Real-world cost-effectiveness

POTENTIAL & PITFALLS IN THE CONTEXT OF CONDITIONAL REIMBURSEMENT

CHANTAL VAN GILS

Real-world cost-effectiveness

Potential & pitfalls in the context of conditional reimbursement

Chantal van Gils

Funding
The studies performed in this thesis were supported by the Dutch Healthcare Insurance Board (CVZ) and the Netherlands Organization for Health Research and Development (ZonMW), grant number 152002017.

Publication of this thesis was financially supported by: GlaxoSmithKline, Roche

van Gils, C.W.M.
Real-world cost-effectiveness. Potential & pitfalls in the context of conditional reimbursement
Dissertation Erasmus University Rotterdam, The Netherlands

ISBN: 978-94-6169-453-9

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

REAL-WORLD COST-EFFECTIVENESS

Potential & pitfalls in the context of conditional reimbursement

KOSTENEFFECTIVITEIT GEBASEERD OP DE DAGELIJKSE PRAKTIJK

Potenties en valkuilen in de context van voorwaardelijke vergoedingen

Proefschrift

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

De openbare verdediging zal plaatsvinden op
Donderdag 19 december 2013 om 13:30 uur

door
Chantal Wilhelmina Martina van Gils
geboren te Raamsdonk

PROMOTIECOMMISSIE

Promotoren
Prof.dr. C.A. Uyl-de Groot
Prof.dr. C.J.A. Punt

Overige leden
Prof.dr. M.P.M.H. Rutten-van Mölken
Prof.dr. F.F.H Rutten
Prof.dr. L. Annemans

Copromotor
Dr. W.K. Redekop

CONTENTS

1.	General introduction	7
2.	Adjuvant chemotherapy in stage III colon cancer: Guideline implementation, patterns of use and outcomes in daily practice in The Netherlands	19
3.	Real-world resource use and costs of adjuvant treatment for stage III colon cancer	33
4.	Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: A synthesis of clinical trial and daily practice evidence	55
5.	Practical feasibility of outcomes research in oncology: Lessons learned in assessing drug use and cost-effectiveness in The Netherlands	87
6.	A propensity to get it right: A Monte Carlo simulation study comparing statistical methods to obtain correct cost-effectiveness estimates in observational studies	103
7.	A propensity to get it precise: A simulation study comparing statistical methods to obtain reliable cost-effectiveness estimates in observational studies	131
8.	General discussion	147
	References	167
	Summary	185
	Samenvatting	195
	About the author	205
	List of publications	207
	Dankwoord	213

Chapter 1

General introduction

INTRODUCTION

The development of new and expensive health care technologies has increased pressure on national health care budgets as well as hospital budgets, leading to difficult questions about the affordability of new medicines. In order to strike an optimal balance between ensuring timely access to new drugs and having sufficient evidence of their relative benefits and risks, health care reimbursement policies have been increasingly rationalised and formalised over the past decades. The Netherlands is no exception in this context. Already since the early 1990s, the Dutch government has been developing policies linking reimbursement to evidence requirements aiming to reduce the growth rate of pharmaceutical costs without loss of medical quality. Over the years these evidence requirements as well the methods to demonstrate this evidence have greatly evolved. However, this progress has also led to greater complexity in reimbursement policy.

In 1991, the first drug policy was introduced, the Dutch Price Reference System (in Dutch: geneesmiddelenvergoedingssysteem, GVS). Since the introduction of the GVS new outpatient drugs are only considered for reimbursement if their therapeutic value and efficiency are assessed. Therapeutic value is elaborated into efficacy, safety, experience, ease of use, quality of life, and applicability (e.g. interactions, contra-indications); the efficiency assessment was based on the outpatient drug budget impact only.[1] Since 2005 another efficiency measurement must also be provided when applying for reimbursement of a new drug: a cost-effectiveness analysis. In this analysis, the health effects and efficiency are usually expressed as the incremental cost-effectiveness ratio (ICER), being the extra cost per unit of health gained.

The Dutch Health Care Insurance Board (in Dutch: College voor Zorgverzekeringen, CVZ) is primary responsible for the assessment of new outpatient drugs in the assessment phase. Subsequently, the final reimbursement advice is based on four principles, necessity, effectiveness, cost-effectiveness and feasibility, that are weighted in relation to one another and other possible arguments in the appraisal phase.[2] This is done by the Insured Package Advisory Committee (Advies Commissie Pakket, ACP). Based on this advice, the Minister of Health then makes the final reimbursement decision.

Although the introduction of the GVS was an important step forward towards rationalising health care decision making, it only applied to outpatient drugs. Inpatient drugs (i.e. drugs dispensed via hospitals) were financed at the hospital's discretion from their own fixed yearly hospital budgets. However, through the introduction of new, expensive parenterally administered drugs, this system led to financial difficulties, since budgets did not increase at the same pace as expenses. While the hospitals' fixed budgets controlled costs at the national level, they appeared to lead to differences in patient access to expensive inpatient drugs, as some hospital specialists were reluctant to prescribe expensive drugs.[3]

In a move to avoid unequal access and prevent delays in treatment, The Netherlands introduced in 2002 a new policy for reimbursement and financing of expensive inpatient drugs.[4] The policy enabled hospitals to receive up to 75% additional financing for drugs placed on the 'expensive drug list', above and beyond their yearly fixed budget. The actual percentage of additional funding depended on the negotiations between hospitals and healthcare insurers.

However, when it became clear that this policy still insufficiently ensured equality of access to these expensive drugs, the Dutch healthcare Authority amended this policy on 1st January 2006 and increased additional funding to a fixed 80% of drug costs. Next to this, additional requirements for a drug to be added and stay on the expensive drug list were formalised.

Similar to the inclusion of outpatient drugs in the GVS**,** placement on the expensive drug list requires an assessment of therapeutic value, cost-effectiveness and budget impact. However, there are two important differences. First, placement on the expensive drug list is conditional and will be re-evaluated after a maximum of four years. Second, during these four years it is mandatory to collect information on appropriate drug use and the relative cost-effectiveness. While information on effects and cost is commonly collected during an randomised controlled trial (RCT), the expensive drug policy requires the collection of data regarding Dutch daily clinical practice (i.e. real-world data).[5] The quality of the study proposal of this real-world study needs to be assessed before placement on this list of expensive drug.

After four years, a re-assessment and appraisal are conducted regarding the actual budget impact, real-world therapeutic added value, appropriate use

and cost-effectiveness in daily practice. The decision about whether or not to continue funding is mainly based on evidence from real-world data.

Since its introduction in 2006, the expensive drug policy has undergone several adjustments (e.g. changes in financial structures, percentage of additional financing and timelines). However, to date, a positive final reimbursement decision remains conditional upon the demonstration of an acceptable real-world cost-effectiveness.

The obligation to use real-world data in the expensive drug policy context has the potential of providing the decision maker with more relevant and realistic information on the costs and effects in daily practice and the pressure on the health care budget than RCT based analyses. However, using real-world data comes along with several challenges since it has important consequences for the methods to collect and analyse cost-effectiveness data.

This thesis demonstrates the added value (potential) of conducting real-world studies and addresses several related methodological challenges (pitfalls) in the field of expensive inpatient oncologic drugs. This is done in three linked parts. Part 1 consists of a detailed description of a case study where the use, effectiveness and cost-effectiveness of the drug oxaliplatin in the adjuvant treatment of stage III colon cancer are studied. Part 2 addresses the feasibility and several related methodological challenges of the use of real-world data. Three different case studies are used as empirical input: i) oxaliplatin in stage III colon cancer (case study of part 1), ii) oxaliplatin in metastatic colorectal cancer and iii) bortezomib in patients with multiple myeloma. Next to this a simulation study is conducted in part 3. This simulation study is also based on one of the case studies (oxaliplatin in metastatic colorectal cancer) and aims to further explore one particular methodological challenge: to correctly adjust for confounding by indication.

First, however, some background to the thesis will be provided by introducing the possible potential and pitfalls of real-world data in cost-effectiveness analyses. Subsequently, the case study of part 1 will be introduced.

POTENTIAL AND PITFALLS OF USING REAL-WORLD DATA

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. In ideal randomised experiments (i.e. large sample size, no selection of subgroups, adequately stratified, double blinded, no loss to follow up, full adherence to assigned treatment, etc.) internal validity is assured.

However, the patients enrolled in clinical trials may not be representative of the patients who will receive the therapy in daily practice. As a consequence, the treatment effectiveness and healthcare resource use in clinical trials may be different from those in daily practice. Differences between RCTs and real-world practice may exist in patient selection criteria, dosing regiments, the use of supportive care, and the intensity of follow-up.[6] Such differences can lead to important differences between cost-effectiveness outcomes based on RCT-data and outcomes based on real-world data. This is increasingly recognised by researchers, policy makers and decision makers responsible for pricing and reimbursement of healthcare interventions. Evaluations using data from real-world practice can provide policymakers with results that are more relevant and applicable and can address uncertainties arising from the gap between clinical trials and daily practice.[7-13]

Once a drug is available for prescription in daily practice, it is usually considered unethical to randomise patients to predefined treatments, which excludes a RCT –like design. For this reason, observational studies are mostly used to evaluate treatments in real-world practice. To adequately reflect daily practice, these observational studies do not have restrictive in- and exclusion criteria, they just monitor real-world patients, drug use, effectiveness, toxicities and costs of drugs and their comparators in daily practice.[14] Subsequently, one could estimate the real-world relative cost-effectiveness of this new drug.

Yet the approach of using observational studies for this purpose has its own methodological challenges to its feasibility. Based on theoretical expectations and expert opinion, the lack of a randomised controlled setting is one of the major concerns.[15-18]

Instead of a random assignment of treatments, in observational studies physicians decide on a patient's treatment 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. This means that treatment assignment and prognostic factors may be associated. If this so-called confounding by indication is not removed or reduced, the estimated treatment effect will be biased. Confounding by indication is one important reason to be very cautious in directly comparing two treatments using data from daily practice, because the internal validity of the comparison is not ensured by an observational study design.

To overcome these problems with internal validity, it is key to eliminate the impact of confounding, which can, to a certain extent, be done via several statistical adjustment methods. Traditionally, regression methods have been used for this purpose. More recently, methods based on propensity scores have become increasingly popular. The propensity score was defined by Rosenbaum en Rubin in 1983[19] as a patients' probability of receiving a specific treatment assignment, based on certain observed characteristics of that patient. In RCTs treatments are randomly assigned, therefore the propensity score is determined by the study design. In real life, however, the change to receive a treatment is based on specific characteristics such as disease severity and age. The propensity score is therefore not known, but can be estimated. This is most often done by using a logistic regression model, in which treatment status is regressed on the observed baseline characteristics. The idea behind propensity score techniques is that patients with similar (predefined) characteristics would have a similar likelihood of receiving a certain treatment. Propensity score techniques can be used in several ways to address confounding. Techniques include matching, stratification on the propensity score, inverse probability of treatment weighting using the propensity score, weighting by the odds, and regression adjustment with the propensity score as a covariate.[20-23] Based on propensity score matching, for instance, it is possible to select two comparable treatment arms without conducting an RCT. This, potentially, makes propensity score techniques particularly useful in real-life data studies.

Propensity score 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 propensity score methods was produced in the same context.[21, 22, 24] Although examples of the application of propensity

score methods also exist in cost-effectiveness analysis,[25-34] they are still relatively scarce. Two methods, inverse probability of treatment weighting and double robustness, where regression and weighting by the propensity score are combined in one equation, have not been used in observational cost-effectiveness studies. Furthermore, the properties of propensity score methods in this field have not been investigated extensively. 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 propensity score methods. First, cost data have different properties than data on clinical effectiveness. More importantly, health economic studies examine two outcomes simultaneously, incremental costs and incremental effects, to estimate incremental cost-effectiveness ratios (ICERs). This might have implications for the choice of method when correcting for confounding by indication in this type of studies.

Next to potential problems with internal validity due to confounding by indication, also other aspects of real-world data use have been debated in the literature, such as the role of data synthesis and modelling,[8, 17, 35, 36] use of existing data sources,[35, 37-40] applicability of outcome measures,[8, 12] timeliness[9, 41] and generalisability.[42] However, these issues have mainly been discussed on the basis of theoretical expectations and expert opinions; case studies actual implementing real-world data in determining appropriate drug use and its relative cost-effectiveness are scarce.

CASE-STUDY: DRUG EVALUATION IN COLORECTAL CANCER

Colorectal cancer (CRC) places a considerable burden on individuals and society in Europe, being the second most common cause of cancer-related death in the region.[43, 44] In The Netherlands, each year about 10,000 people are diagnosed with colorectal carcinoma. The number of patients diagnosed with colorectal carcinoma is expected to rise to about 14,000 in 2015 due to a slowly increasing incidence (particularly among men), the growing population and the ageing population.[45]

Colorectal carcinoma usually develops from a polyp, which is a growth or thickening of the mucous membrane that lines the intestines. While most polyps are usually benign, some develop into cancer over time. Stage I and II

invasive colorectal cancers confined to the wall of the colorectum comprise 40% of all cases of colorectal cancer. In stage III, occurring in 37% of the patients, the carcinoma extends to the regional lymphatic glands. In stage IV patients, who comprise 19% of all colorectal cancer patients at diagnosis, the cancer has already spread to distant sites.

The stage of disease plays an important role in treatment options. In stage I, II and III patients, surgery is the recommended treatment option and is performed with curative intent. However, nearly half of the patients who undergo curative surgery will ultimately relapse and die of metastatic disease.[46, 47] Stage IV patients have advanced or metastatic disease which is usually not curable. For decades, chemotherapy is part of the standard treatment of metastatic colorectal cancer, since it has clearly been shown to lengthen the progression-free survival and overall survival.

For stage III colon cancer patients, chemotherapy adjuvant to initial surgery has clearly been shown to lengthen disease-free survival and overall survival and 6 months of adjuvant chemotherapy has been part of the standard therapy since the nineties.[48-52] Until 2005, treatment with intravenous 5-fluorouracil and leucovorin (5FU/LV), with a 3-year survival of 65% was the only available effective adjuvant chemotherapy for patients with stage III colon carcinoma.[50-52]

In 2004, the chemotherapy drug oxaliplatin obtained registration for the adjuvant treatment of stage III colon cancer, based on a large phase III clinical trial in colon cancer (Multicentre International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer [MOSAIC]). This MOSAIC trial had shown that the addition of oxaliplatin to 5FU/LV significantly improved disease-free survival compared to 5FU/LV alone, against acceptable toxicity. Among patients with stage III disease, the hazard ratio for relapse was 0.76 (95 percent confidence interval, 0.62 to 0.92) in the group given 5FU/LV plus oxaliplatin, as compared with the 5FU/LV group, and the three-year disease-free survival rate was 72.2 percent and 65.3 percent, respectively. At that time, however, no improved overall survival benefit could be demonstrated, due to immaturity of the data on overall survival.

