. Eur J Epidemiol. 2006; 21: 803-813.
23 Griffin J, Lee S, Caiado M, et al. Comparison of tiotropium bromide and combined ipratropium/salbutamol for the treatment of COPD: a UK General Practice Research Database 12-month follow-up study. Prim Care Respir J. 2008; 17: 104-110.
24 Decramer M, Celli B, Tashkin DP, et al. Clinical trial design considerations in assessing long-term functional impacts of tiotropium in COPD: the UPLIFT trial. COPD. 2004; 1: 303-312.
25 van Schayck CP, Halbert RJ, Nordyke RJ, et al. Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting. Respirology. 2005; 10: 323-333.
26 Tinkelman DG, Price DB, Nordyke RJ, et al. Symptom-based questionnaire for differentiating COPD and asthma. Respiration. 2006; 73: 296-305.
27 Price DB, Tinkelman DG, Nordyke RJ, et al. Scoring system and clinical application of COPD diagnostic questionnaires. Chest. 2006; 129: 1531-1539.
28 Martinez FJ, Foster G, Curtis JL, et al. Predictors of mortality in patients with emphysema and severe airflow obstruction. Am J Respir Crit Care Med. 2006; 173: 1326-1334.
29 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40: 373-83.
30 Long JS. Regression Models for Categorical and Limited Dependent Variables (Advanced Quantitative Techniques in the Social Sciences) . Thousands Oaks: Sage, 1997.
31 Williams R. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. The Stata Journal. 2006; 6: 58-82.
32 Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates & Proportions (3rd ed.). New York: John Wiley & Sons, 2003.
33 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159-174.
34 Vasankari TM, Impivaara O, Heliovaara M, et al. No increase in the prevalence of COPD in two decades. Eur Respir J. 2010; 36(4): 766-73.
35 Hoogendoorn M, Feenstra TL, Schermer TR, et al. Severity distribution of chronic obstructive pulmonary disease (COPD) in Dutch general practice. Respir Med. 2006; 100: 83-86.
36 Minas M, Hatzoglou C, Karetsi E, et al. COPD prevalence and the differences between newly and previously diagnosed COPD patients in a spirometry program. Prim Care Respir J. 2010;
37 Shahab L, Jarvis MJ, Britton J, West R. Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample. Thorax. 2006; 61: 1043-1047.
38 Agusti A, Calverley PM, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. Respir Res. 2010; 11: 122.

General discussion

11

11.1 Introduction

This thesis addressed several methodological issues in the economic evaluation of healthcare. Furthermore, it contained an evaluation of a hospital-at-home program for COPD patients. This chapter provides an overview of the main findings, with an emphasis on the methodological elements, and puts them into perspective.

11.2 Hospital-at-home after a COPD exacerbation

The objective of the first part of the thesis was to perform an economic evaluation of the GO AHEAD trial of early assisted discharge of COPD patients after an exacerbation, to put the estimate of cost savings into perspective and to investigate treatment preferences of patients and informal caregivers.

In the GO AHEAD trial, the costs and health effects of two treatments were compared for patients who were admitted to the hospital with a COPD exacerbation in the Netherlands. Participants stayed in the hospital for seven days, or went home after three days where they were supervised and treated by community nurses.

In chapter 4, we found that transferring hospital care for a COPD exacerbation to the patient's home was likely to lead to modest savings in healthcare costs in the Netherlands, while there is no clear evidence to conclude that it would lead to a slower or worse resolution of the exacerbation. From a societal perspective – when the societal costs of informal care and productivity losses are taken into account – these cost savings decrease considerably or even turn into cost increases.

There were differences in mean generic and disease-specific health-related quality of life at the end of the initial treatment phase, but these were quite small, temporary and only partially significant. By the end of the three-month follow-up period they had disappeared. The early-assisted discharge program was likely to lead to modest savings in healthcare costs.

The health outcomes that we found were consistent with those of other studies of hospital-at-home interventions. However, most other authors found larger cost savings than we did. Part of this inconsistency could be attributed to differences in the design of the early assisted discharge or admission avoidance programs, to different patterns of usual treatment and to different cost structures across countries. In addition, in chapter 2, we also concluded that savings in most studies were overestimated. The quality of most of the papers that we selected in our systematic review was not high. The large majority assumed unit costs per inpatient hospital day that were too high, which led to an overly optimistic estimate of cost savings due to the reduction of length of stay. Many studies did not take into account that treatment intensity decreases over the course of an admission, which makes the last days

less costly than the first days. Furthermore, some studies did not adapt their cost estimate to the particular disease or intervention that is investigated.

Ultimately, cost savings for hospital-at-home occur when the costs of homecare are less than the costs of the avoided hospital days. One way of achieving this is simply to provide less care at home than in the hospital. In the case of COPD exacerbations, that means that there are fewer minutes of contact between the patient and healthcare professionals. The other way is to deliver the same amount of care more efficiently. In the accomplishment of the latter, a number of factors can play a role: lower hourly costs for personnel, higher productivity (the proportion of working time devoted to patient care), lower overhead costs, less expenditure on housing or capital. In the GO AHEAD trial the amount of professional care, measured in days as well as in minutes, was similar for both treatment groups. That means that the cost savings must have come from efficiency gains. Furthermore, the total amount of care, professional and informal combined, was larger for patients who were discharged early.

The GO AHEAD trial was not an equivalence trial. We cannot be completely certain that the early assisted discharge intervention does not lead to worse health outcomes than usual hospital care. Indeed, both health measures indicated that patients in the intervention group had on average a somewhat worse health status on day 7 and the confidence interval of the CCQ score contained clinically relevant values. However, we have argued that this difference might be explained by differences in expectations when leaving the hospital. Furthermore, the difference does not mean that the intervention is unsafe.

Whereas there is no compelling reason – from a medical or economic point of view – to recommend either the early supported discharge treatment or usual hospital care to all patients, many patients and informal caregivers did show a strong preference for one or the other option. This was presented in chapter 5. Only for a minority of respondents this preference could be changed by altering the characteristics of the proposed early assisted discharge treatment. This does not mean that the design of the intervention did not matter to the majority of patients and caregivers. In a choice between two early assisted discharge programs, copayments and the burden on the informal caregiver had a strong impact on preferences.

11.3 The importance of informal care

The analysis from the societal perspective in chapter 4 made clear how crucial the inclusion of informal care costs can be for the conclusions on cost savings. For the initial treatment episode, the projected savings disappeared almost completely when the costs for informal care were added to the calculations. For the full three-month follow-up period, the savings even turned into estimated costs increases.

However this points to one of the problems concerning the costs of informal care. Respondents in the early assisted discharge group reported much larger amounts of informal care than respondents in the usual hospital care group. In contrast, there were no differences in mean health outcomes over the same period. There are two possible explanations for the discrepancy. If the responses reflected an actual difference in care that was provided, this might be explained by a different attitude of the informal caregivers that was formed during the initial treatment period, in which they took over part of the hospital care. Unfortunately, it is also plausible that informal caregivers in the early assisted discharge group were more primed to record their activities as informal care, due to the attention it may have got during visits of the homecare nurse in the initial treatment phase at home.

A second matter is the valuation of the time of the informal caregivers. Following Dutch guideline recommendations, we used a shadow price of €12.50 per hour, which was based on the hourly costs of cleaning services provided by homecare organizations. This amount is in the lower part of a range of studies with different valuation methods in various patient groups [1,2]: €8 to €20, indexed to the 2010 price level. It could be argued that our price led to a somewhat conservative estimate of the costs of informal care and therefore, to a slightly optimistic estimate of the cost-effectiveness of the hospital-at-home program. At any rate, the decision on the price is debatable. However, even if the unit price of an hour of informal care has not been established, it is clear that this time is valuable. Ignoring the costs of informal care would give the incorrect impression that informal care is costless.

11.4 The multidimensionality and heterogeneity of inpatient hospital costs

In economic evaluations, costs are often calculated by multiplying an amount of healthcare resources with their unit price. Inpatient hospital days are a special type of resource use in that context because of their multidimensionality. In fact, an inpatient day is not a 'unit' of healthcare resources. Rather, it is a combination of many different inputs: time from physicians, nurses and other staff, materials, depreciation of medical equipment, medication, food, heating, administrative overhead, cleaning, capital costs and more.

This combination varies per patient, as well as across diseases and interventions. It is therefore obvious that disease-specific unit cost prices must be used, although not all studies in our review of hospital-at-home costs did so. However, they may still be insufficiently tailored to the patient group that is treated. In hospital-at-home studies, for instance, the target population is not the average patient with a certain disease. Patients who are eligible for hospital-at-home are likely to need less care than this average patient. This means that the actual savings for hospital-at-home are likely to be overestimated when average disease-

specific prices per inpatient hospital day are used. This stricter criterion of target-patient unit costs was however not applied in our systematic review in chapter 2.

Of course, the multidimensionality problem applies to other resources besides inpatient hospital days. As an example, the costs of a general practitioner's consult contain more than the physician's time. However, the heterogeneity among individual patients with regards to the contents of a consult is more limited. This means that the average costs per consult can be calculated by averaging the total costs of a practice or the total national costs over the number of consults.

The cost analysis of the GO AHEAD trial in chapter 4 showed that assumptions on the costs of inpatient hospital days can be crucial for the conclusions. In our base case, we took the microcosting approach, in which detailed cost calculations were made for inpatient hospital days in this specific patient population. From a societal perspective, cost savings were negligible for the initial treatment phase and turned into cost increases when the full three-month period was considered. In contrast, the use of average, non-disease-specific reference costs per inpatient day without adjustment for the course of the admission dramatically increased the estimated savings for the initial phase as well as the follow-up period.

11.5 Quantifying uncertainty

The uncertainty around estimates of incremental costs and effects can be analysed by applying the bootstrapping method for sample data and by probabilistic sensitivity analysis in modelling studies. The results of these analyses can be presented graphically in cost effectiveness planes and acceptability curves.

When sample data are used, the uncertainty estimates result from variation within the sample. If the differences in health outcomes or resource use among patients are small, so is the uncertainty around the estimates of the mean effects and costs. The bootstrapping method brings this to the surface by repeatedly taking samples with replacement from the original dataset. The estimates and costs and effects differ for each bootstrap sample, depending on the patients it contains.

However, in order to appreciate the full uncertainty around the estimates, the multidimensionality of inpatient hospital days and their costs must be acknowledged.

The combination of staff time, use of equipment and overhead costs is different for each patient, so it is technically incorrect to apply one estimate of the mean costs to all patients and ignore the variation. Ideally, patient-level data on each element of healthcare resource use should be available for each separate inpatient hospital day. Obviously, the collection of these data is not feasible in most circumstances.

An alternative approach is combining univariate sensitivity analyses with bootstrapping, as we did for the GO AHEAD trial in chapter 4: the bootstrapping procedure was repeated

for various levels of the unit costs per hospital admission. The results were presented in one figure with cost-effectiveness acceptability curves and interpreted in combination. Taken separately, the results from each separate bootstrapping procedure underestimate the uncertainty around the estimated incremental costs and the incremental cost-effectiveness ratio since it is based on a fixed price for the hospital admission. In the GO AHEAD study, this became apparent from the cost-effectiveness planes for the initial treatment episode, in which the costs of inpatient hospital days were one of very few cost drivers. These planes showed virtually no uncertainty about the incremental costs, which is unrealistic and which is why they were not included in this thesis. The planes for the full three-month period appeared correct at first sight, but they also do not express the full extent of the uncertainty involved in estimating the incremental costs.

In the probabilistic sensitivity analysis in health economic models, uncertainty around the costs can be taken into account by applying a distribution from which the costs for hospital admissions are sampled. However, in the absence of patient-level data of unit costs for inpatient hospital days, this distribution would have to be based on even less than expert opinion.

Of course, the underestimation of uncertainty around hospital costs is only a problem when the costs of hospital admissions is an important driver of total cost estimates.

11.6 Quantifying preferences and preference heterogeneity

Patients' preferences for aspects of medical treatments can be investigated in a great many ways [3-5]. The simplest method may be to ask respondents how important each characteristic is to them, and possibly express this on a scale. The great advantage of discrete choice experiments (DCEs) compared to simpler preference questionnaires is that they can be used to quantify the value of treatment characteristics. A DCE makes the trade-off between desirable elements explicit. The results of the analysis show which attributes are considered important, but also what their relative position is among the others and what their value is. It is consistent with notions of scarcity of resources and based on economic and utility theory [6]. A DCE suffers from the same weakness as other questionnaires: it is a stated preference approach [7]. There is no guarantee that respondents would make the same choices if they were offered to them in real life. On the other hand, stated preference data also have advantages compared to observing actual choices (revealed preferences). The most important point is that stated preferences make it possible to evaluate interventions that are not yet as such provided to patients. Furthermore, the characteristics of current real-life interventions may show little variability or be highly collinear, which makes it impossible to correctly analyse the importance of each of them.

The questionnaires in the DCE in chapter 5 described potential future treatments of COPD exacerbations. An analysis of revealed preference would not have been possible.

The data was analysed using latent class conditional logit models. This is quite a rarity in health economics, where the multinomial logit (or conditional logit) and mixed logit are the common techniques for analyzing DCE data. Latent class analysis was applied in health economics only twice before our study [8].

It was shown in chapter 5 of this thesis how using the conventional conditional logit model would have led to false conclusions: it would have appeared as if all respondents could have been persuaded to choose the hospital or the early assisted discharge option, if the characteristics of the homecare program were adjusted to their preferences. In reality, the majority had a fixed preference for one of the option, regardless of the characteristics of the early assisted discharge program. These respondents, however, were willing to choose between different early assisted discharge programs.

Preference heterogeneity across respondents can be addressed in in a mixed logit model – and it often is – which assumes that the individual preferences for each attribute follow a certain distribution. However, the raw data from the experiment in GO AHEAD showed a clear segmentation of respondents rather than differences in strength of preferences. The classes that were identified by the latent class conditional logit models were to a large extent characterized by their *a priori* preference for or aversion to hospital admissions or early assisted discharge.

11.7 Health-related quality of life during moderate COPD exacerbations

The study in chapter 6 showed that the EQ-5D questionnaire for the measurement of health-related quality of life can be used for assessing the health of COPD patients during a moderate exacerbation, which was defined as a worsening of their disease leading to a prescription of antibiotics and/or oral steroids but not to a hospital admission. The EQ-5D scores were responsive to the recovery from a moderate exacerbation. Three-quarters of patients experienced an improvement in health-related quality of life after the worst day of the exacerbation. All five dimensions of the EQ-5D contributed significantly to the mean improvement in the EQ-5D index score. The EQ-5D results were at least as responsive as the symptoms scores and rescue medication use.

In a tentative calculation, it was estimated that the average QALY loss during a moderate exacerbation could be 0.00896, or 3.27 quality-adjusted life days. At first sight, this number may appear negligibly small. However, if a QALY is valued at €20,000, the maximally acceptable costs for avoiding a moderate exacerbation would be €179. It is very plausible that certain therapies could deliver the health gain for these costs.

The aim was not to investigate if the EQ-5D could be useful in studies, not to calculate a standard value of the health losses that are suffered by patients during a moderate COPD

exacerbation. The latter was impossible and inappropriate for several reasons. The utilities that result from the EQ-5D are based on country-specific value sets, so a world-wide value for this health loss does not exist. Nevertheless, this might have been addressed by calculating different values according to different value sets.