Based on the MOSAIC results, 6 months of oxaliplatin combined with 5FU/LV (FOLFOX) became the standard adjuvant treatment in the Netherlands for stage III colon cancer patients as of early 2005.[53] Treatment with 5FU/LV alone remained indicated for patients who were not eligible or refused treatment with

oxaliplatin. At that time the oral fluoropyrimidine capecitabine, alone or in combination with oxaliplatin (CAPOX), also became available as an alternative to 5FU/LV or FOLFOX, as these treatments were found to be equally effective in the treatment of stage III colon or metastatic colorectal cancer.[54-57]

However, the introduction of oxaliplatin in the adjuvant treatment of colon cancer led to a substantial cost increase. The average additional costs of oxaliplatin were estimated at about € 12,500 per six months (based on drug cost only). Based on this, oxaliplatin has been eligible for additional funding and was included in the list for expensive medicines for the adjuvant treatment of stage III colon cancer, early 2005.

The abovementioned knowledge regarding the expected effectiveness and costs was roughly all available evidence on the adjuvant treatment with oxaliplatin at that time. Uncertainty regarding overall survival benefit, real-world appropriate use, real-world (comparative) effectiveness and cost-effectiveness existed. Although in 2005 the conditional reimbursement policy with the requirement of additional real-world data collection did not yet exist, the case of oxaliplatin served as a pilot example and its real-world investigation, during 3 years after its listing as an expensive drug, is topic in part 1 and 2 of this thesis.

QUESTIONS ADDRESSED IN THIS THESIS

Based on the above discussion on the potential and pitfalls of using real-world data in the evaluation of expensive drugs, each of the three parts of this thesis addresses a question:

1. How is oxaliplatin used in the adjuvant treatment of stage III colon cancer in real-world practice and what are the real-world costs, effects and cost-effectiveness?

2. What is the feasibility of assessing appropriate drug use and real-world cost-effectiveness in oncologic drugs and what are the most important challenges in light of the Dutch expensive drug policy? How can these challenges best be addressed?

3. How should propensity-score based adjustment methods be applied in observational cost-effectiveness studies?

By answering these questions, this thesis seeks to contribute to the ongoing debates regarding the potential and pitfalls of using real-world observational studies to demonstrate the effectiveness and cost-effectiveness of drugs.

OUTLINE OF THIS THESIS

This thesis is structured as follows. Part 1 consists of the following three chapters, which are based on a Dutch population-based observational study in stage III colon cancer patients. Chapter 2 evaluates the current guideline for the treatment of stage III colon cancer by examining guideline implementation, treatment patterns and disease-free survival. Chapter 3 describes the real-world resource use and costs of the adjuvant treatment for stage III colon cancer. Subsequently, the real-world cost-effectiveness of the addition of oxaliplatin to the existing adjuvant treatment of stage III colon cancer was investigated in chapter 4. These estimates were based on a combination of published MOSAIC trial data with data from our Dutch population-based observational study.

Part 2 is predominantly based on chapter 5. This chapter describes our experiences in The Netherlands regarding the feasibility of different aspects of real-world cost-effectiveness research. These experiences were obtained from three different case studies, including the case study described in part 1 of this thesis.

Chapter 6 and 7 belong to part 3 of this thesis. Both chapters are based on a simulation study, where several confounding adjustment methods are compared in a hypothetical cohort of patients with metastatic colorectal cancer. Chapter 6 compares the validity and accuracy of the results of several adjustment methods and chapter 7 studies the reliability of these methods in the context of economic evaluations.

Chapter 8 draws together the results of previous chapters, provides a discussion of the results and explores the implications and the limitations of this thesis.

To note, the chapters of this thesis are based on research articles published or planned to be published in scientific peer reviewed journals. As a result, the chapters of this thesis can be read independently and some overlap exists between some of the chapters.

Chapter 2

Adjuvant chemotherapy in stage III colon cancer: Guideline implementation, patterns of use and outcomes in daily practice in The Netherlands

Acta Oncologica 2012; 51: 57–64

Chantal van Gils
Miriam Koopman
Linda Mol
Ken Redekop
Carin Uyl- de Groot
Kees Punt

SUMMARY

Background: Little is known about how well guidelines about adjuvant chemotherapy in colon cancer are followed in daily practice. We evaluated the current guideline, which is based on the MOSAIC trial, by examining implementation, treatment patterns and disease-free survival. Methods: We analysed a population-based cohort of 391 patients treated with adjuvant chemotherapy for stage III colon cancer in 2005-2006. Data were gathered from the Dutch Cancer Registry and medical records of 19 hospitals. Patients were classified according to whether or not they fulfilled MOSAIC trial eligibility criteria. Results: The administered regimens were: 5FU/LV (17 patients), capecitabine (93), FOLFOX (145), and CAPOX (136). After its inclusion in national guidelines, oxaliplatin was prescribed in 16 hospitals within 6 months. Patients receiving oxaliplatin were younger and had less comorbidity than other patients. Dose schedules corresponded well with guidelines. Two-year disease-free survival probability of oxaliplatin patients meeting MOSAIC eligibility criteria was 78.4% (95%CI 72.5–84.3), which was comparable to MOSAIC trial results. Conclusion: Guidelines for adjuvant chemotherapy in stage III colon cancer are generally well followed in daily practice. However, uncertainty remains regarding the optimal treatment of elderly patients and patients with comorbidities, which underscores the need for practical clinical trials including these patients.

INTRODUCTION

Colon and rectal cancers are the second most common cause of death in Western countries.[58] Nearly half of the patients who undergo curative surgery will ultimately relapse and die of metastatic disease.[46] During the 1990s the survival rates of patients with stage III colon cancer significantly improved by the introduction of adjuvant chemotherapy with 5-fluorouracil and leucovorin (5FU/LV).[52]

As result of the publication of the MOSAIC trial in 2004, which demonstrated that adding oxaliplatin to 5FU/LV improved the adjuvant treatment, clinical practice guidelines in The Netherlands were changed early 2005.[53] National guidelines since then have recommended the use of 6 months of treatment with 5FU/LV combined with oxaliplatin (FOLFOX) as the primary treatment option for stage III and possibly high-risk stage II colon cancer patients. In addition, the oral fluoropyrimidine capecitabine was indicated for patients who are not eligible or who refuse treatment with oxaliplatin, based on the X-ACT trial.[59] Also the use of capecitabine combined with oxaliplatin (CAPOX) as an alternative to FOLFOX was supported by the Dutch association for Medical Oncology (NVMO), as these treatments had shown comparable efficacy in metastatic colorectal cancer.[55, 60, 61] Data on the efficacy of CAPOX in the adjuvant setting were not available at that time.

In light of more recent evidence from RCTs, this strategy proved to be valid and in line with the current international clinical practice guidelines.[62-65]

However, the nationwide level of implementation of the primarily RCT-based guidelines and its impact on population-based clinical outcomes is unknown. For instance, differences between RCTs and daily practice may exist in the patient selection criteria, dosing regimens, the use of supportive care, and the intensity of follow-up.[6] Observational studies including detailed information on chemotherapy use in daily practice can complement findings from RCTs and allow post-implementation evaluation of guidelines.[14]

In our study we retrospectively analysed population-based data of stage III colon cancer patients treated with adjuvant chemotherapy in the first two years after the change of the Dutch clinical practice guideline. The aim was to examine the speed of guideline implementation and to compare the guideline to chemotherapy use in daily practice with respect to treatment choice, patient

characteristics and dosage quantities. In addition, we compared the disease-free survival outcomes of patients receiving FOLFOX and CAPOX in Dutch daily practice with the outcomes of patients receiving FOLFOX in the MOSAIC trial.

METHODS

Data and cohort construction

The primary data source for the population-based study was the Netherlands Cancer Registry (NCR), which registers information on demographics, tumour characteristics and survival outcomes of more than 95% of all new cancer cases in the Netherlands. All stage III colon cancer patients (pTanyN1, 2M0, ICD-O C18-C19.9) who were diagnosed in 2005 or 2006, and who received adjuvant chemotherapy were identified in the NCR. To gather additional information, we approached 72% of the hospitals in The Netherlands and included the 19 first responding hospitals into our study. The medical files of all identified patients were reviewed in these 19 selected hospitals (3 university hospitals, 9 large teaching hospitals, and 7 general hospitals dispersed over The Netherlands), which together were considered to be a good representation of clinical daily practice in The Netherlands. Data were collected on baseline characteristics, eligibility criteria used in the MOSAIC trial, treatment schedules and disease-free survival. We recorded comorbid conditions using a slightly adapted version of the Charlson index, which classifies all serious comorbid conditions based on possible prognostic impact into eight groups (i.e. previous malignancies, chronic obstructive pulmonary diseases, cardiovascular diseases, cerebro-vascular diseases, hypertension, diabetes mellitus, digestive tract diseases and other).[66, 67] Reasons for not prescribing oxaliplatin were also recorded. Additional information on treatment schedules, dose reduction, delay and/or interruptions of treatment and its reasons was recorded in a randomly selected subset of patients.

Statistical Analyses

To check the representativeness of the 19 selected hospitals for The Netherlands, we first compared the average percentage of patients receiving adjuvant chemotherapy among hospitals included in our study versus other hospitals

in the Netherlands by means of the Students t-test. Also their median age was compared. Next, we assessed the frequency of administration of treatment in the selected hospitals. Two groups of patients were created ("receiving oxaliplatin based regimens" and "receiving regimens without oxaliplatin") and the baseline characteristics of these two groups were compared using the Students t-test for continuous variables and the $\chi 2$ test for dichotomous or nominal values. Next, we investigated the uptake of new treatments as recommended by the new guidelines by the hospitals. We used the Cochran-Armitage trend test to test for time trends in the use of different treatment regimens. Reasons for not prescribing oxaliplatin were explored using descriptive statistics. A multivariate logistic regression analysis was performed to identify independent predictors of non-prescription of oxaliplatin. Dose schedules and modifications were compared using the tests for continuous and categorical variables mentioned above. Subsequently, disease-free survival was calculated from the date that chemotherapy started until relapse or death or censored on the date last known to be alive. Based on the published MOSAIC trial, patients receiving oxaliplatin were grouped into "fulfilling MOSAIC eligibility criteria" and "not fulfilling MOSAIC eligibility criteria".[53] Per group, the 2-year disease-free survival rate was estimated using the Kaplan-Meier method and compared to the 2-year disease-free survival rate of the stage III patients receiving oxaliplatin in the MOSAIC trial by means of the $\chi 2$ test. The MOSAIC 2 year-disease-free survival rate and standard error were derived from the published Kaplan-Meier curve and number of patients at risk at 24 months.[64] In all analyses, statistical significance was assumed if the two-tailed probability value was less than 0.05. The SAS computer package (version 8.2) was used for all statistical analyses (SAS Institute Inc., Cary, NC, USA, 1999).

RESULTS

Patients and treatments

Between January 2005 and December 2006, 4010 patients were diagnosed with stage III colon cancer in the Netherlands, of whom 2249 were treated with adjuvant chemotherapy (Figure 2.1). A total of 423 patients were treated at one of the 19 hospitals included in our study. The average percentage of patients receiving

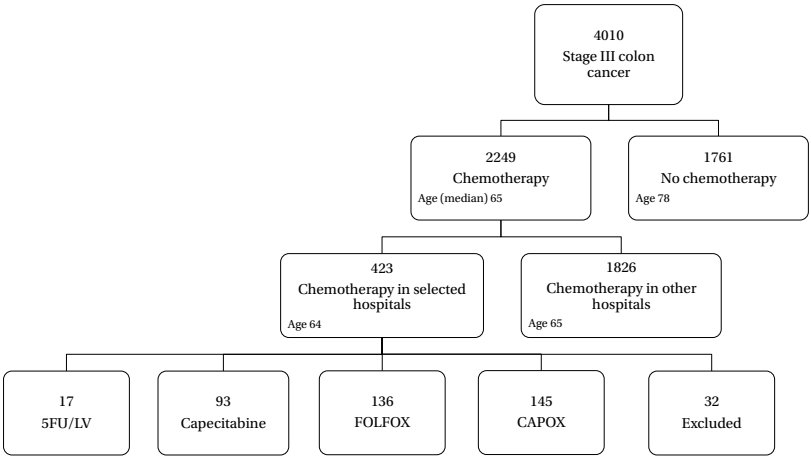


Figure 2.1 Study profile. Number of patients registered by The Netherlands Cancer Registry in 2005-2006.

adjuvant chemotherapy in the 19 included hospitals was 53% versus 57% in 92 not included hospitals (p = 0.17). Also the median age of the selected versus non-selected patients was comparable (64 versus 65 years). The four most commonly administered regimens were: 5FU/LV (17 patients), capecitabine (93), FOLFOX (145), and CAPOX (136). Five patients were excluded from further analysis: 3 patients received bevacizumab and 2 patients UFT as adjuvant chemotherapy (<2%). Furthermore, 27 patients were excluded because of inclusion in clinical trials (16), diagnosis of a second malignancy in the past five years (9), and missing files (2).

We observed a rapid adoption of oxaliplatin in the period shortly after Dutch national guidelines recommended it for the adjuvant treatment of colon cancer at the start of 2005 (Figure 2.2). Of the 19 hospitals included in our survey, 8 were already using oxaliplatin in the first quarter of 2005, followed by a total of 16 hospitals using oxaliplatin during the second quarter of 2005. By January 2006, oxaliplatin was standard therapy in all 19 hospitals. Furthermore, a significant trend from FOLFOX use to CAPOX use was observed between January 2005 and December 2006 (Ptrend < 0.001). In the second quarter of 2005, 82% of the patients receiving oxaliplatin were treated with FOLFOX versus 18% with CAPOX. By the start of 2007 only 27% were being treated with FOLFOX versus 73% with

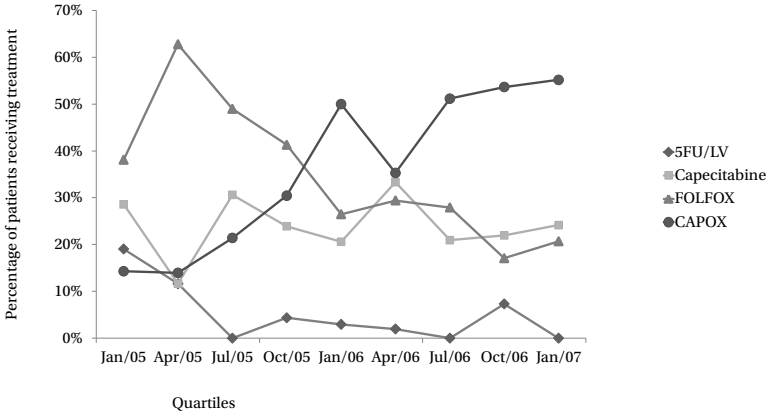


Figure 2.2 Distribution of regimen use from the first quartile of 2005 to the first quartile of 2007, by treatment group.

CAPOX. However, despite the rapid adoption of oxaliplatin use on a hospital level, a substantial proportion of the patients did not receive oxaliplatin-based regimens. The percentage of patients not receiving oxaliplatin was 28% and this percentage did not change over time (Ptrend = 0.77). Already since the first quarter of 2005, the majority of these patients was treated with capecitabine instead of 5FU/LV.