Still, this country-specific estimate for the average health loss would be based on the health states in a specific sample. That means that it is, in principle, applicable only to similar patients. Quality of life and and changes between measurements are likely to be associated with the characteristics of the patients in the sample, especially severity of disease and co-morbidities. Even within a country, there is not one value for the health loss due to a moderate COPD exacerbation.

A third reason for not presenting the health loss we calculated as the best available estimate of the QALY loss due to a moderate exacerbation is that we did not have estimates of respondents' health-related quality of life during the stable phase of their disease. This would have been the correct value for the comparison with the scores during the exacerbation. Instead, we used the measurement after the exacerbation had been resolved as a proxy.

In this version of the EQ-5D, respondents chose between three levels on each dimension (no/some/severe problems). The new version of the questionnaire has five levels (no/slight/moderate/severe/extreme problems). It is conceivable that this EQ-5D 5L is even more sensitive to the recovery from a moderate exacerbation than the three-level version if respondents are more willing to switch from, for instance, 'no problems' to 'slight problems' than to 'some problems'. In that case, the new version would be able to detect smaller changes in health status. However, if patients were already willing to switch in the three-level version, the measured changes in the five-level version would be smaller and the estimated health losses due to a moderate exacerbation would become smaller than in the three-level EQ-5D.

11.8 The use of propensity scores in economic evaluations

As an alternative to conventional regression analysis, propensity-score based methods have gained popularity in observational studies in recent decades. These techniques are primarily applied in epidemiological and, to a lesser extent, in medical studies. Examples of their use in economic evaluations are still scarce. The objective of chapters 8 and 9 was to compare the performance of several propensity-score adjustment methods and conventional regression in the context of an economic evaluation based on observational time-to-event data.

The propensity score is defined as the probability of receiving a certain treatment conditional on a patient's characteristics. It can be used in several ways to overcome differences in observed characteristics of patients across treatment groups. Covariate adjustment is the most popular application: the propensity score is used in a regression model instead of the baseline characteristics themselves. Other applications include weighting and matching.

It is conceptually important to distinguish between methods that estimate an Average Treatment effect in the Treated patients (ATT) and methods that estimate an Average Treatment Effect in the population as a whole (ATE). The ATT is focused on the patients who actually received the treatment of study: what would have been the difference if they had been treated differently. The ATE is the expected gain for a randomly selected subject from the population. It can equivalently be defined as the difference between the hypothetical situations that either of the treatments had been assigned to the entire population. ATT and ATE are different from each other if treatment is more effective in patients with certain characteristics than in others, i.e. when effect modification occurs.

When the data can be described by a parametric multiplicative model, such as the Weibull survival models in this thesis, effect modification is always present. This is due to the assumption that the treatment effect is proportional to the severity of the impairment in health status in absence of treatment. The hazard of dying is decreased by a proportion of the original hazard – this is the proportional hazards assumption; life is prolonged by a proportion of the prognosis without treatment. Even if this proportion is constant across subgroups of patients – so the hazard ratio or the regression coefficient of treatment is not affected by effect modification – the health gains still vary in absolute terms. Patients with a good prognosis have more survival time to gain in absolute terms, while the proportional gains are the same for other subgroups.

In the context of the Dutch policy for conditional reimbursement to hospitals of the costs of expensive medications, the ATT is probably the relevant treatment effect: what is the effect in the patient group that has been given the new treatment? This requires two assumption or observations: the treated subjects form a good representation of the patients to which the new treatment should be targeted, and the control group contains comparable patients.

In different circumstances, the ATE may be more relevant. This could be the case, for instance, for preventive programs, when the indication for the intervention is going to be expanded. Another example is when the treated sample is somehow different from the population of interest.

It is usually not acknowledged in cost-effectiveness studies that the most widely used methods to adjust for baseline unbalances lead to estimates of the ATE. These are conventional regression methods and covariate adjustment on the propensity score [9,10]. Inverse probability-of-treatment weighting and double robustness also estimate this ATE. Matching methods, in contrast, result in an ATT estimate. In practice, the ATE and ATT may not always be substantially different. Indeed, in the simulations in this thesis, some ATT methods appeared to perform better at estimating ATEs than some ATE methods, and vice versa. However, this must be ascribed to the specific circumstances of this situation. It is not a justification for confusing ATE and ATT methods.

Given the conceptual and potentially practical difference between ATE and ATT, it is important that one set of propensity scores is used in the estimation of both cost and ef-

fects. This prevents that analyses of costs and effects are performed on different samples of patients.

Although no method led to the best results for ATTs and ATEs in all circumstances, some lessons could be drawn from the simulations that were performed. Support was found for the pre-processing approach, which means that a fully specified regression model (step 2) is applied on the data after these have been balanced by using a propensity score method, preferably kernel matching for an ATT and inverse probability-of-treatment weighting (IPTW)) for an ATE (step 1). This approach lessens the demands on the performance of each of the two steps. On the one hand, some residual imbalance in the data after step 1 is not a catastrophe since it can be adjusted in the next step. On the other hand, a misspecification in the regression model is less of a problem if the data has been balanced to a sufficient degree. The preprocessing approach performed more consistently than the double robustness method, which is another technique that is aimed at mitigating the consequences of unsuccessful adjustment in either part of the model. It also led to better results than the direct regression methods: conventional regression and covariate adjustment on the propensity score.

These direct regression methods have more serious disadvantages. Since they do not explicitly lead to an RCT-like design of the dataset, it is not possible to determine whether the adjustment on observed characteristics has been successful. In contrast, after matching or weighting, the covariate balance across the treatment groups can be inspected. Furthermore, the un-biasedness of the results depends on the correctness of the model, in particular the functional form of the link between the propensity score or the other covariates and the health or cost outcome. In our simulations, and in most other studies, this functional form is linear. Of course, other forms can be used and compared, but it is not possible to determine if the adjustment has been sufficient.

11.9 Censoring, non-collapsibility and the analysis of time-to-event data

Non-collapsibility and non-independent censoring have often been lumped together as a source of actual or apparent bias in time-to-event analysis. The analysis in chapter 7 of this thesis showed that only one of them leads to biased estimates: non-independent censoring.

Censoring may occur independently from health outcomes. For instance, it can be due to the loss of paperwork in the mail. Many trials, however, are designed to end before all patients have experienced the event of interest. This means that patients with a better survival prognosis are more likely to be censored: they have not yet died at the end of the trial. When the event is recovery, patients with a worse prognosis are more likely to be censored, i.e. they have not recovered during the trial period.

In the presence of non-independent censoring, relevant prognostic variables must be included in the regression model. Otherwise, projected mean survival for both treatment groups as well as the treatment effect will be underestimated. Furthermore, the hazard ratios in the model will be biased as well. Even in RCT data, in which all measured and unmeasured covariates are expected to be balanced, the omission of prognostic covariates will lead to a biased estimate of the treatment effect. The severity of this bias depends on the proportion of censoring and on the importance of the omitted variables.

In contrast, non-collapsibility does not lead to incorrect estimates. Non-collapsibility occurs when models with different numbers of covariates may lead to different estimates of the hazard ratio for treatment. However, as long as confounding has been addressed successfully, these hazard ratios are all valid albeit different measures of the treatment effect. In other words, the hazard ratios provide answers to different questions: the effect of treatment within strata of the covariates (conditional effect) versus the treatment effect averaged over all strata (marginal effect). A conditional effect is estimated when covariates are added to the marginal model, which contains only treatment. The effects are different when the regression function is non-linear. Marginal and conditional effects are different when the outcome measure is the hazard ratio, but they are equal when the outcome measure is mean time-to-event.

These findings have implications for the analysis of data from RCTs and observational studies. The first implication is that, in principle, time-to-event analysis can be used to provide estimates of the incremental treatment effect when a difference in time is the outcome measure. This estimate is not affected by the model specification once confounding has been eliminated: it is collapsible.

The second implication is that all prognostic covariates should be included in the analysis when the data is censored non-independently. This means that it is impossible to estimate a marginal – or population-averaged – hazard ratio when censoring is present, even in data from an RCT.

The third, and related, implication is that only endpoint models containing all important prognostic factor should be used when propensity-score based methods are applied on censored time-to-event data. Achieving balance of observed baseline covariates by selecting an extensive propensity-score model is not sufficient, since this can only solve the problem of confounding. This is an extra argument for the use of the preprocessing approach, in addition to the arguments in the previous section.

11.10 Adjusting for severity of COPD in database studies

A rich and well-collected dataset from a large clinical trial provided the opportunity to examine whether the severity of COPD, defined by the extent to which a patient suffers from

pulmonary obstruction, could be approached in the absence of lung function measurements. These data are lacking from many registries and administrative databases. However, spirometrically determined disease severity is associated with many health outcomes, especially in the longer term. For this reason, several observational studies have attempted to adjust for possible differences in COPD severity by relying on data on healthcare resource use, medication, sex, age, body-mass index and smoking status.

The study in chapter 10 led to the conclusion that these variables are insufficient to adequately adjust for COPD severity. It is true that an association with lung function was found: more severe COPD was more likely to be found in male patients, with a relatively low age, a lower BMI, who had quit smoking, suffered from osteoporosis, had been hospitalized in the previous year, used oxygen therapy and certain types of respiratory maintenance medication. However, this association was not sufficiently strong as to enable confident classification of patients in categories of disease severity based on the amount of airflow obstruction. Many patients who exhibited the characteristics that were mostly linked to a certain severity stage, were actually in a different stage.

This might not be a problem if the association of COPD severity and the adjusting covariates were consistent across subpopulations, countries and time. In that case, achieving a balanced distribution of covariates across the subpopulations would lead to asymptotic balance on disease severity, as well. In that case, it would still be impossible to determine with some certainty to which individual patients belonged, but wrong categorizations for different stages would cancel out against each other. Unfortunately, the joint distributions of severity stage and other characteristics cannot be observed when spirometric data are unavailable.

11.11 Final remarks: underestimated uncertainties

This discussion has highlighted a number of issues that increase the uncertainty around estimates of incremental costs, effects and cost-effectiveness. These uncertainties are often not fully acknowledged in economic evaluations.

The first uncertainty is about the validness of the estimate of context-specific unit prices for inpatient hospital days. Since an inpatient day may consist of numerous actions and may involve several devices and facilities, its exact contents vary from condition to condition and, crucially, even across patients of different disease severity. ‘The’ inpatient hospital day does not exist, which makes the use of a standard price inappropriate from a societal, healthcare of hospital perspective when the costs of hospital admission are an important driver of costs. However, calculating an appropriate, context-specific price is challenging.

The second uncertainty regards the variation in inpatient hospital costs across patients. When a fixed price is used, this variation – and the stochastic uncertainty around the mean

costs that follows from it – is not noticed in statistical and sensitivity analyses. This means that the actual uncertainty around the estimates of incremental costs and cost-effectiveness ratios is greater than it appears.

The third uncertainty is a result of censored data. Estimates of a treatment effect in time-to-event analysis are only valid in the absence of censoring or when all relevant prognostic covariates are included in the model. However, it is never certain that all of the covariates have been identified.

The fourth uncertainty involves the patient group at which an intervention is targeted. The incremental effects and costs of treatment are not necessarily constant across patients with different characteristics. In observational studies, this is reflected in the difference between ATT and ATE. In RCTs, it is often less obvious. If the trial population is different from the real-life patient population, the real-life effectiveness and cost-effectiveness will be different from the trial estimates. Although this is widely acknowledged in principle, cost-effectiveness analyses are often piggy-backed on RCTs.

The fifth uncertainty concerns the adjustment approach in observational studies. In most or all circumstances, the analyst has to choose between several options. Some approaches are better than others, on average, but all methods occasionally lead to very incorrect estimates. It cannot always be concluded from the results whether this has happened. Nevertheless, this should not be understood as a plea against observational studies, which can result in valuable information.

Some of these uncertainties can be reduced, some can be explored, and some remain unknown. Observational studies can help in estimating real-life treatment effects when this is not possible in an RCT. In that respect, observational studies are not second best, compared to the 'gold standard' of RCTs. In the GO AHEAD trial, however, it was possible to randomise a sample of patients that resembled the patients for whom the treatment would be appropriate in real life. Furthermore, the intervention was designed to follow daily practice as closely as possible.

When there are substantial differences between an RCT protocol and daily practice, univariate sensitivity analyses can be used to explore the results of different assumptions on the effectiveness of the treatments under comparison. The same approach can be used to investigate the impact of a different unit price. Univariate sensitivity analysis is a well-established technique, which is often overshadowed by probabilistic sensitivity analysis. Nevertheless, it can be very useful, especially when the uncertainty is of a structural and not merely a stochastic character.

Reducing the amount of censoring could reduce the uncertainty about the validity of the estimated treatment effect as well as the stochastic uncertainty. However, as it is the case when sample sizes are determined, ethical and financial considerations lead to a pressure to terminate a trial as early as possible in order to limit costs and the exposure of patients to an inferior treatment.

The uncertainty about the validity of estimates from different methods in observational studies can be reduced by diligently checking the balance of covariates after applying a matching or weighting approach. The remaining uncertainty can be explored by using several approaches and comparing the results. If these results are not very dissimilar, this could enhance confidence in their correctness. However, they may also be quite different and lead to different conclusions. In that case, all that is certain is that the real treatment effect remains uncertain.

11.12 References

1 de Meijer C, Brouwer W, Koopmanschap M, et al. The value of informal care--a further investigation of the feasibility of contingent valuation in informal caregivers. Health Econ. 2010; 19: 755-771.

2 Koopmanschap MA, van Exel JN, van den Berg B, Brouwer WB. An overview of methods and applications to value informal care in economic evaluations of healthcare. Pharmacoeconomics. 2008; 26: 269-280.

3 Blinman P, King M, Norman R, et al. Preferences for cancer treatments: an overview of methods and applications in oncology. Ann Oncol. 2012; 23: 1104-1110.

4 Mullen PM. Public involvement in health care priority setting: an overview of methods for eliciting values. Health Expect. 1999; 2: 222-234.

5 Shackley P, Ryan M. Involving consumers in health care decision making. Health Care Anal. 1995; 3: 196-204.

6 Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ. 2000; 320: 1530-1533.

7 Louviere JL, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications . Cambridge: Cambridge University Press, 2000.

8 de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012; 21: 145-172.

9 Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006; 59: 437-447.

10 Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol. 2005; 58: 550-559.

Summary
Samenvatting

S

Summary

Many economic evaluations are based on or conducted in combination with medical studies, in particular phase III randomised controlled trials, also in COPD. However, an economic evaluation is more than a medical study with costs as an additional outcome measure. It is more complicated than combining the effectiveness results from a trial with the invoices from physicians, hospitals and pharmacies. Economic evaluations are often characterized by challenges that do not occur in purely clinical trials.

Firstly, in many cases costs cannot be observed directly. From a societal or healthcare perspective, they do not equal payments or tariffs. Secondly, the objective of many, although not all, clinical trials are to find statistical proof that one intervention's effectiveness is superior to another one. In an economic evaluation the *size* of the treatment effect is essential, not merely its existence. In order to estimate this size with sufficient external validity, an observational study can be conducted on patients in real life. However, this study design lacks the benefit of randomization. Adjustment for possible differences between treatment groups is required. Thirdly, the size of the treatment effect must be estimated in terms of natural units of health in order to be useful in an economic evaluation. However, the results of medical and epidemiological studies are often expressed in hazard ratios – in time-to-event analyses – or odds ratios – for binary count data. These cannot be used in economic evaluations, in which outcome measures such as survival time and numbers of successes and failures are required instead. Fourthly, in order to make the results of economic evaluations comparable across interventions and diseases, health gains must be expressed in a common outcome measure, such as the QALY. These challenges were central to this thesis.