The baseline characteristics of the patients in the four treatment groups are summarised in Table 2.1 Patients receiving oxaliplatin were significantly younger (P < 0.0001) and had fewer comorbidities (P = 0.001) than patients who did not receive oxaliplatin. Furthermore, patients receiving oxaliplatin more often had well-differentiated tumour histology (P = 0.007) and higher serum carcinoembryonic antigen (CEA) levels (P = 0.028) than other patients. Additional stratification by age (older versus younger than 70 years of age) revealed that these differences in tumour differentiation and CEA levels could be explained by the different age distribution in the two groups.

We next explored reasons why some patients did not receive oxaliplatin (111 patients). The reasons for non-prescription were: prescription not in line with hospital policy (23%), advanced age (21%), patient refusal (18%), comorbidity or poor health status (10%), specific contra-indications for oxaliplatin (2%),

Table 2.1 Baseline characteristics of stage III colon cancer patients receiving chemotherapy in Dutch daily clinical practice.

Baseline Characteristics	Regimens without oxaliplatin		Oxaliplatin regimens		P-values
	5FUL/LV	Capecitabine	FOLFOX	CAPOX	
	N = 17	N = 93	N = 136	N = 145	
Age - yr					
Median	71	73	61	62	< 0.0001
Range	41-80	46-85	30-78	22-82	
Age group - no. (%)					
< 70	7 (41.2)	31 (33.3)	118 (86.8)	123 (84.8)	< 0.0001
≥ 70	10 (58.8)	62 (66.7)	18 (13.2)	22 (15.2)	
Comorbid conditions - no. (%)					
0 - 1	12 (70.6)	71 (76.3)	115 (84.6)	134 (92.4)	0.001
2+	5 (29.4)	22 (23.7)	21 (15.4)	11 (7.6)	
Sex - no. (%)					
male	9 (52.9)	47 (50.5)	72 (52.9)	81 (55.9)	0.528
female	8 (47.1)	46 (49.5)	64 (47.1)	64 (44.1)	
Depth of invasion - no. (%)					
T2 -T3	15 (88.2)	82 (89.1)	116 (85.3)	123 (84.8)	0.313
T4	2 (11.8)	10 (10.9)	20 (14.7)	22 (15.2)	
Unknown		1			
Nodes involved - no. (%)					
N1	11 (64.7)	62 (66.7)	84 (61.8)	85 (58.6)	0.255
N2	6 (35.3)	31 (33.3)	52 (38.2)	60 (41.4)	
Histologic appearance - no. (%)					
Well differentiated	11 (64.7)	75 (81.5)	111 (86.7)	125 (91.9)	0.007
poorly differentiated	6 (35.3)	17 (18.5)	17 (13.3)	11 (8.1)	
Unknown		1	8	9	
CEA level - no.					
< 5 ng/ml (ULN)	10 (83.3)	69 (93.2)	94 (83.2)	105 (80.8)	0.028
≥ 5 ng/ml (ULN)	2 (16.7)	5 (6.8)	19 (16.8)	25 (19.2)	
Unknown	5	19	23	15	

CEA, carcinoembryonic antigen; ULN, upper limit of normal

combination of these factors (7%), and unknown (23%). To identify independent predictors of non-prescription of oxaliplatin, we performed a multivariate logistic regression on baseline characteristics and included the variables age, presence of comorbid conditions, gender, depth of invasion of primary tumour (T-stage), lymph node involvement (N-stage), differentiation and serum CEA level. The multivariate analysis identified only age and comorbidity as being independent predictors of non-prescription of oxaliplatin (OR [95 CI] of 0.765 [0.708-0.826] and 0.426 [0.169-1.075], respectively).

Dose schedules

To evaluate dose schedules, additional data were collected from the medical records of a randomly selected subset of 206 patients. This selection was also stratified by hospital and oxaliplatin use to ensure equal numbers of patients that did and did not receive oxaliplatin.

Table 2.2 provides an overview of the patterns of use of the different treatment regimens in daily practice. With six months of chemotherapy being accepted as the standard duration of adjuvant treatment, and a treatment cycle of 2 weeks

Table 2.2 Planned and actually delivered dose in Dutch daily clinical practice.

			FOLFOX		CAPOX	
	5FUL/LV	Capecitabine	5FU/LV	Oxaliplatin	Capecitabine	Oxaliplatin
	N = 15	N = 89	N = 37		N = 65	
Median nr of cycles received (planned nr of cycles)	**	8 (8)	12 (12)	12 (12)	8 (8)	7 (8)
Dose planned according to guidelines in mg/m^2/wk	**	11666	1000 / 200	43	9333	43
Mean dose over all cycles given in mg/m^2/wk	**	9659	890 / 178	42	8049	42
Mean dose over all planned cycles in mg/m^2/wk	**	8250	800 / 160	36*1	7052	30*1
% of patients requiring modification (for dose reduction or interrupation)	53%	57%	54%	59%*2	50%	70%*2
% of planned dose given	72%	83%	84%	84%*3	79%	71%*3

** Not reported because of diversity of dose schedules and low patient numbers; *1 p - value = 0.0213, *2 p - value = 0.2661, *3 p - value = 0.0896, Oxaliplatin in FOLFOX versus CAPOX.

for FOLFOX and 3 weeks for CAPOX and capecitabine, the median number of planned cycles was 12, 8 and 8, for FOLFOX, CAPOX and capecitabine, respectively. The median number of cycles received equalled the planned number of cycles in FOLFOX and capecitabine, indicating that at least 50% of the patients completed the number of cycles that was expected according to the protocol. The median number of oxaliplatin cycles for patients receiving the CAPOX regimen was 7.

The planned dose for each regimen was equal to the dosing recommendations of the national guidelines. The mean dosages in milligrams per square metre per week across all cycles administered were slightly lower than the planned dosages. When we calculated the mean dose in the administered cycles as a percentage of the mean dose advised in the guidelines, we found that 83% of the recommended dose was given in capecitabine monotherapy and even 98% of the oxaliplatin in the FOLFOX and CAPOX regimens. However, regarding the mean dose over all planned cycles, we found that the mean dose of oxaliplatin in CAPOX was significantly lower than that in FOLFOX, with 30 mg/m2/wk versus 36 mg/m2/wk, respectively (P = 0.0213). This also resulted in a significantly lower mean cumulative dose of oxaliplatin (CAPOX: 780 mg/m2 versus FOLFOX: 936 mg/m2, P = 0,002). In total, 71% of the planned doses amongst CAPOX-treated patients were administered versus 84% amongst FOLFOX-treated patients (P = 0.0896).

Disease-free survival

The recommendation to use oxaliplatin in the adjuvant treatment of stage III colon cancer in The Netherlands was based on the results of the MOSAIC trial. For this reason we compared the 2-year disease-free survival (DFS) rates of the patients receiving oxaliplatin-based regimens in Dutch daily practice to the results of the MOSAIC trial.

Of the 281 patients receiving oxaliplatin-based regimens, 200 met the MOSAIC eligibility criteria, 43 did not and 38 could not be classified due to missing data. They were excluded from further analysis. Reasons for ineligibility were: chemotherapy treatment did not start within seven weeks after surgery (27 patients), CEA levels above 10ng/ml (12), age older than 75 years (1), and combination of mentioned reasons (3). The 2-year probability of DFS of eligible and ineligible patients was 78.4% (95% CI 72.5 - 84.3) and 56.7% (95% CI 41.4

- 72.0), respectively (Figure 2.3). The published Kaplan-Meier curve for stage III patients receiving oxaliplatin in the MOSAIC trial showed a 2-year DFS probability of 79.5% (95% CI 75.6- 83.4). This probability was not significantly different from the 2-year DFS of eligible oxaliplatin patients from Dutch daily practice (P = 0.32).

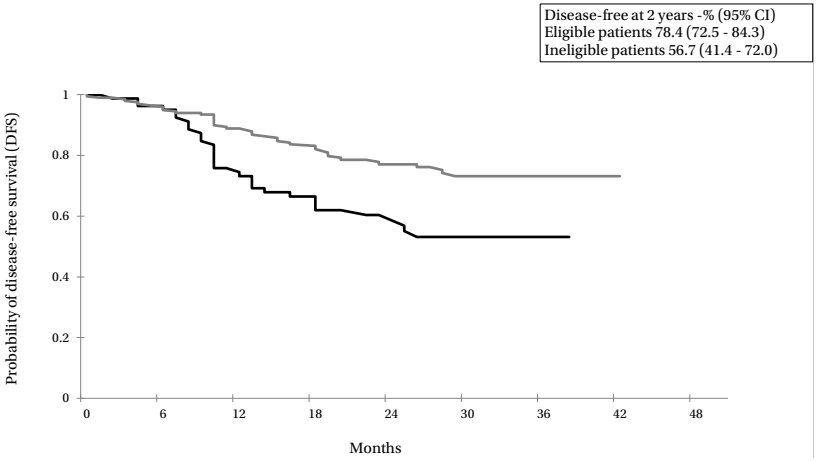


Figure 2.3 Kaplan-Meier curves of disease-free survival in the groups of oxaliplatin patients that did and did not meet the MOSAIC eligibility criteria.

DISCUSSION

In this study analysing population-based data from Dutch daily practice in 2005-2006, we evaluated the clinical practice guideline for the adjuvant treatment of stage III colon cancer.

When treatment with oxaliplatin, either in the FOLFOX or CAPOX regimen, became the new standard therapy in early 2005, we observed a quick implementation with the majority of the hospitals already using oxaliplatin in the second quarter of 2005. This rapid adoption is most likely due to the extensive experience that the physicians already had with oxaliplatin as an important treatment in advanced colorectal cancer.[60] Over time, we found an increasing preference for the use of CAPOX over FOLFOX. This preference was probably

due to the need for intravenous access devices in the administration of 5FU/LV in the FOLFOX regimen.

We also observed a rapid adoption of capecitabine monotherapy as an alternative to 5FU/LV. Like in oxaliplatin, this was most likely due to the extensive experience already obtained in advanced colorectal cancer and because of the ease of the oral capecitabine administration.[60]

The Dutch guideline does not specifically indicate who is eligible for oxaliplatin and who is not, leaving this decision up to the judgement of the physicians. We observed that physicians were reluctant to prescribe oxaliplatin to patients with advanced age or serious comorbidities. One other population-based study reported on prescription of oxaliplatin as adjuvant chemotherapy and found similar results.[68] The median age of the patients not receiving oxaliplatin was 73, whereas this was 61 in patients receiving oxaliplatin in daily practice. The latter number equals the median age of the patients randomised in the MOSAIC trial suggesting that physicians considered the MOSAIC criteria when deciding on the prescription of oxaliplatin. Over the past years several studies have reported conflicting results regarding the efficacy and safety of oxaliplatin-based chemotherapy in older patients.[69-71] As a consequence, uncertainty remains regarding the question whether the lower rates of oxaliplatin treatment we observed in older patients represent wise clinical judgement or undertreatment.

Specific guideline recommendations are also lacking regarding the eligibility for capecitabine or 5FU/LV without oxaliplatin as adjuvant treatment option. We observed that a substantial part of the stage III colon cancer population in The Netherlands did not receive any adjuvant chemotherapy in 2005 and 2006. They were older than patients receiving chemotherapy (median age 78). This finding is in line with results from studies that found that age and comorbidities were associated with the decision to prescribe 5FU/LV or no adjuvant therapy in stage III colon cancer.[72, 73] However, since patients not receiving any adjuvant chemotherapy were not included for additional data collection, this chapter provides no further insights in this elderly patient population.

In general, the observed dose schedules demonstrated a good adherence to existing guidelines. The mean dosages in milligrams per square metre per week across all cycles administered were only slightly lower than those recommended by the guidelines. However, more than half of the patients in all regimens also needed dose modifications resulting in lower total cumulative

dosages than could theoretically have been administered. But even in trials the administered dose is usually lower than the planned dose. For FOLFOX, the dosages of oxaliplatin given in daily practice were similar to the dosages reported in the MOSAIC trial.[53] Also for capecitabine monotherapy these dosages were similar in RCT findings.[54] However, regarding CAPOX, patients in daily practice received on average 71% of the planned dose, whilst the literature reports this to be 87%, although part of this difference can probably be explained by the fact that the latter number reflects the median percentage rather than the average.[63] Also when comparing the FOLFOX and CAPOX regimens in our study, we found a lower mean dose of oxaliplatin, more dose modifications and a lower percentage of planned dose given in de CAPOX regimen, all together suggesting that oxaliplatin is less well tolerated in CAPOX versus FOLFOX. A pooled analysis in advanced colorectal cancer also found that some toxicities were slightly but consistently more prominent in capecitabine containing regimens.[74] However, dosage comparisons in our study need to be interpret with caution since the decision to select patients using random sampling that was stratified by hospital and oxaliplatin use, resulted in the selection of only 37 patients who received FOLFOX.

The disease-free survival of the eligible patients receiving oxaliplatin was comparable to that of the MOSAIC patients receiving oxaliplatin. Our result supports the external validity of the MOSAIC trial results, which in general has been a matter of concern in RCTs.[6] A limitation of our study here is that although we used the same definition of disease-free survival as presented in the MOSAIC trial, we cannot guarantee that the same method is used in both studies. The estimated time to occurrence depends on the monitoring of the patients during follow-up which might have been less intense in daily practice as compared to trial monitoring. This might have resulted in a delayed diagnosis of relapse in daily practice. However, the proposed follow-up schedule in the Dutch guideline does not differ from the follow-up schedule followed in the MOSAIC trial. Moreover, we calculated that a delay of 3 months would not have an effect on the conclusion that the disease-free survival is similar in both studies. Between 2005 and 2010, two other RCTs reported similar outcomes when using oxaliplatin in the adjuvant treatment of colon cancer.[62, 75]

Our finding of a decreased disease-free survival in patients who did not meet the trial eligibility criteria underscores the fact that trial results should not be

extrapolated to other patient categories in daily practice.[76] The less favourable prognosis in non-eligible patients can mainly be explained by the significantly higher CEA values of these patients. Although it is uncertain whether adjuvant therapy with oxaliplatin has an added value here, it is unlikely that oxaliplatin would be unfavourable for this group of patients as oxaliplatin also plays an important role in the treatment of metastatic colon carcinoma.

In conclusion, our results point towards a quick nation-wide implementation of the stage III colon cancer clinical guideline after its change early 2005. We observed a good concordance of practice with the RCT-based treatment recommendations and similar disease-free survival outcomes of trial eligible patients receiving oxaliplatin in daily practice versus patients receiving oxaliplatin in RCTs. However, uncertainty regarding the optimal treatment for elderly patients or patients with serious comorbidities is still present today. The lack of specific guideline recommendations for this large and increasing patient population underscores that practical clinical trials for elderly patients with stage III colon cancer are needed.

Chapter 3

Real-world resource use and costs of adjuvant treatment for stage III colon cancer

Accepted for publication in European Journal of Cancer Care

Chantal van Gils
Saskia de Groot
Siok Swan Tan
Ken Redekop
Miriam Koopman
Kees Punt
Carin Uyl- de Groot

SUMMARY

Since the generalisability of trial-based economic evaluations may be limited, there is an increasing focus on real-world cost-effectiveness. Real-world studies involve evaluating the effects and costs of treatments in daily clinical practice. This study reports on the real-world resource use and costs of adjuvant treatments of stage III colon cancer in a population–based observational study. Analyses were based on a detailed retrospective medical chart review which was conducted for 206 patients with colon cancer treated in 2005 and 2006 in the Netherlands. Mean total costs per patient were €9,681 for 5-FU/LV, €9,736 for capecitabine, €32,793 for FOLFOX and €18,361 for CAPOX. Drug costs and the costs related to hospitalisations for chemotherapy administration were the main cost drivers. We identified a potential for substantial cost-savings when the 48 hour administration of 5FU/LV in the FOLFOX regimen were to take place in an outpatient setting or be replaced by oral capecitabine as in the CAPOX regimen. This analysis based on detailed real-life data clearly indicates that clinical choices made in oncology based on efficacy of therapy have economic consequences. Considering today's reality of finite healthcare resources, these economic consequences deserve a formal role in clinical decision making, for instance in guideline development.

INTRODUCTION

Colon and rectal cancers are among the most common types of cancer in the western world with more 215,000 deaths in Europe in 2012.[77]

For stage III colon cancer, surgical removal of the tumour is the primary treatment option. However, nearly half of the patients who undergo curative surgery as sole treatment will ultimately relapse and die of metastatic disease.[46]

During the 1990s the survival rates of patients with stage III colon cancer significantly improved by the introduction of adjuvant chemotherapy with 5-fluorouracil and leucovorin (5FU/LV).[49] In 2004, the oral fluoropyrimidine capecitabine demonstrated at least equivalent clinical benefit.[59] Subsequently, the MOSAIC trial demonstrated that the addition of oxaliplatin to 5FU/LV chemotherapy (FOLFOX) could further improve survival.[53] More recently, the clinical benefit of adding oxaliplatin to capecitabine chemotherapy (CAPOX) compared to 5FU/LV alone was also demonstrated.[78]

Since 2005, national guidelines in the Netherlands have recommended the use of oxaliplatin in combination with 5-FU/LV or capecitabine as the primary adjuvant treatment option for stage III colon cancer. Capecitabine is the preferred treatment option when oxaliplatin is not indicated.[60]

The improved treatment options and the nationwide implementation of oxaliplatin and capecitabine for patient with stage III colon cancer have increased survival rates, but also led to a substantial cost increase. In 2005, costs of colorectal cancer in the Netherlands were estimated at 273 million euro, which equals 10% of the total health expenditures of cancers and 0.4% of total health expenditures.[79] Annual costs of colorectal cancer are expected to rise due to the growth and ageing of the population.[45] Moreover, the implementation and expanded use of new expensive drugs in this disease area will further raise the costs.[80, 81] This will pose an increasing financial burden on health care systems.

With the aim to make better use of a limited health care budget, decision makers carefully balance the benefits of a treatment against its costs in economic evaluations when making reimbursement decisions. To inform these decision makers, several economic evaluations have examined the cost-effectiveness of adjuvant treatments in stage III colon cancer. Commonly, the cost-effectiveness outcomes in these evaluations are based on data collected within randomised controlled trials (RCTs).[82-88] However, the generalisability

of RCT-based economic evaluations into real-world practice may be limited.[8] Differences between RCTs and real-world practice may exist in patient selection criteria, dosing regiments, the use of supportive care, and the intensity of follow-up.[6]

Such differences can lead to strong differences between cost (-effectiveness) outcomes based on RCT-data and outcomes based on real-world data. For this reason, decision makers are increasingly interested in effectiveness and cost-effectiveness studies that make use of real-world data. This allows them to make better decisions regarding the reimbursement of new drugs, based on the population that is actually using these drugs.

In the past few years, several papers have reported on the effectiveness of adjuvant chemotherapy for stage III colon cancer in real-world practice.[67, 89-91] In an earlier publication describing patterns of chemotherapy use and outcomes in daily practice, we already showed that Dutch patients receiving adjuvant chemotherapy are different from patients in RCTs.[92] For example, Dutch patients receiving oxaliplatin in daily practice more often had elevated serum carcinoembryonic antigen levels, which resulted in less favourable disease-free survival outcomes of these patients compared to trial patients. Taking these real-world patients into account will increase the representativeness of the effectiveness estimates in real-world practice.

To our knowledge, only two studies report on real world resource use and costs of stage III colon cancer: one Chinese and one Italian study.[93, 94] A study reporting on Dutch real-world resource use and costs of the adjuvant treatment of colon cancer has not been published, yet. Therefore, we aimed to investigate the real-world resource use and costs of the adjuvant treatment for stage III colon cancer. To meet this aim we conducted a population based observational study of Dutch patients.

METHODS

Patient population

The primary data source for this study was the Netherlands Cancer Registry (NCR), which collects information on demographics, tumour characteristics and survival outcomes of more than 95% of all new cancer cases in the Neth-

erlands. All stage III colon cancer patients (pTanyN1, 2M0, ICD-O C18-C19.9) who were diagnosed in 2005 or 2006 were identified in the NCR (n = 4010). A total of 2249 were treated with adjuvant chemotherapy, of whom 423 patients were treated in one of the 19 hospitals included in our study (3 university hospitals, 9 large teaching hospitals, and 7 general hospitals), which comprised approximately 20% of the Dutch hospitals and together were considered to be a good representation of clinical daily practice in The Netherlands. To gather additional information, we reviewed medical files of all identified patients in the 19 selected hospitals. A total of 32 patients were additionally excluded from further analysis because of administration of bevacizumab or uracil/tegafur as adjuvant treatment (5), inclusion in clinical trials (27), diagnosis of a second malignancy in the past five years (9), and missing files (2).

Data were collected on baseline characteristics, treatment schedules and choices, and disease-free survival. We recorded comorbid conditions using a slightly adapted version of the Charlson index, which classifies all serious comorbid conditions based on possible prognostic impact into eight groups (i.e. previous malignancies, chronic obstructive pulmonary diseases, cardiovascular diseases, cerebrovascular diseases, hypertension, diabetes mellitus, digestive tract diseases and other).[66] For efficiency reasons, additional detailed information on resource use associated with each treatment and follow-up was collected in a randomly selected subset of 206 patients. These patients were randomly selected, stratified by hospital and oxaliplatin use to ensure equal numbers of patients receiving and not receiving oxaliplatin. The present paper describes the results of the cost analysis based on this subset of patients.

Evaluation of a retrospective database study using anonymised data is until now in the Netherlands not considered an interventional trial according to Directive 2001/20/EC and to the Dutch legislation (WMO). Consequently a request to a medical ethics committee is not required.

Resource use and unit prices

The cost analysis was conducted from a hospital perspective. Resource use data were drawn from individual medical records. Mean costs per patient were calculated for the four most common treatment regimens seen in daily practice: 5-FU/LV, capecitabine, 5-FU/LV with oxaliplatin (FOLFOX) and capecitabine with oxaliplatin (CAPOX). Total costs for individual patients were determined

by the identification of resource use and unit costs of the following cost components: chemotherapy, inpatient hospital days, intensive care days, outpatient visits, consultations by telephone, daycare treatments, emergency room visits, radiotherapy, surgical procedures, laboratory services, medical imaging services, intravenous access devices and concomitant medications.

Resource use was divided into two time periods, a treatment phase (period 1) and a follow-up phase (Period 2). Period 1 began on day 1 of the first administration of adjuvant chemotherapy. To capture resource use resulting from treatment related toxicity, period 1 ended one month after the last administration of chemotherapy. Period 2 started one month after the last administration of chemotherapy and lasted until disease progression (or end of the follow up duration of our study).

Table 3.1 presents the unit costs of inpatient hospital days, intensive care days, outpatient visits, consultations by telephone, daycare treatments and emergency room visits. The unit cost calculations were based on detailed microcosting studies reflecting full hospital costs, including overhead costs.[95, 96] Unit costs of in- and outpatient visits and daycare treatments were weighted according to their type of hospital: 33% of the unit costs were based on data from the university hospitals and 67% on those from general hospitals. These shares reflect the distribution of patients among hospitals in Dutch daily prac-

Table 3.1 Unit costs (Euro 2012)

Oncology inpatient hospital day[a]	€ 520
Intensive care unit day[b]	€ 2,209
Oncology day-care treatment[a]	€ 177
Oncology outpatient visit[a]	€ 104
Consultation by telephone[a]	€ 14
Emergency room visits[c]	€ 185
5-Fluorouracil (mg)[d]	€ 0.01
Leucovorin (mg)[d]	€ 0.29
Capecitabine (mg)[d]	€ 0.01
Oxaliplatin (mg)[d]	€ 4.62
Uracil/tegafur (mg)[d]	€ 0.05

[a]Tan et al., 2010[96]; [b]Tan et al., 2008[95]; [c]Dutch Healthcare Authority, 2009[4]; [d]Pharmacotherapeutical Aid Committee[97]

tice. The unit costs of surgical procedures, laboratory tests and medical imaging tests were based on the fees as issued by the Dutch Healthcare Authority.[4]Unit costs of chemotherapy are shown in Table 1. Unit costs of chemotherapy and concomitant medications were acquired from the Committee Pharmacotherapeutical Aid.[97]

All costs were based on Euro 2012 cost data. Where necessary, costs were adjusted to 2012 using the consumer price index from the Dutch Central Bureau of Statistics.[98]

Statistical Analyses

In addition to descriptive statistics, differences between the four treatment groups were assessed by means of the one-way analysis of variance (ANOVA) test for variables showing a normal distribution, the Kruskal-Wallis test for variables not normally distributed and the Pearson Chi-square test for variable fractions, followed by pair-wise tests with the Mann-Whitney test to identify which groups differed significantly from others. Statistically significant differences were concluded from comparisons with a P-value less than or equal to an alpha of 0.05.

The association between prognostic baseline characteristics (independent variables) and the average total cost of period 1 (dependent variables), controlled for type of chemotherapy regimen, was tested by means of a generalised linear regression model (GLM) with a gamma distribution and a log link function. Only prognostic factors with 10% or fewer missing values were entered into the multivariate model. The final model was chosen based on the likelihood ratio test for goodness-of fit.

All statistical analyses were conducted using the SAS computer package version 9.2 (SAS Institute Inc., Cary, NC, USA, 1999).

RESULTS

Treatments and patients

The four most commonly administered regimens in the 391 patients included in our study were: 5FU/LV (4% of patients), capecitabine (24%), FOLFOX (35%), and CAPOX (37%). An overview of the patient flow is given in Fig. 3.1 CAPOX

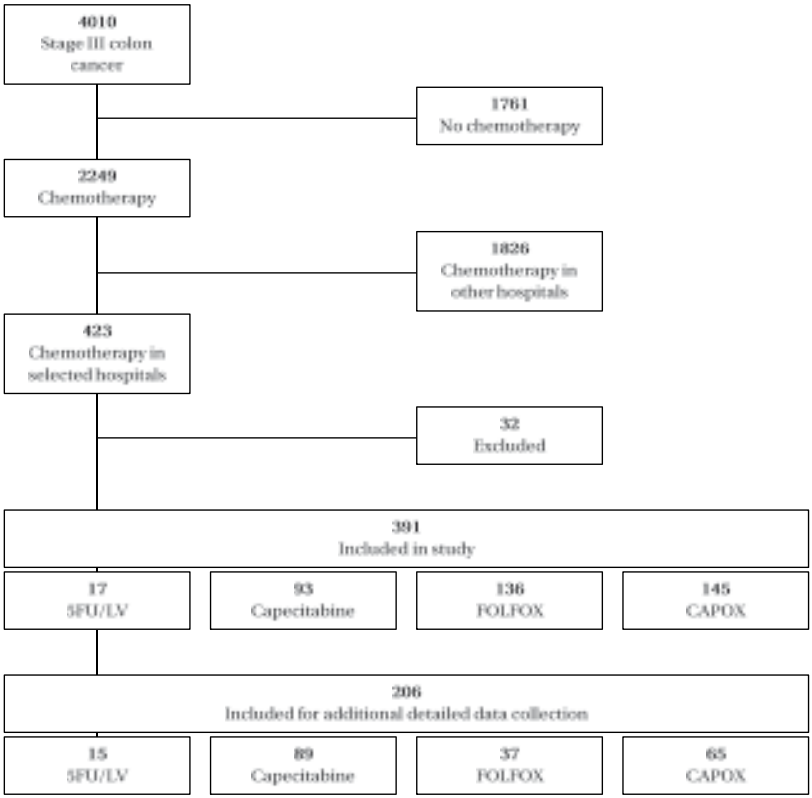


Figure 3.1 Patient flowchart

was relatively often administered in the three university hospitals and FOLFOX in the seven general hospitals. In the nine large teaching hospitals a more mixed pattern could be observed.

In the subset of 206 patients, 15 patients received 5-FU/LV, 89 received capecitabine, 37 received FOLFOX and 65 received CAPOX. The patient characteristics of the total population (n=391) as well as those of the subset of 206 patients stratified by treatment are summarised in Table 3.2 patients receiving oxaliplatin were significantly younger ($P < 0.001$) and had fewer comorbidities ($P = 0.019$) than patients who did not receive oxaliplatin. Patient characteristics such as age and sex were comparable between the different types of treatment centres.

Table 3.2 Patient characteristics at baseline

	Without oxaliplatin		With oxaliplatin		
	5FU/LV	Capecitabine	FOLFOX	CAPOX	
	N = 15	N = 89	N = 37	N = 65	*P-values*[a]
Age - yr					
Median	71	73	60	60	*< 0.001*
Range	41-79	58-85	34-46	22-82	
Age group - no.					
< 70 (%)	6 (40)	29 (33)	31 (84)	54 (83)	*< 0.001*
≥ 70 (%)	9 (60)	60 (67)	6 (16)	11 (17)	
Comorbid conditions - no.					
0 – 1 (%)	11 (73)	67 (75)	31 (84)	59 (91)	*0.019*
2+ (%)	4 (27)	22 (25)	6 (16)	6 (9)	
Sex - no. (%)					
Male (%)	9 (60)	44 (49)	20 (54)	36 (55)	*0.580*
Female (%)	6 (40)	45 (51)	17 (46)	29 (45)	
Depth of invasion - no.					
T2 -T3 (%)	13 (87)	78 (88)	34 (92)	55 (85)	*0.834*
T4 (%)	2 (13)	10 (11)	3 (8)	10 (15)	
Unknown		1			
Nodes involved - no.					
N1 (%)	10 (67)	59 (66)	23 (62)	38 (58)	*0.387*
N2 (%)	5 (33)	30 (34)	14 (38)	27 (42)	
Histologic appearance - no.					
Well differentiated (%)	10 (67)	71 (81)	29 (78)	55 (90)	*0.204*
Poorly differentiated (%)	5 (33)	17 (19)	8 (22)	6 (10)	
Unknown				4	
CEA level - no.					
< 5 ng/ml (ULN) (%)	5 (71)	43 (77)	12 (75)	24 (55)	*0.080*
≥ 5 ng/ml (ULN) (%)	2 (29)	13 (23)	4 (25)	20 (45)	
Unknown	8	33	21	21	

[a] P-values based on comparisons between patients receiving oxaliplatin based regimens versus patients who did not receive oxaliplatin; FOLFOX, 5FU/LV combined with oxaliplatin; CAPOX, capecitabine combined with oxaliplatin; CEA, carcinoembryonic antigen; ULN, upper limit of normal.