The first part of the thesis focused on hospital-at-home, in particular but not restricted to COPD. Hospital-at-home aims to avoid admission, or reduce length of stay (early assisted discharge schemes) by delivering formal healthcare in the patient's home.

In chapter 2, a systematic review of costs studies for hospital-at-home was performed. We identified 29 studies for acute conditions that were published from 1996 through 2011. There is a risk that cost savings due to shortening hospital admissions are overestimated. Cost calculations were considered incorrect if they failed to meet four criteria. Costs per inpatient hospital day had to be disease-specific. The decreasing intensity of care over the course of an admission had to be reflected in costs of inpatient days, since the last days in the hospital tend to be the less costly than an average day. In studies from the societal perspective, informal care costs had to be included. Violating any of these three criteria leads to overestimation of savings. Finally, follow-up had to be at least one month in order to capture relevant downstream costs, such as re-admissions. Only five studies met all criteria. The most frequent problems were the use of average costs per inpatient hospital day and too short follow-up. This leads to the conclusion that, while most studies found that

hospital-at-home was cost-saving, these savings were probably overestimated. Savings may not disappear after adjusted calculations, but they will be considerably smaller.

Chapter 3, 4 and 5 described a randomized controlled trial of early assisted discharge compared to usual hospital treatment for patients with a COPD exacerbation in the Netherlands. In this GO AHEAD trial (Assessment of GOing Home under Early Assisted Discharge), participants stayed in the hospital for seven days, or went home after three days where they were supervised and treated by community nurses. The protocol was described in chapter 3.

In chapter 4, we found that transferring hospital care for a COPD exacerbation to the patient's home was likely to lead to modest costs savings from a healthcare perspective, while there was no clear evidence to conclude that it would lead to a slower or worse resolution of the exacerbation. From a societal perspective – when the societal costs of informal care and productivity losses were taken into account – these cost savings decreased considerably or even turned into increases.

The first day of the admission was the most costly. After that, the intensity of care by physicians and nurses decreased, which was reflected in lower costs per day.

The preferences of patients and informal caregivers for early assisted discharge and usual hospital treatment were examined in chapter 5 by means of a discrete choice experiment. The results showed that the average patient and informal caregiver do not exist. For both groups, four distinct classes were identified, which had different attitudes towards being treated at home or in the hospital. Large proportions of respondents had a preference for either treatment option that could not be influenced by changes in the characteristics of the early assisted discharge treatment. The most attractive hospital-at-home program would be operated by pulmonary nurses and have no co-payments and a light burden on informal caregivers. Patients would not be visited by more than two different nurses and the pulmonary ward of the hospital would be reachable 24-hours a day in case of sudden worsening of the disease.

The second part of this thesis was devoted to three additional methodological issues: quality-of-life measurement, natural units of health gain, and observational studies.

Chapter 6 demonstrated that the EQ-5D questionnaire can be used to measure health-related quality of life during the recovery from a moderate COPD exacerbation. When the exacerbation was at its worst, the average EQ-5D utility score was 0.651. It increased to 0.768 on day 14, after which it remained largely stable until the final visit at 6 weeks. The greatest improvement in utility scores was reached within two weeks after the onset of the exacerbation. The EQ-5D was more responsive than expiratory peak flow, rescue medication and sputum symptom scores and equally responsive as cough and dyspnoea symptoms scores. All five dimensions of the EQ-5D contributed significantly to the mean improvement in the EQ-5D score.

In chapter 7, the concepts of censoring and non-collapsibility in time-to-event analysis were disentangled and clarified by analyzing synthesized data. Earlier studies have some-

times confused these phenomena and misrepresented the problems that non-collapsibility and censoring can induce. Non-collapsibility exists when the estimated treatment effect changes as prognostic covariates are added to the regression model, even when confounding is absent. All estimates may be valid. In contrast, censoring threatens the validity of estimates when prognostic factors are omitted from the regression model. Hazard ratios are non-collapsible, but estimated time-to-event is collapsible.

Chapters 8 and 9 investigated the validity, accuracy and precision of conventional and propensity-score-based adjustment methods in observational cost-effectiveness analysis. The PS-methods were: PS matching (kernel and one-to-one), covariate adjustment using PS, inverse probability-of-treatment weighting (IPTW) and double robustness, each with several specifications. The propensity score is the conditional probability of receiving the treatment under study, given a patients characteristics.

In chapter 8, conventional regression was less likely to lead to accurate and unbiased estimates than propensity-score methods. On the other hand, in chapter 9, it performed well with regards to precision. The pre-processing approach, in which a fully specified regression model is applied after a matching or weighting on the propensity score, combined accuracy with relative precision. Covariate adjustment using the propensity score performed badly on both counts.

In chapter 10 it was shown that it is not possible to reliably adjust for the severity of COPD in observational studies in the absence of data on lung function. Almost all classifications of COPD severity are at least partly based on lung function, but this information is not available in administrative and clinical databases and registries. Previous retrospective database research has approximated COPD severity using demographic and healthcare resource utilization data. However, our partial proportional odds model analysis made clear that lung function cannot be reliably predicted from these.

The final chapter highlighted six uncertainties that are generally underestimated in the economic evaluation of healthcare. The first uncertainty is about the validity of the estimate of context-specific unit prices for inpatient hospital days. Since an inpatient day may consist of numerous actions and may involve several devices and facilities, its exact contents vary from condition to condition and even across patients of different disease severity. The second uncertainty regards the variation in inpatient hospital costs across patients. When a fixed price is used, this variation – and the stochastic uncertainty around the mean costs that follows from it – is not noticed in statistical and probabilistic sensitivity analyses. The third uncertainty is a result of censored data. Estimates of a treatment effect in time-to-event analysis are only valid in the absence of censoring or when all relevant prognostic covariates are included in the model. The fourth uncertainty involves the patient group at which an intervention is targeted. Real-life effectiveness and cost-effectiveness may be different from the trial estimates. The fifth uncertainty concerns the adjustment approach in observational

studies. Some approaches are better than others, on average, but all methods occasionally lead to very incorrect estimates.

Samenvatting

Veel economische evaluaties zijn gebaseerd op of worden uitgevoerd in combinatie met klinische studies, in het bijzonder gerandomiseerde onderzoeken met een controlegroep. Toch is een economische evaluatie is meer dan de combinatie van een klinische studie met de rekeningen van artsen, ziekenhuizen en apotheken. Het onderzoek wordt dikwijls gekenmerkt door problemen die niet optreden in puur medische studies.

Ten eerste kunnen kosten vaak niet direct worden waargenomen. Vanuit het maatschappelijk of gezondheidszorgperspectief komen kosten niet per se overeen met betalingen of tarieven. Ten tweede is het doel van de veel klinische studies het statistisch aantonen dat een interventie beter werkt dan een andere. In een economische evaluatie is de *omvang* van dit behandeleffect essentieel, niet alleen het bestaan ervan. Een extern valide schatting van deze omvang kan worden gemaakt met een observationele studie. Deze studieopzet heeft echt niet het voordeel van randomisatie. Correcties voor mogelijke verschillen tussen de behandelgroepen zijn daardoor noodzakelijk. Ten derde kan de schatting van het behandeleffect alleen worden toegepast in een economische evaluatie als het ze wordt uitgedrukt in eenheden van gezondheidswinst. De resultaten van klinische en epidemiologische studies worden echter vaak uitgedrukt in *hazard ratios* – in duurmodellen – of *odds ratios* – voor dichotome uitkomsten. In plaats daarvan vereist economische evaluatie uitkomstmaten als levensverwachting of aantallen geslaagde behandelingen. Ten vierde moet gezondheidswinst kunnen worden uitgedrukt in een uitkomstmaat die de resultaten van verschillende interventies en voor verschillende aandoeningen vergelijkbaar maakt, zoals de QALY. Deze kwesties stonden centraal in dit proefschrift.