Resource use and costs

Period 1: the treatment phase, starting from the first administration of chemotherapy to one month after the last administration

In line with the standard duration of adjuvant therapy of six months in all four chemotherapy regimens, the mean follow up durations for period 1 were 5.3 ± 2.3 months for patients receiving 5-FU/LV, 5.5 ± 1.7 months for capecitabine, 5.8 ± 1.0 months for FOLFOX and 5.9 ± 1.6 months for CAPOX.

Table 3.3 presents the resource use and costs for period 1 of the four treatment groups. Resource use and costs were grouped in: use and costs of chemotherapy, the administration of chemotherapy and use and costs for managing adverse events/monitoring plus other resources.

Mean total costs per patient in period 1 were €6,163 for 5-FU/LV, €5,229 for capecitabine, €27,446 for FOLFOX and €14,783 for CAPOX. Mean total costs for FOLFOX and CAPOX were significantly different ($P < 0.001$), where mean costs for 5-FU/LV and capecitabine were not significantly different ($P = 0.060$).

As reflected by the large standard deviations, substantial variations were found in the total resources used and costs obtained for individual patients within treatment groups as well as in each individual resource component. Chemotherapy drug costs and the administration of chemotherapy (including intravenous access devices, inpatient hospital days, day care treatments and outpatient visits) were the most important cost drivers.

Patients treated with 5FU/LV as monotherapy received substantially less 5FU/LV than FOLFOX-treated patients ($P < 0.001$). Also the total cumulative dose of oxaliplatin in FOLFOX versus CAPOX significantly differed (1,922 mg versus 1,520 mg respectively, $P = 0.0143$). Mean costs for chemotherapy amounted to €504 for 5-FU/LV, €2,328 for capecitabine, €11,443 for FOLFOX and €9,289 for CAPOX. For the FOLFOX treatment group, oxaliplatin alone accounted for 78% of the chemotherapy costs and 32% of the total treatment costs. For the CAPOX group, these proportions equalled 76% and 48%.

As can be seen in Table 3.3, patients treated with FOLFOX were admitted for the administration of chemotherapy for an average of 19.3 (SD 16.3) inpatient days, compared to 0.8 (SD 3.1), 0.0 and 1.7 (SD 5.5) days in the other three treatment groups ($P < 0.001$). Inpatient stay costs related to the administration of chemotherapy were €416 in 5-FU/LV, €0 in capecitabine, €10,029 in FOLFOX

Table 3.3 Resource use and costs in the treatment phase (period 1, Euro 2012)

	Resource use (numbers)				Costs (euro)			
	Without oxaliplatin		With oxaliplatin		Without oxaliplatin		With oxaliplatin	
	5FU/LV	Capecitabine	FOLFOX	CAPOX	5FU/LV	Capecitabine	FOLFOX	CAPOX
	N = 15	N = 89	N = 37	N = 65	N = 15	N = 89	N = 37	N = 65
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Chemotherapy drug (mg)								
5- Fluorouracil	17,058 (8,360)	319 (2,292)	42,207 (15,297)	1,514 (5,789)	105 (51)b	2 (14)	260 (95)b	8 (36)
Leucovorin	1,356 (760)	42 (302)	7,476 (2,665)	340 (1,208)	399 (224)b	13 (89)	2,204 (786)b	100 (356)
Capecitabine	0 (0)	361,400 (134,325)	17,239 (61,605)	335,140 (123,581)	0 (0)	2,303 (856)b	109 (393)	2,136 (787)b
Oxaliplatin	NA (NA)	NA (NA)	1,922 (750)	1,520 (615)	0 (0)	0 (0)	8,870 (3,462)c	7,017 (2,839)c
Uracil/tegafur	0 (0)	180 (1780)	0 (0)	520 (4140)	0 (0)	10 (95)	0 (0)	28 (220)
Administration of chemotherapy								
Intravenous access	0.0 (0.0)	0.0 (0.0)	0.4 (0.5)	0.0 (0.1)	0 (0)	0 (0)	159 (206)	7 (52)
Inpatient hospital days	0.8 (3.1)	0.0 (0.0)	19.3 (16.3)	1.7 (5.5)	416 (1,612)	0 (0)	10,029 (8,487)a	889 (2,884)
Day-care treatments	10.1 (12.6)	0.8 (2.2)	7.0 (12.4)	5.7 (3.7)	1,797 (2,198)b	150 (386)	1,247 (2,196)	1,013 (647)
Outpatient visits	10.9 (13.3)	8.4 (4.1)	10.1 (6.5)	8.1 (6.1)	1,131 (1,386)	876 (426)	1,046 (679)	840 (637)
Adverse events/Monitoring/Other								
Inpatient hospital days	1.3 (3.1)	2.0 (5.9)	1.7 (5.4)	2.8 (7.8)	660 (1,592)	1,019 (3,070)	858 (2,796)	1,457 (4,056)
Intensive care unit days	0.0 (0.0)	0.0 (0.3)	0.0 (0.0)	0.0 (0.0)	0 (0)	65 (607)	0 (0)	0 (0)
Emergency room visits	2.1 (7.7)	0.6 (1.5)	0.1 (0.5)	0.4 (0.9)	394 (1,427)	108 (281)	25 (89)	68 (174)
Consultations by telephone	0.8 (1.4)	1.4 (2.3)	1.0 (3.2)	2.0 (3.2)	11 (20)	19 (32)	14 (44)	27 (44)
Concomitant medication	1.7 (4.3)	1.6 (4.5)	3.6 (9.0)	4.3 (3.3)	151 (379)	107 (298)	1,912 (4,763)a	360 (275)

Table 3.3 Resource use and costs in the treatment phase (period 1, Euro 2012) (table continued)

	Resource use (numbers)				Costs (euro)			
	Without oxaliplatin		With oxaliplatin		Without oxaliplatin		With oxaliplatin	
	5FU/LV	Capecitabine	FOLFOX	CAPOX	5FU/LV	Capecitabine	FOLFOX	CAPOX
	N = 15	N = 89	N = 37	N = 65	N = 15	N = 89	N = 37	N = 65
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Laboratory	15.1 (10.0)	8.3 (4.8)	11.9 (4.9)	11.1 (6.3)	568 (378)	311 (182)	449 (185)	419 (236)
Radiology	0.7 (1.1)	1.1 (1.3)	1.4 (1.3)	1.5 (1.8)	157 (420)	172 (489)	152 (173)	194 (286)
Colonoscopy	0.3 (1.0)	0.1 (0.4)	0.1 (0.3)	0.3 (0.6)	126 (489)	69 (183)	64 (164)	119 (289)
Radiotherapy	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.1)	0 (0)	0 (0)	0 (0)	28 (222)
Surgical procedures	0.1 (0.2)	0.0 (0.1)	0.1 (0.2)	0.0 (0.1)	245 (761)	5 (44)	47 (132)	72 (489)
Total costs								
Mean (SD)					6,163 (3,075)	5,229 (3,356)	27,446 (11,938)bc	14,783 (5,944)c
Median					6623	4433	28873	14673
Minimum					1137	336	2773	1814
Maximum					12881	20426	63891	35655

Due to violation of the normality assumption, the nonparametric Kruskal-Wallis test for two or more related samples was used to compare the four groups. For cost components that differed significantly at $\alpha = 0.05$, pairwise tests by means of the Mann-Whitney test were performed to identify which groups differed. No resource use significance results are presented since these equal the cost P-values.

a: Cost component was significantly higher compared to all other groups in pair-wise comparisons (P-value <0.001).

b: Cost component was significantly higher in pairwise comparisons between the 5FU/LV or FOLFOX and capecitabine or CAPOX groups (within "without oxaliplatin" or within "with oxaliplatin", P-value <0.001).

c: Cost component was significantly higher in pairwise comparisons between the "with oxaliplatin" and the "without oxaliplatin" groups (P-value <0.001).

FOLFOX, 5FU/LV combined with oxaliplatin; CAPOX, capecitabine combined with oxaliplatin

and €889 in CAPOX. Furthermore, 38% of the patients treated with FOLFOX received an intravenous access via a port-a-cath system, resulting in an additional mean cost per patient of €159, compared to maximum €7 in the other treatment groups (P < 0.001).

Day care treatments were of minor importance in the capecitabine treatment group (average costs of day care treatments of €150), since capecitabine is administered orally during outpatient visits. In contrast, costs for day care treatments were much higher in the other three treatment groups (€1,797 for 5-FU/LV, €1,247 for FOLFOX and €1,013 for CAPOX; P < 0.001). A substantial variation was found in number of day care treatments per individual patient (range: 0-46).

The number of outpatient visits did not differ significantly in the four treatment groups (P = 0.177). The proportion of outpatient visits in the total costs related to chemotherapy was 29% in 5-FU/LV, 26% in capecitabine, 4% in FOLFOX and 7% in CAPOX.

Costs not directly related to the administration of chemotherapy, such as cost of managing adverse events and the cost of monitoring were €2,312 for 5FU/LV, €1,875 for capecitabine, €3,521 for FOLFOX and €2,744 for CAPOX. Mean non-chemotherapy-related costs of patients treated with FOLFOX or CAPOX were significantly different from costs in patients treated with 5FU/LV or capecitabine (P < 0.002). The most important driver for this difference was the cost of concomitant medications. Patients receiving FOLFOX and CAPOX often received IV granisetron or ondansetron (anti-emetica) for nausea. Three patients receiving FOLFOX were treated with pegfilgrastim which cost more than €15,000 per patient. Only one of the 206 patients was admitted to the intensive care unit. This patient was treated with capecitabine and developed sepsis during admission for dehydration from diarrhoea.

To further explore the costs of hospitalisations related to the administration of chemotherapy, we compared patients treated with the FOLFOX regimen administered in an outpatient setting with an inpatient setting. In total, 13 FOLFOX-treated patients (35%) never had any hospitalisation for therapy administration, while the other 24 FOLFOX-treated patients (65%) were frequently admitted. Total costs of period 1 for the 13 patients were € 18,868 versus € 32,092 for the 24 patients with frequent admissions. These numbers illustrate the strong impact of hospital admittance on cost-outcomes. The differences

in admission rates may be related to differences in health status. However, it may also be induced by differences in treatment centre policies. To illustrate, we observed that the majority of the 24 frequently admitted patients received treatment in general hospitals.

Period 2: the follow-up phase, starting from one month after the last administration of chemotherapy until disease progression or end of follow-up (data collection)

The mean follow up durations for period 2 were 23.6 ± 10.7 months for patients receiving 5-FU/LV, 20.5 ± 10.6 months for patients receiving capecitabine, 22.2 ± 12.5 months for patients receiving FOLFOX and 17.5 ± 9.8 months for patients receiving CAPOX. Follow-up differed significantly between regimens (p = 0.04), which can be explained by the fact that physicians started using CAPOX later in time than the other regimens.

Table 3.4 presents the resource use and costs for period 2 considering all patients within the four treatment groups (n=206). Mean costs per patient were €3,518 for 5-FU/LV, €4,507 for capecitabine, €5,347 for FOLFOX and €3,578 for CAPOX and did not significantly differ between regimens (P = 0.56). When comparing mean cost per month per regimen, to correct for differences in follow-up duration, mean costs per patients were still not significantly different.

Inpatient hospital days, outpatient visits, radiology and colonoscopies were the most important cost drivers. The costs of these cost drivers did not significantly differ between the four different treatment regimens in period 2.

Regression of prognostic factors on costs

Our GLM regression results show that mean total costs per patient during period 1 were significantly associated with age and comorbidities, controlling for type of chemotherapy regimen (5FU/LV, capecitabine, FOLFOX and CAPOX). Having more than one comorbid condition was associated with a 27% increase in mean total cost of period 1 (P = 0.018). A one-year increase in age was associated with a 1% decrease in costs (P = 0.019).

To further explore these associations, we compared costs and resource use between patients below and above age 70 and costs between patients with or without comorbidities (Table 3.5). Regarding patients below and above age 70, total costs were similar, but some of the cost components showed differ-

Table 3.4 Resource use and costs in the follow-up phase (period 2, Euro 2012)

	Resource use (numbers)				Cost (Euro)			
	Without oxaliplatin		With oxaliplatin		Without oxaliplatin		With oxaliplatin	
	5FU/LV	Capecitabine	FOLFOX	CAPOX	5FU/LV	Capecitabine	FOLFOX	CAPOX
	N = 15	N = 89	N = 37	N = 65	N = 15	N = 89	N = 37	N = 65
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Inpatient hospital days	2.3 (5.1)	2.5 (6.4)	4.4 (10)	1.8 (4.2)	1,214 (2,674)	1,316 (3,364)	2,293 (5,219)	937 (2,208)
Intensive care unit days	0 (0)	0.3 (2.6)	0 (0)	0 (0)	0 (0)	603 (5690)	0 (0)	0 (0)
Outpatient visits	9.2 (10.9)	6.8 (5.4)	8.6 (5.5)	7 (4.8)	964 (1,137)	710 (559)	894 (575)	724 (501)
Day-care treatments	0.3 (1.1)	0.4 (0.9)	0.3 (0.6)	0.4 (0.9)	59 (186)	69 (152)	45 (105)	68 (165)
Consultations by telephone	0.5 (0.9)	0.6 (1.3)	0.6 (1.3)	0.6 (1.4)	7 (13)	7 (17)	8 (17)	8 (19)
Emergency room visits	0.1 (0.5)	0.4 (1)	0.2 (0.5)	0 (0.3)	24 (96)	82 (189)	41 (88)	16 (50)
Laboratory	8.1 (9.2)	10.2 (19.9)	8.1 (6.6)	7.9 (10)	305 (347)	383 (751)	306 (246)	298 (376)
Radiology	4.7 (4.4)	4.6 (3.8)	4.3 (3.5)	4.2 (3.4)	534 (882)	639 (870)	551 (385)	590 (706)
Colonoscopy	0.9 (1.1)	1 (1)	1.5 (3.6)	1.1 (1.3)	410 (532)	452 (467)	691 (1691)	526 (603)
Radiotherapy	0 (0)	0 (0)	0 (0)	0.0 (0.1)	0 (0)	0 (0)	0 (0)	28 (222)
Surgical procedures	0 (0)	0.1 (0.1)	0.2 (0.8)	0.1 (0.4)	0 (0)	246 (883)	517 (1,621)	383 (1,534)
Total costs								
Mean (SD)					3,518 (3,863)	4,507 (8,083)	5,347 (7,511)	3,578 (4,240)
Median					1,926	2,240	2,373	2,317
Minimum					0	0	246	0
Maximum					10,711	66,640	41,049	21,263

None of the differences reached statistical significance

Table 3.5 Treatment cost (period 1) stratified by age and comorbidities (Euro 2012)

	Without oxaliplatin		With oxaliplatin	
	all patients		*all patients*	
	N = 104		N = 102	
Chemotherapy drug	2,065 (1,008)		10,072 (3,710)	
Administration of chemotherapy	1,360 (1,254)		6,280 (6,849)	
Adverse events/Monitoring/Other	1,939 (3,317)		3,024 (4,819)	
Mean total costs	5,364 (3,318)		19,377 (10,518)	
Median	4,604		15,521	
	< 70 years	*≥ 70 years*	*< 70 years*	*≥ 70 years*
	N = 35	N = 69	N = 85	N = 17
Chemotherapy drug	2,122 (1,138)	2,035 (942)	10,362 (3,612)	8,619 (3,966)
Administration of chemotherapy	1,485 (1,440)	1,297 (1,155)	6,158 (6,565)	6,891 (8,325)
Adverse events/Monitoring/Other	2,303 (3,450)	1,755 (3,258)	2,985 (4,850)	3,226 (4,799)
Mean total costs	5,909 (3,252)	5,087 (3,341)	19,504 (10,291)	18,735 (11,907)
Median	5,045	4,378	15,521	16,609
	No comorbidities	*Comorbidities*	*No comorbidities*	*Comorbidities*
	N = 78	N = 26	N = 90	N = 12
Chemotherapy drug	2,018 (1,014)	2,203 (994)	9926 (3769)	11,162 (3,157)
Administration of chemotherapy	1,271 (1,177)	1,625 (1,455)	5880 (6334)	9,278 (9,745)
Adverse events/Monitoring/Other	1,680 (2,884)	2,714 (4,350)	3066 (4897)	2,718 (4,374)
Mean total costs	4,971 (2,806)	6,542 (4,381)	18872 (10225)	23,159 (12,339)
Median	4,603	4,587	15,521	19,126

None of the differences reached statistical significance

ences. Older patients had lower cost of chemotherapy drugs and higher cost of administration of chemotherapy due to a higher frequency of hospital admissions. However, these differences were not significant. Regarding patients with comorbidities we found that patients with 2 or more comorbidities had substantial higher cost of administration of chemotherapy, similar to older patients mainly explained by an increased frequency of hospital admissions. Again, these differences were not significant.

DISCUSSION

Since the generalisability of RCT- based economic evaluations may be limited, there is an increasing focus on real-world effectiveness and cost-effectiveness. This paper estimated the real-world resource use and costs of the adjuvant treatment of stage III colon cancer.

We found that the use of oxaliplatin in the adjuvant treatment of stage III colon cancer leads to a substantial increase in treatment costs compared to treatments with fluoropyrimidines alone.

The combination of 5-FU/LV and oxaliplatin was observed to be the most expensive treatment option. Mean total costs of FOLFOX in periods 1 and 2 were €23,112 more expensive than 5-FU/LV without oxaliplatin, predominantly owing to the cost of oxaliplatin and the inpatient administration of chemotherapy. Similar to our real-world findings, earlier economic evaluations based on the MOSAIC trial have also demonstrated favourable costs for 5FU/LV over FOLFOX, although the size of the differences varied between studies. Reported differences between 5FU/LV and FOLFOX were respectively €10,779, €13,370, €11,041, €4,969, and €6,763 in a Canadian, US, Japanese and two UK studies (country specific currencies converted to Euro 2012).[82, 83, 88, 99, 100] These cost differences are not directly comparable to the €23,112 we found in our study, since our time horizon consisted of approximately six months, whereas these studies applied a life-time horizon. It may therefore be that over longer time periods the differences in costs decrease.

The combination of capecitabine and oxaliplatin was €14,432 less expensive than the costs related to FOLFOX, mainly due to the higher number of patients with FOLFOX being admitted for the administration of chemotherapy than with CAPOX. A Greek RCT-based cost analysis also found substantial cost differences between treatment with CAPOX and FOLFOX.[87] Similar to our study, the higher costs for FOLFOX were almost entirely explained by higher hospitalisation costs. However, the number of hospitalisations in this trial was substantially less than we found in our Dutch real-world study. Their FOLFOX group was hospitalised for an average of 10.7 inpatient days (versus 20.3 in our study), while the CAPOX group was hospitalised for an average of 2.2 inpatient days (versus 4.3 in our study).[87] One possible explanation for the difference in the FOLFOX group can be found in the duration of the intravenous administra-

tion of the FOLFOX regimen, which takes 48 hours. In our study this often led to a 3-day hospital admission per cycle, whereas this might have been 2 days in the Greek study. In the MOSAIC-based economic evaluation of Pandor et. al., the authors reported that the administration of oxaliplatin and 5FU/LV took place via day-ward visits without requiring any inpatient hospital admissions.[88] The patients in the MOSAIC trial all had infusional devices that allowed the administration in an outpatient setting.

We found that the choice to administer the 48 hours administration of FOLFOX as inpatients or outpatients had a great impact on costs. Among patients receiving FOLFOX, the costs of the treatment phase were € 18,868 for patients in an outpatient setting, versus € 32,092 for patients who were hospitalised. Given the fact that it has been demonstrated that the FOLFOX regimen may be safely administered in an outpatient setting, it is questionable whether expensive hospitalisations are necessary for its administration. In this light it should be noted that there may be a financial incentive (reimbursement) for Dutch hospitals to hospitalise patients. If the actual costs of these hospitalisation days are as high as the costs of an average hospitalisation day, which reflects the unit costs used in this paper, potential cost savings up to € 12,480 per patient may be achievable if all FOLFOX patients would receive their treatment in an outpatient setting.

In our study, the costs of CAPOX treatment were estimated at € 14,783. Given the observation that the costs of FOLFOX treatment, even without hospitalisations, are € 18,868, CAPOX remains considerably cheaper than FOLFOX. This means that from an economic viewpoint, treatment with CAPOX is preferred over FOLFOX. Moreover, patients may also prefer CAPOX based on the fact that capecitabine is an oral drug while 5FU/LV is administered intravenously over a 48-hour period. Regarding efficacy, CAPOX regimens are considered to have at least comparable efficacy to FOLFOX regimens, with a HR for overall survival of 0.94 (95% confidence interval 0.89–1.00; P = 0.0489) in a meta-analysis of 67171 patients with both stage III and distant metastatic CRC.[101] However, the toxicity profile differs between these regimens, with CAPOX resulting in a higher incidence of grade 3/4 thrombocytopenia and diarrhea and grade 2/3 handfoot syndrome, and FOLFOX in a higher incidence of neutropenia.[74] The published effectiveness results of present real-world study also implied that FOLFOX and CAPOX are comparable.[92] Combining current clinical and economic evidence therefore may shift the balance even more in favour of CAPOX.

Regarding the comparison of 5FU/LV and capecitabine monotherapies, we also found that the cost of drug administration was an influencing factor. Although drug cost of capecitabine were more expensive then 5FU/LV, their total treatment cost was still comparable due to the higher administration costs of 5FU/LV. This finding is also seen in trial-based cost analyses.[56, 84-86, 88] These results imply that from a real-world economic perspective there is no preference in either administering capecitabine or 5FU/LV as a monotherapy. Since Dutch clinicians prefer capecitabine over 5FU/LV[92] and since capecitabine has a more patient friendly mode of administration, there seems little support for 5FU/LV.[102, 103]

Our results are in line with those of other European studies carried out in predominantly metastatic colorectal patients.[94, 104-106] Comparable to our study, an Italian study on adjuvant colon cancer patients also reported savings of approximately € 1400 per patient in using CAPOX instead of FOLFOX. In this study these saving were similarly explained by the higher costs of administration of FOLFOX.[94] In a Chinese study these savings were €2277 (country specific currencies converted to Euro 2012). In this study, the savings were explained by (again) cost of administration, but also by fewer costs related to adverse events in CAPOX versus FOLFOX.[93] A retrospective study carried out in France estimated a saving of € 2,000 - € 7,200 when using capecitabine instead of 5FU/LV.[105] An English report estimated a saving per patient for capecitabine compared with a modified de Gramont regimen carried out in hospital of € 5,100.[106]

We believe that the detailed insights about resource use and costs of expensive treatments gained in our study are valuable in several ways. Firstly, this study provides insights regarding potential cost-savings by modifying the adjuvant treatment of colon cancer. This can stimulate physicians to carefully consider the necessity of hospital admissions for the administration of FOLFOX before making choices that may put an unnecessary burden on the Dutch health care system.

Secondly, the results of this study were directly usable as input in a subsequent study estimating the real-world cost-effectiveness of oxaliplatin in the adjuvant treatment of stage III colon cancer.[107] Using real-world data in cost-effectiveness models can result in better evidence for Dutch policy makers by addressing uncertainty in outcomes arising from the gap between clinical trials and everyday practice. ICERs from this real-world cost-effectiveness model

ranged from €8,388 to €12,746 in different scenarios. They were all considered acceptable and support the use of oxaliplatin in the adjuvant treatment of stage III colon cancer.

Thirdly, the findings of this study could be used in the development of clinical guidelines. The awareness of clinicians regarding the importance of making cost conscious choices is growing. Moreover, cost (-effectiveness) is starting to play a formal role in health care decision making. For instance, in the Netherlands, health economists are increasingly asked for their advice during the process of guideline development.[108]

A limitation of our study is that we included a relatively low number of patients. As a result, it was difficult to make sufficiently powered comparisons between subgroups. Some of our differences (e.g. the difference in oxaliplatin doses between the younger and the older patients) may have been statistically nonsignificant due to the small sample size. We also did not adjust for multiple testing in our study. However, our statistical inferences were mostly based on P-values lower than 0.0001.

Moreover, it should be noted that caution is needed when using real-world resource use for valid cost comparisons between treatment groups, because the baseline (cost) prognosis of the treatment groups may not be comparable. For example, in our study population, patients receiving oxaliplatin were significantly younger and had fewer comorbidities than patients not receiving oxaliplatin. Clearly, the medical specialist did not randomly assign the different treatment options. The resulting imbalances regarding prognostic factors may have led to invalid cost comparisons. In the regression analysis we indeed found that older patients incurred lower costs and that the presence of more than 1 comorbid condition was associated with increased costs of period 1. These two effects actually work in opposite directions since older patients had more comorbidities. This probably explains the relatively small differences in costs found when the cost comparison was stratified by either age or comorbidities alone.

Conclusion

This study demonstrated that drug costs and the costs related to the number of hospitalisation days needed to administer the chemotherapy were the main cost drivers of the total costs of adjuvant treatment in colon cancer. We

identified a potential for substantial cost savings when the 48 hour during administration of 5FU/LV in the FOLFOX regimen would be administered in an outpatient setting or would be replaced by the oral drug capecitabine like in the CAPOX regimen. This analysis based on detailed real-life data clearly indicates that clinical choices made in oncology based on efficacy of therapy can have economic consequences and that determining the appropriateness of treatment (within the context of finite healthcare) should take this aspect into account after considering the efficacy and safety of the different treatment options.

Chapter 4

Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: A synthesis of clinical trial and daily practice evidence

Pharmacoeconomics 2013 Aug; 31(8): 703-18

Chantal van Gils
Saskia de Groot
Ken Redekop
Miriam Koopman
Kees Punt
Carin Uyl- de Groot

SUMMARY

Objectives**:** Previous cost-effectiveness analyses of oxaliplatin have been based on randomised trials whereas current Dutch policy requires evidence from daily practice. The objective of this study was to examine the real-world cost-effectiveness of oxaliplatin plus fluoropyrimidines (FL) versus FL only, as adjuvant treatment of stage III colon cancer. Methods**:** A Markov model was developed to estimate lifetime cost and quality-adjusted life-years (QALYs) from a hospital perspective. The effectiveness of the oxaliplatin arm was modelled by combining published efficacy data from the pivotal clinical registration trial (MOSAIC trial) with real-world (RW) data from a Dutch population- based observational study. RW patients were categorized into "eligible" or "ineligible", depending on whether the patients fulfilled the MOSAIC trial eligibility criteria. Ineligible RW patients (18%) had a poorer prognosis than eligible RW patients (82%) and MOSAIC trial patients. The effectiveness of the comparator was modelled using MOSAIC trial results. All cost inputs were based on RW patients and reported in Euro 2012. Cost-effectiveness analyses were performed for four different scenarios: (1) cost-effectiveness analyses based on MOSAIC trial patients; (2) cost-effectiveness analyses using MOSAIC and eligible RW patients; (3) cost-effectiveness analyses using MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had an equal effect in ineligible and eligible patients;(4) cost-effectiveness analyses using MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had no effect amongst ineligibles. For each scenario, univariate and probabilistic sensitivity analyses were undertaken. Results: MOSAIC trial patients and eligible RW patients treated with oxaliplatin had comparable two-year disease-free survivals (79.5% vs. 78.4%). Oxaliplatin showed an incremental QALY gain of 1.02, 1.13, 1.17 and 0.93, and incremental cost of €9,961, €11,055, €9,814 and €11,854 in scenarios 1 to 4, respectively. The corresponding incremental cost-effectiveness ratios (ICERs) were €9,766, €9,783, €8,388 and €12,746 in scenarios 1 to 4, respectively. In all scenarios, univariate and probabilistic sensitivity analyses indicated that the ICERs are acceptable and robust under a wide range of model assumptions. Conclusions**:** The ICERS of the different scenarios that resulted from combining MOSAIC trial data with data from Dutch daily practice, all suggest that FL + oxaliplatin is cost-effective versus FL alone in the adjuvant treatment of colon

cancer. This chapter illustrates how one could design and implement a real-world cost-effectiveness study to yield internally valid results that could also be generalisable.

INTRODUCTION

Colon and rectal cancers are among the most common types of cancer in the western world with 215,000 deaths in Europe in 2012.[77] Nearly half of the patients who undergo curative surgery will ultimately relapse and die of metastatic disease.[46] During the 1990s the survival rates of patients with stage III colon cancer significantly improved by the introduction of adjuvant chemotherapy with intravenous fluoropyrimidine, 5-fluorouracil, and leucovorin (5FU/LV).[49, 52] In 2004, a large phase III clinical trial in colon cancer (Multicenter International Study of Oxaliplatin/5-Fluoroucacil/Leucovorin in the Adjuvant Treatment of Colon Cancer [MOSAIC]) demonstrated that the addition of oxaliplatin to 5FU/LV significantly improved disease-free survival compared to 5FU/LV alone. As a result, 6 months of oxaliplatin combined with 5FU/LV (FOLFOX) became the standard adjuvant treatment in the Netherlands for stage III colon cancer patients as of early 2005.[53] Treatment with 5FU/LV only remained indicated for patients who were not eligible or refused treatment with oxaliplatin. At that time the oral fluoropyrimidine capecitabine, alone or in combination with oxaliplatin (CAPOX), also became available as an alternative to 5FU/LV or FOLFOX, as these treatments were found to be equally effective in the treatment of stage III colon or metastatic colorectal cancer.[54, 55, 57, 59]

Next to oxaliplatin, many other expensive drugs have become available the past decade posing a substantial financial burden on the health care sector. In striking an optimal balance between ensuring timely access of these medicines and having sufficient evidence regarding its effectiveness and cost-effectiveness, Dutch policy regulations for expensive inpatient drugs have been implemented since 2006.[5, 109] This policy enables new expensive medicine to be conditionally reimbursed for a period of four years without restriction. However, one important condition is that additional evidence, including evidence from Dutch daily practice, is required to assess appropriate use and to estimate the cost-effectiveness of the expensive drug in daily practice

(Real-world cost-effectiveness). The decision about whether or not to continue funding beyond the first four years will be mainly based on the real-world (RW) cost-effectiveness.