Het eerste deel van het proefschrift ging over ziekenhuisverplaatste zorg, in het bijzonder voor patiënten met een COPD-exacerbatie. Het doel van ziekenhuisverplaatste zorg is het voorkomen van ziekenhuisopnames of het verkorten van de opnameduur (vervroegd en begeleid ontslag) door formele zorg te leveren bij de patiënt thuis.

In hoofdstuk 2 werd een systematisch literatuuronderzoek van kostenstudies over ziekenhuisverplaatste zorg uitgevoerd. Er is een risico dat kostenbesparingen door dit concept worden overschat. We vonden 29 studies die werden gepubliceerd van 1996 tot 2011. De berekeningen werden als incorrect beschouwd als ze niet voldeden aan een van de volgende vier criteria. De kostprijzen van verpleegdagen moesten zijn toegespitst op de specifieke ziekte van de studie.

De afnemende intensiteit van de zorg gedurende de opname moest tot uiting komen in de kostprijzen, aangezien de laatste dagen van een verblijf in het ziekenhuis vaak minder duur zijn dan de gemiddelde dag. In studies met een maatschappelijk perspectief moesten de kosten van mantelzorg worden meegerekend. Het schenden van een van deze drie criteria leidt tot een overschatting van de kostenbesparingen. Ten vierde moesten de kosten, bijvoorbeeld die van heropnames, zijn geregistreerd tot ten minste een maand na de behandeling.

Slechts vijf studies voldeden aan alle criteria. De meest voorkomende problemen waren het gebruik van gemiddelde kosten per verpleegdag en een te korte periode van follow-up. Dit leidde tot de conclusie dat de besparingen die in de meeste studies worden waargenomen, waarschijnlijk zijn overschat. De besparingen zouden in een juiste berekening niet noodzakelijkerwijs verdwijnen, maar wel aanzienlijk kleiner worden.

De hoofdstukken 3, 4 en 5 gingen over een gerandomiseerde studie met controlegroep, waarin vervroegd en begeleid ontslag werd vergeleken met de reguliere ziekenhuisopname voor patiënten met een COPD-exacerbatie. In deze GO AHEAD-studie (Assessment of GOing Home under Early Assisted Discharge) verbleven patiënten ofwel zeven dagen in het ziekenhuis of gingen ze na drie dagen naar huis, waar ze verder warden behandeld en gecontroleerd door thuiszorgverpleegkundigen. Het studieprotocol staat in hoofdstuk 3.

In hoofdstuk 4 werd vastgesteld dat het vervangen van een deel van de ziekenhuisopname door zorg bij de patiënt thuis waarschijnlijk leidt tot bescheiden kostenbesparing binnen de gezondheidszorg. Er waren geen duidelijke aanwijzingen dat het zou leiden tot een langzamer of minder goed herstel van de exacerbatie. Vanuit maatschappelijk perspectief – wanneer ook de kosten van mantelzorg en productiviteitsverliezen in aanmerking worden genomen – waren de besparingen aanzienlijk kleiner of veranderden ze in meerkosten.

De eerste dag van de ziekenhuisopname was de meest kostbare. Daarna werd de intensiteit van de zorg door artsen en verpleegkundigen minder, wat neersloeg in een lagere kostprijs per dag.

De voorkeuren van patiënten en mantelzorgers voor vervroegd en begeleid ontslag en voor de reguliere ziekenhuisopname werden onderzocht in een discrete-keuze-experiment in hoofdstuk 5. Uit de resultaten bleek dat de gemiddelde patiënt en mantelzorger niet bestaan. Beide groepen konden worden verdeeld in vier klassen, elk met een andere houding tegenover behandelingen thuis en in het ziekenhuis. Een groot deel van de respondenten had een onwrikbare voorkeur voor een van beide mogelijkheden. Veranderingen in de aangeboden behandeling met vervroegd en begeleid ontslag hadden daarop geen invloed. De meest aantrekkelijke thuisbehandeling zou worden uitgevoerd door gespecialiseerde longverpleegkundigen, zou geen eigen bijdrage vergen en zou geen zware last voor de mantelzorgers opleveren. Een patiënt zou door ten hoogste twee verschillende verpleegkundigen worden bezocht en de longafdeling van het ziekenhuis zou 24 uur per dag bereikbaar moeten zijn voor overleg over een eventuele onvoorziene verslechtering van de toestand van de patiënt.

Het tweede gedeelte van het proefschrift is gewijd aan drie andere methodologische onderwerpen: de meting van kwaliteit van leven, eenheden van gezondheidswinst en de analyse van gegevens uit observationele studies.

In hoofdstuk 6 werd aangetoond dat de EQ-5D-vragenlijst kan worden gebruikt om de kwaliteit van leven van een patiënt te meten tijdens het herstel van een matige COPD-exacerbatie. Op het dieptepunt van de exacerbatie, was de gemiddelde EQ-5D-nutsscore

0,651. Op dag 14 was de het nut opgelopen tot gemiddeld 0,786, waarna het stabiel de volgende zes werken in grote mate stabiel bleef. De belangrijkste verbetering in nut werd bereikt binnen twee weken na het begin van de exacerbatie. De EQ-5D reageerde sterker op het herstel van de exacerbatie dan metingen van de *peak expiratory flow*, opgehoest slijm en het gebruik van noodmedicatie. De symptomen hoesten en ademnood reageerden even sterk als de EQ-5D. Alle dimensies van de EQ-5D droegen significant bij aan de verbetering van de gemiddelde EQ-5D score.

In hoofdstuk 7 zijn de concepten *censoring* en *non-collapsibility* in overlevingsmodellen van elkaar onderscheiden en verduidelijkt aan de hand van een analyse van gesynthetiseerde gegevens. Eerder studies hebben deze fenomenen soms verward. Dat leidde tot een verkeerde voorstelling van de problemen waartoe *censoring* en *non-collapsibility* kunnen leiden. *Non-collapsibility* treedt op wanneer een schatting van het behandeleffect verandert met de toevoeging van prognostische covariaten aan het regressie model, ook wanneer er geen *confounding* is. Al deze schattingen kunnen valide zijn. *Censoring* vormt daarentegen wel degelijk een bedreiging voor de validiteit van effectschattingen wanneer belangrijke prognostische factoren worden weggelaten uit het regressiemodel. *Hazard ratios* zijn *non-collapsible*, maar de gemiddelde voorspelde overlevingsduur is wel *collapsible*.

In de hoofdstukken 8 en 9 werden de validiteit, accuratesse en precisie van verscheidene statistische analysemethoden onderzocht in observationeel kosteneffectiviteitsonderzoek: conventionele methoden en van technieken die zijn gebaseerd op *propensity scores*. Deze laatste technieken waren: *propensity score matching* (met *kernel* en 1-op-1), regressie met de *propensity score* als covariaat, *inverse probability-of- treatment weighting* (IPTW) en *double robustness*, elk met verschillende specificaties. De *propensity score* van een patiënt is de berekende kans dat iemand met dezelfde kenmerken als deze patiënt de onderzochte behandeling krijgt.

In hoofdstuk 8 bleek dat conventionele regressiemethoden minder vaak tot een accurate en niet-vertekende schatting leidden dan de methoden gebaseerd op *propensity scores*. Daartegenover staat dat conventionele regressie in hoofdstuk 9 wel goed presteerde op het gebied van precisie. De voorbehandelingsbenadering, waarbij een volledig regressiemodel wordt toegepast nadat eerst matching of weging heeft plaatsgevonden, combineerde accuratesse met relatieve precisie. Regressie met de *propensity score* als covariaat presteerde slecht op beide criteria.

In hoofdstuk 10 werd aangetoond dat het in observationele studies niet mogelijk is correcties uit te voeren voor verschillen in ernst van COPD wanneer er geen gegevens over longfunctie beschikbaar zijn. Vrijwel alle classificatiesystemen voor de ernst van deze ziekte zijn ten minste gedeeltelijk gebaseerd op longfunctiemetingen, maar deze informatie is vaak niet terug te vinden opgenomen in administratieve en klinische bestanden en registers. In eerdere onderzoeken van bestaande bestanden is geprobeerd de ernst te benaderen met behulp van demografische gegevens en informatie over het zorggebruik van de patiënt. De

analyse van ons *partial proportional odds model* maakte echter duidelijk dat longfunctie daarmee niet betrouwbaar kan worden voorspeld.

In het slothoofdstuk werden zes onzekerheden gemarkeerd, die over het algemeen onderschat worden in economische evaluaties in de gezondheidszorg.

De eerste onzekerheid gaat over de validiteit van schattingen van de context-specifieke prijzen voor verpleegdagen in het ziekenhuis. Een verpleegdag kan bestaan uit talloze activiteiten en kan het gebruik van verschillende apparaten en faciliteiten met zich mee brengen. De exacte inhoud van zo'n dag verschilt dan ook van aandoening tot aandoening en zelfs tussen patiënten met een ernstiger of minder ernstige variant van dezelfde aandoening. De tweede onzekerheid heft betrekking op de variatie in verpleegdagkosten tussen patiënten. Wanneer een vast prijs wordt gebruikt, wordt deze variatie niet opgemerkt in statistische analyses en probabilistische gevoeligheidsanalyses. Hetzelfde geldt voor de stochastische onzekerheid over de gemiddelde totale kosten die het gevolg is van deze variatie. De derde onzekerheid is een resultaat van *censoring* van de gegevens. Schattingen van een behandeleffect in overlevingsanalyses zijn alleen valide wanneer er geen censoring optreedt of wanneer alle relevante prognostische covariaten zijn opgenomen in het model. De vierde onzekerheid gaat om de groep patiënten waarop een interventie wordt gericht. Effectiviteit en kosteneffectiviteit kunnen in de praktijk anders zijn dan in klinische studies. De vijfde onzekerheid betreft de correctiemethode in observationele studies. Sommige methoden zijn beter dan andere, maar alle methoden leiden in sommige gevallen tot zeer incorrecte schattingen.

D

Dankwoord

PhD Portfolio

Curriculum vitae

Dankwoord

De voorbereidingen voor dit proefschrift begonnen in mei 2003, al wist ik dat toen nog niet. Ik was bij een voorlichtingsavond over de nieuwe masterstudie Health Economics, Policy and Law aan de Erasmus Universiteit. Eigenlijk had ik op dat moment geen enkele goede reden om die opleiding te gaan volgen – ik had al een beroep en een baan – maar het aanstekelijke verhaal van Wynand van de Ven als opleidingscoördinator trok me over de streep. Ik heb er geen spijt van dat ik me na die avond heb ingeschreven voor de studie.

René van Vliet en Johan Polder, tijdens het schrijven van mijn masterscriptie onder jullie begeleiding kreeg ik voor het eerst idee dat dit werk mijn beroep zou kunnen worden. Werner Brouwer, bedankt dat je me eind 2005 belde met de vraag of ik misschien geïnteresseerd zou zijn in een eventuele functie als junior onderzoeker bij het iMTA. Daarover hoefde ik maar een paar seconden over na te denken.

De leden van de kleine en grote promotiecommissie dank ik hartelijk voor het beoordelen van mijn proefschrift en de oppositie tijdens de verdediging ervan. De coauteurs van de artikelen waarin ik meewerkte, binnen en buiten dit proefschrift, hebben een belangrijke bijdrage geleverd.

Brigitta Monz, I really enjoyed working with you on studies commissioned by Boehringer Ingelheim, two of which were included in this thesis.

Cecile, jarenlang hebben we samengewerkt in de GO AHEAD-studie. Dankjewel voor al het werk dat je hebt verzet bij het voorbereiden en coördineren van de studie en het verzamelen van alle gegevens bij de deelnemende ziekenhuizen en thuiszorgorganisaties. Onze discussies over statistische analyses, formuleringen en conclusies in onze artikelen waren soms heftig, meestal leuk, en altijd interessant. Ook de artsen en verpleegkundigen die bij de studie betrokken waren, ben ik veel dank verschuldigd, met name Walter van Litsenburg, Frank Smeenk, Wiel Seezink en Monique van Vliet. En natuurlijk aan de deelnemende patiënten en hun mantelzorgers.

Maureen, ik denk niet dat er een betere promotor bestaat dan jij. Je hebt een opgewekt soort discipline, eist veel van jezelf en bent als vanzelfsprekend vastbesloten ook uit anderen het beste te halen. Dat is uitdagend en inspirerend. Het duurde een hele tijd tot ik je eindelijk eens iets kon vertellen wat je zelf niet allang veel beter wist. Bedankt voor het delen van je kennis en kunde, je warme belangstelling en je vertrouwen.

Onno, weinig wetenschappers zijn zo druk bezet als jij. Het lijkt wel of heel de wereld je weet te vinden. Ik ben dan ook erg blij dat je tijd vrijmaakte om mijn tweede promotor te zijn. De manier waarop je commentaar gaf op mijn teksten was bijzonder motiverend.

Ken, toen ik je vroeg mijn co-promotor te zijn, antwoordde je meteen: “Dat zou ik een grote eer vinden.” Die reactie is typisch voor jou, maar het is natuurlijk precies andersom. Bedankt voor alles. En voortaan denk ik bij elk artikel dat ik schrijf: “What’s the take-home message?”

Apostolos, Steef en Maiwenn, ik heb jullie vaak lastiggevallen als ik iets schijnbaar onmogelijks wilde doen, of juist als ik op zoek was naar iets uit de cursus statistiek voor beginners. Ik kwam nooit voor niets, zelfs als jullie dachten dat we geen stap verder waren gekomen. Iets voorleggen aan iemand die verstandige dingen terugzegt, dat helpt echt.

Marieke, mijn kamergenote aan de zonnige kant van het J-gebouw. We tonen elke dag wetenschappelijk aan dat het mogelijk is: onophoudelijk praten en ondertussen gewoon doorwerken. Je begon een maand eerder dan ik bij de universiteit en je proefschrift was ook een maand eerder af, zodat we ons lekker konden bemoeien met elkaars lettertypes, kaft en dankwoord. Het is een plezier om met jou een kamer te delen.

Chantal en Pepijn, het lag voor de hand dat ik jullie als mijn paranimfen zou vragen.

Pepijn, wat was het leuk om samen met jou door ondoorgrondelijke rekenmodellen heen te worstelen, op internet nieuwe huizen te bekijken in de buurt van Groningen, te stappen in Praag en te zwemmen in Madrid. Het is jammer dat je naar de andere kant van het land bent verhuisd.

Chantal, zonder jou zouden drie van de hoofdstukken in dit proefschrift niet zijn geschreven. Vier jaar geleden bedachten we dat we samen 'iets' wilden doen, met propensity scores of zo. We wisten een onderzoekssubsidie voor het project in de wacht te slepen en hebben er avonden en avonden aan doorgewerkt, samen aan één toetsenbord of via de telefoon. Ik ben nog steeds uitermate boos over je vertrek bij de universiteit...

Alle andere vrienden en collega's bij de sectie GE-iMTA, jullie maken het werk nog leuker dan het anders al zou zijn, en niet alleen op vrijdag.