Currently, little is known about the use of real-world (RW) data in cost-effectiveness estimations of expensive medicines. For example, previous cost-effectiveness analyses of oxaliplatin have been solely based on randomised trials.[82, 83, 86, 99, 100] These results can substantially differ from RW cost-effectiveness results because differences between RCTs and daily practice may exist in patient selection criteria, dosing regiments, the use of supportive care, and the intensity of follow-up.[6]

The aim of this study was to estimate the RW cost-effectiveness of fluoropyrimidines (FL) plus oxaliplatin versus FL only, as adjuvant treatment of stage III colon cancer. These estimates were based on a combination of published MOSAIC trial data with data from a Dutch population-based observational study.[53, 92]

METHODS

Data sources

Two data sources were used to assess the RW cost-effectiveness of adding oxaliplatin to FL in patients with stage III colon cancer; the MOSAIC trial and a Dutch population-based observational study.[53, 64, 92, 110]

The MOSAIC trial was a multicentre international randomised controlled trial enrolling 2246 patients with both stage II and III colon cancer. We examined the subgroup of 1347 stage III patients of which 672 were randomised to adjuvant treatment with FL + oxaliplatin, and 675 to FL only. Since completion of the study, three-year disease-free survival (DFS) as well as final results of this study including 6-year overall survival (OS) and 5-year updated DFS have been published.[53, 64]

The Dutch population-based observational study involved a retrospective analysis of the treatment outcomes of 391 stage III colon cancer patients diagnosed in 2005 and 2006. The study population was created by selecting a representative sample of 19 Dutch hospitals and then gathering RW data from the Dutch Cancer Registry and medical records of all patients treated in these

Table 4.1 Comparison of the MOSAIC trial and the Dutch observational study

	MOSAIC trial		RW observational study			
	FL + oxaliplatin	FL alone	FL + oxaliplatin		FL alone	
	n = 672	n = 675	n = 281		n = 110	
Treatments	FOLFOX	5FU-LV	FOLFOX or CAPOX		5FU/LV or capecitabine	
Planned cumulative dose - mg/m^2						
Oxaliplatin	1,020	.[a]	1,020		.[a]	
5FU/LV	12,000/2,400	12,000/2,400	12,000/2,400		12,000/2,400	
Capecitabine	.[a]	.[a]	224,000		280,000	
Planned dose given - %	84%	81%	81% (FOLFOX)		72% (5FU/LV)	
			71% (CAPOX)		83% (capecitabine)	
Eligibility for MOSAIC trial						
Eligible - % (no. of patients)	100% (672)	100% (675)	82% (200)		63% (54)	
Ineligible - % (no. of patients)	0% (0)	0% (0)	18% (43)		37% (32)	
Unknown (no. of patients)			(38)[b]		(24)[b]	
Fulfilling eligibility criterium - %			Eligible	Ineligible	Eligible	Ineligible
Chemotherapy within 7 wks after surgery	100%	100%	100%	30%	100%	50%
Age < 75	100%	100%	100%	93%	100%	31%
Good performance score (WHO 0-1)	100%	100%	.[c]	.[c]	.[c]	.[c]
Low tumour marker (CEA level < 10 ng/ml)	100%	100%	100%	70%	100%	94%
Baseline characteristics						
Age - median	61	60	61	62	71	77
range	(19-75)	(20-75)	(22-75)	(32-82)	(55-75)	(59-85)

Table 4.1 Comparison of the MOSAIC trial and the Dutch observational study (table continued)

	MOSAIC trial		RW observational study			
	FL + oxaliplatin	FL alone	FL + oxaliplatin		FL alone	
	n = 672	n = 675	n = 281		n = 110	
No. of comorbid conditions - % 0-1 (versus 2+)	.[d]	.[d]	93.0%	69.8%	72.2%	75.0%
Gender - % male (versus female)	56.1%	52.4%	53.5%	51.2%	53.7%	46.9%
Depth of invasion - % T4 (versus T2-T3)	19.1%	17.9%	13.0%	25.6%	7.5%	15.6%
No. Of nodes involved - % 1-4 (versus >4)	74.2%	76.0%	70.5%	76.7%	72.2%	68.8%
Tumour differentiation - % well (versus poorly)	81.7%	79.4%	92.6%	85.0%	77.4%	78.1%
CEA level - % < 5ng/ml (versus ≥ 5 ng/ml)	.[d]	.[d]	86.5%	60.5%	92.6%	90.6%
Clinical effects						
DFS after 2 years - %	79.5%	72.5%	78.4%	56.7%	82.8%	83.7%
(Lower limit 95% CI)	.[d]	.[d]	(72.5%)	(41.4%)	(72.5%)	(70.5%)
(Upper limit 95% CI)	.[d]	.[d]	(84.3%)	(72.0%)	(93.0%)	(96.8%)
DFS after 5 years - %	66.4%	58.9%	.[c]	.[c]	.[c]	.[c]
Hazard ratio (95% CI)	0.78 (0.65 - 0.93)		.[c]		.[c]	
OS after 2 years - %	90.0%	88.0%	(92.1%)	(74.7%)	(94.0%)	(83.7%)
(Lower limit 95% CI)	.[d]	.[d]	(88.3%)	(61.0%)	(87.6%)	(70.5%)
(Upper limit 95% CI)	.[d]	.[d]	(96.0%)	(88.4%)	(100%)	(96.8%)
OS after 6 years - %	72.9%	68.7%	.[c]	.[c]	.[c]	.[c]
Hazard ratio (95% CI)	0.80 (0.65 - 0.97)		.[c]		.[c]	

RW, real world; CEA, carcinoembryonic antigen level; DFS, disease-free survival; OS, overall survival; CI, confidence interval; FL, fluoropyrimidines; FU, fluorouracil; LV, leucovorin; FOLFOX, 5FU/LV + oxaliplatin; CAPOX, capecitabine + oxaliplatin

[a] Not applicable; [b] Due to missing CEA values; [c] Not measured; [d] Not reported;

19 hospitals. Detailed methods of this study and its results regarding patient characteristics, chemotherapy use and 2-year DFS have been published.[92, 110]

In order to determine how to combine these two sources of evidence in a model, we first compared the two studies on the following subjects: 1) treatments, 2) patients, and 3) disease-free and overall survival outcomes. An overview is provided in Table 4.1.

Treatments

In the MOSAIC trial, patients were randomised to receive FL only (n = 675) or FL + oxaliplatin (n = 672). Although treatment choice was not fixed in the observational study, we found that also RW patients received FL only or FL + oxaliplatin. Since treatment allocation was not randomised in these RW patients, the decision whether a patient was treated with FL only or FL + oxaliplatin was left up to the judgement of the physicians. Specific reasons for not prescribing oxaliplatin were not often retrievable from patient records. In RW patients, FL + oxaliplatin was prescribed more often than FL only (n = 281 versus n = 110, respectively).

Regarding the specific FL regimen used, all MOSAIC patients received 5FU/LV or FOLFOX. In RW practice more variation was seen. For FL only treatment, oral capecitabine was most frequently used, with few patients receiving 5FU/LV. For FL + oxaliplatin treatment, both FOLFOX (48%) and CAPOX (52%) regimens were prescribed in RW oxaliplatin patients. Dosages given in daily practice corresponded well to those seen in the MOSAIC trial.[92]

Patients

The target population of both studies consisted of patients diagnosed with stage III colon cancer who underwent curative surgery. In contrast to the Dutch observational study where no additional entrance criteria were used, the MOSAIC trial only enrolled patients who fulfilled certain eligibility criteria: patients had to start the treatment within seven weeks after surgery, be younger than 75 years, and have a good performance score and low carcinoembryonic antigen (CEA) value (Table 4.1). We applied these eligibility criteria to the RW population to allow better comparisons between the MOSAIC trial and the RW study. However, we were unable to compare performance score since no it is not routinely collected in daily practice.

Of the 281 patients who received FL + oxaliplatin, 200 met the MOSAIC eligibility criteria (eligible RW oxaliplatin patients) and 43 did not (ineligible RW oxaliplatin patients). Of the 110 patients who received FL only, 54 met the eligibility criteria (eligible RW FL only patients) and 32 did not (ineligible RW FL only patients). A total of 62 patients could not be classified due to missing carcinoembryonic antigen (CEA) values and they were excluded from further analysis. In total, 67% of these missing values were found in just 4 hospitals, where CEA levels were not routinely measured. The most important reasons for ineligibility in the FL + oxaliplatin group were chemotherapy not starting within 7 weeks and high CEA levels. In the FL only group the most important reason was age above 75 years.

The differences between RW FL+ oxaliplatin and RW FL-only patients regarding reasons for MOSAIC trial ineligibility resulted from the non-randomised assignment of the treatments in the Dutch observational study. Physicians in daily practice appeared to be reluctant to prescribe oxaliplatin to older patients. As a consequence, the baseline characteristics of RW patients receiving FL + oxaliplatin were different from those of RW patients receiving FL only. Differences amongst the RW eligible patients were also found: eligible RW oxaliplatin patients were significantly younger and had fewer comorbidities than eligible RW FL only patients.[92] In contrast, eligible RW oxaliplatin patients had the same median age as the patients included in the MOSAIC trial (both 61 years) and also ineligible RW oxaliplatin patients showed a comparable median age (62 years), indicating that physicians probably considered comparable kind of patients for inclusion in the MOSAIC trial and adjuvant treatment with oxaliplatin in RW practice. Regarding other baseline characteristics, such as age, depth of invasion, number of involved nodes and tumour differentiation, which were considered not related to treatment assignment, we did notice some baseline differences (Table 4.1). However, given their relatively mild prognostic value in differing directions, we expected a comparable baseline prognosis of the MOSAIC trial patients and RW eligible oxaliplatin patients. However, RW ineligible oxaliplatin patients were expected to have a worse baseline prognosis because of their significantly higher CEA values, which is an important unfavourable prognostic factor.[111] Similarly, we expected patients receiving FL only (both RW eligible and ineligible FL only patients) to have a more favourable baseline

prognosis than MOSAIC trial and RW eligible oxaliplatin patients since they had significantly lower percentages of elevated CEA levels.

In conclusion, MOSAIC trial patients and eligible RW patients receiving FL + oxaliplatin were expected to have a comparable baseline prognosis while all other RW patients were expected to have a different baseline prognosis.

Disease-free and overall survival outcomes

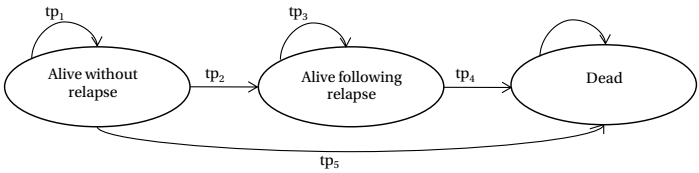
The MOSAIC trial patients randomised in the FL + oxaliplatin arm had a significantly improved 5-year DFS compared to the patients receiving only FL (HR 0.78, p = 0.005, Table 4.1). In RW patients, a comparison of these treatments could not directly be made due to incomparability of the RW FL only and RW FL + oxaliplatin groups caused by the reluctance of physicians to use FL + oxaliplatin in older patients with more comorbidity (see previous section). In the RW study, the FL only patients (both eligible and ineligible) had an even better 2-year DFS than all other groups, which can probably be largely explained by the significantly lower percentage of patients with elevated CEA levels (table 4.1). In an attempt to correct for this incomparability when estimating the 2-year DFS of FL + oxaliplatin versus FL in the RW eligible patients, different adjustment methods were applied to the Cox multivariate regression model, such as average covariate adjustment, regression adjustment on propensity scores and propensity score matched survival analysis. However, these models failed to produce meaningful results since the sample size was not powered for this purpose. This, in combination with missing data on potentially other unmeasured confounding variables, resulted in possibly biased estimates with wide confidence intervals.[110]

Regarding the comparison of MOSAIC trial patients receiving FL + oxaliplatin and eligible RW patients receiving FL + oxaliplatin, comparisons were considered justified and showed comparable 2-year DFS results of 79.5% and 78.4% respectively (p = 0.32). As expected, the 2-year DFS was significantly worse in ineligible RW patients receiving FL + oxaliplatin (56.7%, p <0.01).[92] This can probably be largely explained by the significantly higher CEA levels that were present in these ineligible RW oxaliplatin patients compared to eligible RW oxaliplatin patients. Conclusions regarding the 2-year OS were comparable with those for DFS.

DFS results from 2-5 years and OS results from 2-6 years were available for the MOSAIC trial population (Table 4.1).

Model Structure

A Markov model was developed to estimate the clinical and cost consequences of oxaliplatin over the remaining lifetime of the patients. The model simulated the transition of patients receiving adjuvant chemotherapy for stage III colon cancer through three health states that are typically observed in a clinical setting: alive without relapse, alive following relapse, and dead. For all scenarios, the primary health economic outcome was the marginal cost per quality adjusted life-year (QALY) gained for oxaliplatin combined with FL versus FL alone. Figure 4.1 provides an overview of the health economic model. A similar model was used by Pandor et al. 2006.[88]



tp_1 = transition probability of staying in the disease-free state: estimated by the MOSAIC trial and outcomes research
tp_2 = transition probability of moving from the disease-free state to the relapse state: estimated by subtracting tp_5 and tp_1 from 1
tp_3 = transition probability of staying in the relapse state: estimated by subtracting tp_4 from 1
tp_4 = transition probability of moving from the relapse state to the dead state: estimated by the MOSAIC trial and from mortality rates from the general Dutch population
tp_5 = transition probability of death amongst patients without relapse: estimated by the MOSAIC trial and mortality rates from the general Dutch population

Figure 4.1 Schematic of the health economic model

Transitions between health states were derived from (disease-free) survival curves using a 3 month cycle length, which was considered short enough to avoid multiple transitions between health states within a single cycle. The assumptions which underpin the health economic analysis are presented in Table 4.2. Future cost and benefits were discounted at a rate of 4% and 1.5%,

Table 4.2 Model assumptions

- All relapses were assumed to occur within five years following resection of the primary tumour. Clinical evidence from long-term follow-up of patients undergoing adjuvant chemotherapy supports this assumption.[51]
- All deaths due to colon cancer were assumed to occur within seven years following resection of the primary tumour This assumption follows the assumption of unlikely relapses after five years and a limited life expectancy after relapse.[118]
- The patients in the model were assumed to have the same background mortality as an age-matched population of Dutch individuals with no history of colon cancer
- Follow-up duration was assumed to last for five years and the 3-monthly costs for monitoring visits and diagnostic tests after the first two years were assumed to be € 147. Both assumptions are consistent with Dutch guidelines.[117]
- Patients who relapsed were assumed to receive fluoropyrimidines, oxaliplatin and/or irinotecan in different treatment lines as found in a Dutch observational study involving metastatic patients.[110, 119]

respectively, consistent with current Dutch guidelines.[112]The Markov model was validated by comparing the model results to the results observed in the MOSAIC trial and in RW patients. The Markov model was developed using decision analysis software (TreeAgePro 2009 5Suite, release 1.2, TreeAge software, Inc, Williamstown, MA).