PhD portfolio

Training

Repeated measurements in clinical studies (Emmanuel Lesaffre, Erasmus Medical Centre), Netherlands Institute for Health Sciences, 2011

Missing values in clinical research (Geert Molenberghs, Catholic University of Leuven), Netherlands Institute for Health Sciences, 2010

Causal inference, (Miguel Hernàn, Harvard University), Erasmus Summer Programme, Netherlands Institute for Health Sciences, Rotterdam, 2008

Discrete choice analysis and choice experiment design (John Rose & Michiel Bliemer, University of Sydney), Executive course, Erasmus University Rotterdam, 2007

iBMG didactic courses, Erasmus University Rotterdam, 2012

Probleemgestuurd onderwijs [Problem-oriented teaching], Erasmus University, 2009

Basiscursus didactiek [Basic didactics], Erasmus University Rotterdam, 2008

Academic writing in English, Erasmus University Rotterdam, 2008

Teaching

Kwantitatief gezondheidszorgonderzoek (Methoden en technieken van onderzoek 4) [Quantitative research in healthcare (Research methodology 4)], bachelor program Health Sciences, Policy and Management: lecture on causal structures and model selection; supervision of student research projects, 2006-present

Supervision and co-supervision of several theses, bachelor program Health Sciences, Policy and Management and master Health Economics, Policy and Law, 2008-present

Health technology assessment, master program Health Economics, Policy and Law: supervision of modelling practicums, 2007-present

Advanced economic evaluation, master program Health Economics, Policy and Law: supervision of modelling practicums, 2007-present

Presentations

Cost-effectiveness of early assisted discharge for COPD exacerbations in the Netherlands, Longdagen Utrecht, 2012

Cost-effectiveness of early assisted discharge for COPD exacerbations in the Netherlands, GE-iMTA lunch lecture, Rotterdam, 2012

Collapsibility and censoring: what's the bias in estimating survival time? GE-iMTA lunch lecture, 2012

Measuring health by prediction, memory or experience? An experience sampling study, Low Lands Health Economist Study Group, Almen, 2012

A propensity to get it right. A comparison of different adjustment methods for obtaining correct cost-effectiveness estimates in observational studies: a simulation study, Low Lands Health Economist Study Group, Almen, 2012

Research posters

A propensity to get it right. Comparing statistical methods for observational cost-effectiveness studies. ISPOR 15th annual European Congress, Berlin, 2012 (Poster award finalist)

Non-collapsibility and censoring: What's the bias in estimated survival time? ISPOR 15th annual European Congress, Berlin, 2012 (Poster award finalist)

Hospital-at-home: cost savings are overestimated. ISPOR 15th annual European Congress, Berlin, 2012

Latent class analysis of discrete choice data: an application on hospital-at-home for COPD patients. ISPOR 14th annual European Congress, Madrid, 2011

Is the EQ-5D responsive to recovery from a moderate COPD exacerbation? ISPOR 11th annual European Congress, Athens, 2008

No substitute for spirometry: adjusting for COPD severity in database research. ISPOR 13th annual European Congress, Prague, 2010 (Poster award finalist)

Publications in English

Rutten-van Mölken MPMH, Goossens LMA. Cost effectiveness of pharmacological maintenance treatment for chronic obstructive pulmonary disease: a review of the evidence and methodological issues. Pharmacoeconomics. 2012; 30(4):271-302.

Utens CMA, Goossens LMA, Smeenk FW, Rutten-van Mölken MPMH, van Vliet M, Braken MW, van Eijsden LM, van Schayck OC. Early assisted discharge with generic community nursing for chronic obstructive pulmonary disease exacerbations: results of a randomised controlled trial. BMJ Open. 2012 Oct 16;2(5). doi:pii: e001684.

Goossens LMA, Baker CL, Monz BU, Zou KH, Rutten-van Mölken MPMH. Adjusting for COPD severity in database research: developing and validating an algorithm. Int J Chron Obstruct Pulmon Dis. 2011; 6:669-78.

Goossens LMA, Nivens MC, Sachs P, Monz BU, Rutten-van Mölken MPMH. Is the EQ-5D responsive to recovery from a moderate COPD exacerbation? Respir Med. 2011;105(8):1195-202.

Utens CM, Goossens LM, Smeenk FW, van Schayck OC, van Litsenburg W, Janssen A, van Vliet M, Seezink W, Demunck DR, van de Pas B, de Bruijn PJ, van der Pouw A, Retera JM, de Laat-Bierings P, van Eijsden L, Braken M, Eijsermans R, Rutten-van Mölken MP. Effectiveness and cost-effectiveness of early assisted discharge for chronic obstructive pulmonary disease exacerbations: the design of a randomised controlled trial. BMC Public Health. 2010;10:618.

Goossens LMA, Standaert B, Hartwig N, Hövels AM, Al MJ. Conclusion on cost-effectiveness of rotavirus vaccination highly dependent on assumptions. Vaccine. 2009 Apr 28; 27(19):2531-2.

Goossens LMA, Riemersma RA, Postma DS, van der Molen T, Rutten-van Mölken MPMH. An economic evaluation of budesonide/formoterol for maintenance and reliever treatment in asthma in general practice. Adv Ther. 2009; 26(9):872-85.

Goossens LMA, Standaert B, Hartwig N, Hövels AM, Al MJ. The cost-utility of rotavirus vaccination with Rotarix (RIX4414) in the Netherlands. Vaccine. 2008; 26(8):1118-27.

Publications in Dutch

Hakkaart-Van Roijen L, Tan SS, Goossens LMA, Schawo S, Brouwer WBF, Rutten-van Mölken MPMH, Rutten FFH. Doelmatigheid in praktijkrichtlijnen voor medicijnen: resultaten van een 'quickscan' [Efficiency in practice guidelines for medication: results of a quickscan]. TSG, Tijdschrift voor Gezondheidswetenschappen 2010; 88(4): 172-181 (in Dutch)

Goossens LMA, Knoester P, Rutten-van Mölken MPMH Mölken. Avastin (bevacizumab), Macugen (pegaptanib) en Lucentis (ranibizumab): vergelijking van medicatiekosten [Avastin (bevacizumab), Macugen (pegaptanib) en Lucentis (ranibizumab): a comparison of medication costs]. Pharmaceutisch Weekblad Wetenschappelijk Platform. 2008: 2(4):75-79. (in Dutch)

Goossens LMA, Niessen LW. Rosuvastatine, atorvastatine en de richtlijn: kosteneffectiviteit in de secundaire preventie [Rosuvastatin, atorvastatin and Dutch guidelines: cost-effectiveness in secondary prevention]. Pharmaceutisch Weekblad Wetenschappelijk Platform. 2008; 2(3):56-61. (in Dutch)

Curriculum vitae

Lucas Goossens was born in Oss on the 18th of November, 1973. He grew up in Uden, where he graduated from VWO at Kruisheren Kollege in 1991.

He studied history at Radboud University Nijmegen (MA in 1995), where he specialised in economic and social history, American history and the economics of public finance. His master thesis described the rise of the supermarket in the Netherlands and its political fall-out. In 1996, he was admitted to the Post-doctorale Opleiding Journalistiek (PDOJ) of Erasmus University.

This led to a career in newspaper journalism at the national daily Algemeen Dagblad. From 1997 to 2003, he reported on national politics from The Hague. After this, he selected and edited news stories and composed headlines as an editor of domestic news at the central desk in Rotterdam for three years.

Meanwhile, he studied Health Economics, Policy and Law at Erasmus University (MSc cum laude in 2005), from which he graduated with a thesis on the consequences of population ageing and falling mortality for health care expenditure.

Since 2006, he has worked at the Erasmus University Institute for Medical Technology Assessment/Institute of Health Policy and Management (iMTA/iBMG). His research interests include the cost-effectiveness of treatment for pulmonary diseases, statistical and epidemiological methods and observational studies.