Model scenarios

Cost-effectiveness analyses (CEAs) were performed for three different scenarios: (1) CEA based on MOSAIC trial patients only; (2) CEA using MOSAIC trial patients and eligible RW patients; (3) CEA using MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had an equal effect in ineligible and eligible patients; and (4) CEA using MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had no effect amongst ineligibles (Table 4.3).

Model Input – Clinical Effectiveness

The transition probabilities related to relapse and death are time-dependent variables in which transitions take place every three months. Patients in the alive without relapse state may stay in that phase (tp1), relapse (tp2), or die (tp5). Patients in the alive following relapse state may stay in that phase (tp3), or die (tp4).

Table 4.3 Description of the scenarios

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
FL + oxaliplatin	MOSAIC patients	MOSAIC and eligible RW patients	MOSAIC and eligible RW patients	MOSAIC and eligible RW patients
	n = 672	n = 672 + 200	n = 672 + 200	n = 672 + 200
			ineligible RW patients	ineligible RW patients
			n = 43	n = 43
Transition probabilities				
0-2 years	MOSAIC	MOSAIC + eligible RW	82% MOSAIC + eligible RW	82% MOSAIC + eligible RW
			18% ineligible RW	18% ineligible RW
beyond 2 years	MOSAIC	MOSAIC	82% MOSAIC	82% MOSAIC
			18% MOSAIC, multiplied by three	18% MOSAIC, multiplied by three
FL	MOSAIC patients	MOSAIC patients	MOSAIC patients	MOSAIC patients
	n = 675	n = 675	n= 675	n= 675
Transition probabilities				
0-2 years	MOSAIC	MOSAIC	82% MOSAIC	82% MOSAIC
			18% MOSAIC, multiplied by HR	18% MOSAIC
beyond 2 years	MOSAIC	MOSAIC	82% MOSAIC	82% MOSAIC
			18% MOSAIC multiplied by HR	18% MOSAIC
Assumptions			Addition of oxaliplatin has same benificial effect in ineligible patients	Addition of oxaliplatin has no benificial effect in ineligible patients

RW, real world; FL, fluoropyrimidines; tp, transition probabilities

Scenario 1

In scenario 1, tp1 was solely based on the published MOSAIC 5-year DFS Kaplan Meier (KM) curves of both treatment groups.[64] The cumulative DFS probabilities were read from the published KM Figure and converted to transition probabilities using the following formula: pt = Pt / Pt-1, where Pt and Pt-1 denote the cumulative probability of surviving at the end of times t and t-1, respectively; pt denotes the transition probability for time t. After 5 years, the probability of relapse (tp2) was assumed to be zero. During the MOSAIC trial (up to four years of follow-up), 28 of the 1347 (2%) randomised patients died without having

relapsed.[83] These deaths were equally distributed among both treatment groups and thus seemed independent from prescription of oxaliplatin.[53] We used this percentage to calculate tp5 over the first four years. Beyond this period, we assumed that the patients in our model have the same background mortality as an age-matched population of Dutch individuals with no history of colon cancer. Dutch vital statistics were used to calculate tp5 after four years.[98]

Mortality rates amongst patients in the relapsed state (tp4) were based on published MOSAIC overall survival KM curves, which showed the probability of surviving up to 7 years after randomisation.[64] However, the probability of dying without relapse was also included in the published KM Figure. In order to arrive at tp4, we subtracted tp5 from the OS curves. We assumed tp4 after 7 years to equal background mortality (tp5).

Scenario 2

In scenario 2 the FL alone arm was modelled as in scenario 1. In the FL + oxaliplatin arm, results from eligible RW patients receiving FL + oxaliplatin were added. Transition probabilities in the FL + oxaliplatin arm were based on both the MOSAIC trial (see scenario 1) and the 2-year DFS and OS KM curves of eligible RW patients. At each time point, the transition probabilities from the MOSAIC trial and RW patients were combined via meta-analysis. For all transition probabilities, Q statistics were negative and a fixed effect model could be used. Inverse variance weighted pooled transition probabilities and pooled standard errors were calculated accordingly.[113] Since in RW patients, only 2-year DFS and OS KM curves were available, transition probabilities beyond 2 years were taken from the MOSAIC trial.

Scenario 3 and 4

Scenario 3 and 4 incorporated ineligible patients who comprised 18% of the RW oxaliplatin patients. To reflect Dutch RW clinical practice of patients receiving FL + oxaliplatin, 82% of the cohort was modelled as if it was eligible, and 18% as if it was ineligible for the MOSAIC trial. The eligible cohort was modelled according to scenario 2.

For ineligible patients receiving FL + oxaliplatin, transition probabilities up to 2 years were derived from the 2-year DFS and OS Kaplan Meier curves derived from RW ineligible oxaliplatin patients. Like in scenario 2, all transition

probabilities beyond 2 years were derived from the MOSAIC trial. However, ineligible patients had a worse prognosis than the eligible MOSAIC trial patients, so transition probabilities beyond 2 years could not be taken directly from the MOSAIC trial. We therefore used the following two-step approach. We first adjusted the 3-monthly transition probabilities to match the worse prognosis of the ineligible RW oxaliplatin patient population. Comparisons of the 2-year DFS curves of eligible and ineligible RW oxaliplatin patients revealed that the 3-monthly transition probability of relapse was on average three times higher amongst the ineligible patients than it was amongst the eligible patients. Since this ratio was relatively constant over the 2-year period for which data was available, we assumed that this ratio would remain at this level beyond 2 years and therefore multiplied the transition probabilities taken from the MOSAIC trial (tp2) by three. Regarding mortality amongst patients in the relapsed state (tp4), there was no difference between eligible and ineligible RW oxaliplatin patients.

In scenarios 3 and 4, the FL alone arm was modelled as in scenario 1 and 2. However, given the worse prognosis of the 18% ineligible RW oxaliplatin patients, transition probabilities for the ineligible FL only patients were estimated using different assumptions regarding the position of the DFS and OS KM curves for these patients in two scenarios (3 and 4).

In scenario 3 we assumed that the addition of oxaliplatin to FL would be as beneficial in ineligible patients as in eligible patients. In the MOSAIC trial, the hazard ratios of DFS and OS were 0.78 and 0.80, respectively, for both FL only and FL + oxaliplatin patients (Table 4.1). To obtain transition probabilities for an ineligible FL only arm, we applied these hazard ratios to the transition probabilities of the ineligible FL + oxaliplatin group.

In scenario 4 we assumed that the addition of oxaliplatin to FL would not have any added benefit in ineligible patients. The transition probabilities for the ineligible FL only cohort were made equal to the transition probabilities of the ineligible RW patients treated with FL + oxaliplatin.

Model Input – Health State Utilities

Health utility scores were not collected in either MOSAIC or RW patients. Modelled survival estimates were adjusted to account for the patients' health related quality of life using published colorectal cancer utility estimates. Like

in the analysis of Eggington et al., a utility score of 0.7 was assigned to patients receiving adjuvant treatment who experienced no significant adverse events, while patients who suffered significant adverse events were assigned a utility score of 0.63 for the duration of the treatment course.[86, 114] The percentage of patients experiencing significant adverse events was taken from the observational study and estimated to be 56% of the RW patients treated with FL only and 65% of the RW patients treated with FL + oxaliplatin. An adverse event was considered significant if it led to a dose modification of the adjuvant chemotherapy regimens. Patients who remained disease-free following adjuvant treatment were assigned a utility score of 0.92, while patients who relapsed were assigned a utility score of 0.24.[114, 115] Details on how these utility scores were derived can be found in appendix 4A.

Model Input – Cost

This health economic analysis was conducted using a hospital's perspective. Resource use was based on RW patients and included patient level chemotherapy usage, inpatient hospital days (associated with administration of treatment and/or serious adverse events), outpatient visits, daycare treatments, laboratory services, imaging services, radiotherapy, surgical and other procedures, and concomitant medications (prophylactic with chemotherapy, and for the treatment of adverse events). The unit cost calculations of inpatient hospital days, outpatient visits and daycare treatments were based on detailed micro-costing studies reflecting full hospital costs, including overhead costs.[95, 96] The resource use of surgical procedures, laboratory services and imaging services was valued using the fees as issued by the NZa.[4] Unit costs of chemotherapy and concomitant medications were acquired from the Pharmacotherapeutic Aid Committee.[97] Costs were reported in Euro 2012. Where necessary, costs were adjusted to 2012 using the general price index from the Dutch Central bureau of Statistics.[116]

Cost of adjuvant treatment

In the model, three costing periods could be distinguished. Firstly, cost of the adjuvant treatment period, i.e. costs incurred from first administration of chemotherapy until 30 days after the last administration. These costs were assigned to the first cycle of the model. Mean total cost of adjuvant treatment

with FL only were calculated at €5,369 (SD €3,302). Mean treatment cost of FL + oxaliplatin were €20,861 (SD €8,555).[110] Although these cost estimates directly came from RW patients and therefore, like the effectiveness estimates, being subject to bias, we found that the mean total treatment costs were not associated with age or comorbidities. For this reason we could directly use these cost estimates in the model.

Cost of follow-up

Second, the costs of follow-up were assigned to subsequent cycles, in the "alive without relapse" health state. The follow-up duration was assumed to last for 5 years. Three-monthly cost of € 596 (SD € 927) was applied over the first two years of follow-up. This estimate was directly derived from the RW patients and found to be independent of type of adjuvant treatment. In subsequent cycles, from year 3-5, no data from RW patients were available. Here, a 3 monthly cost of follow-up of € 147 was assumed. This was based on expected resource use according to current Dutch guidelines.[117]

Cost of relapse

The third period included costs associated with the health state "alive following relapse". The treatment of relapsed disease was assumed to be independent of adjuvant treatment. The costs associated with relapse were taken from a related Dutch observational study evaluating patients with advanced colorectal cancer. Patients in this study were treated with FL, oxaliplatin and/or irinotecan in different treatment lines. Mean total cost were estimated at € 24,814 (SD € 16,967) per patient. The mean survival of these patients was 466 days, resulting in a three-monthly cost of € 4,872.[110] This cost estimate was applied to the "alive following relapse" health state.

More information on the calculation of all cost inputs can be found in appendix 4B.

Sensitivity Analysis

For each scenario, a number of univariate sensitivity analyses were performed to examine the impact of alternative parametric assumptions on the cost-effectiveness estimate. First, zero discount rates for cost and health outcomes were applied. Next, we studied the alternative assumption of having relapse rates

beyond 2 years in the ineligible patient cohort that equal the relapse rates of the eligible patients (instead of multiplying this rate by three). Subsequently, alternative utility scores for patients who remain disease-free and suffer a relapse were used. Furthermore, assumptions concerning the choice for FL + oxaliplatin regimen used (FOLFOX versus CAPOX) were made as well as assumptions concerning the impact of recent guideline changes on the use of cytotoxic regimens for metastatic disease. Alternative percentages of RW patients being eligible for the MOSAIC trial were also applied. Next we conducted one-way sensitivity analyses to explore the impact of alternative durations over which patients may relapse or die from colon cancer on cost-effectiveness estimates,

Table 4.4 Input parameters of the model

	Distribution	Parameters		Source
Transition probability parameters		Alpha	Beta	
Probabilities of staying alive without relapse (tp_1) (time-dependent, differed per scenario)	beta	Derived from KM DFS curves & number of patients at risk		MOSAIC trial & RW patients[53, 92]
Probabilities of moving from relapse state to death (tp_4) (time-dependent, differed per scenario)	beta	Derived from KM DFS curves & number of patients at risk		MOSAIC trial & RW patients[64, 110]
		mean	SD	
Hazard Ratio of DFS FL + oxaliplatin versus FL alone HR = 0.78 (95% CI 0.65-0.93), LN(HR) = -0.249	Lognormal	-0.249	0.093	MOSAIC trial[53]
Cost parameters		mean	SD	
Adjuvant treatment costs, FL alone	Bootstrapping	5,369	3,302	RW patients[110]
Adjuvant treatment costs, FL + oxaliplatin	Bootstrapping	20,861	8,555	RW patients[110]
		alpha	lambda	
Costs of follow-up, first 2 years	gamma	0.413	0.000693	RW patients[110]
Costs of relapse	gamma	2.139	0.000086	RW patients[110]
Utility values		mean	SD	
Utility of being on FL + oxaliplatin treatment	normal	0.65	0.036	Ness et al., 1999 [114]
Utility of being on FL treatment	normal	0.66	0.036	Ness et al., 1999[114]
Utility of being disease free	normal	0.92	0.050	Ramsay et al., 2000[115]
Utility of being in relapse	normal	0.24	0.041	Ness et al., 1999 [114]

FL, fluoropyrimidines; tp, transition probability; RW, real world; HR, hazard ratio; DFS, disease-free survival; OS, overall survival; SD, standard deviation

as well as the impact of assuming a 7-year time horizon for the evaluation of cost and health outcomes.

Probabilistic sensitivity analysis was undertaken to allow the effects of the joint uncertainty across all input parameters. Each model parameter was assigned a unique probability distribution based upon estimates of uncertainty. Bootstrapping was used to reflect the distribution in treatment costs, sampling from the RW patient data. Table 4.4 contains an overview of the input parameters and their corresponding distributions.

Monte Carlo sampling techniques were used (10 000 repeated random samples) to generate distributions of lifetime cost and health outcomes for patients treated with FL + oxaliplatin versus FL alone, for each scenario. Per scenario, the results of the probabilistic sensitivity analysis are presented as confidence intervals surrounding the base-case values, incremental cost-effectiveness planes and acceptability curves.

RESULTS

As illustrated in Figure 4.2, the output of the Markov model matched the results of the MOSAIC trial and the RW patients adequately based on face validity. That is, the 2-year and 6-year survival probabilities of the modelled treatment arms in scenarios 1 and 2, and the 2-year survival probabilities of the modelled ineligible oxaliplatin cohort, as can be derived from the curves presented in Figure 4.2, are very similar to the survival rates observed in the MOSAIC trial and the RW observational study, which are presented in Table 4.1.

Table 4.5 presents the expected discounted and undiscounted cost and health outcomes for each of the four scenarios. Estimates of the LYs are shown in parentheses. For all scenarios, Table 4.5 shows that the addition of oxaliplatin to FL is expected to result in increased costs but also additional QALYs given a lifetime horizon. In scenario 1, oxaliplatin showed an incremental QALY gain of 1.02 and an incremental cost of €9,961; the results in scenario 2 were similar, with an incremental QALY gain of 1.13 and incremental cost of €11,055. The corresponding incremental cost-effectiveness ratios in these two scenarios were very similar as well (€9,766 and €9,783, respectively). This was as expected given the comparable populations used in these scenarios. Scenarios 3 and 4

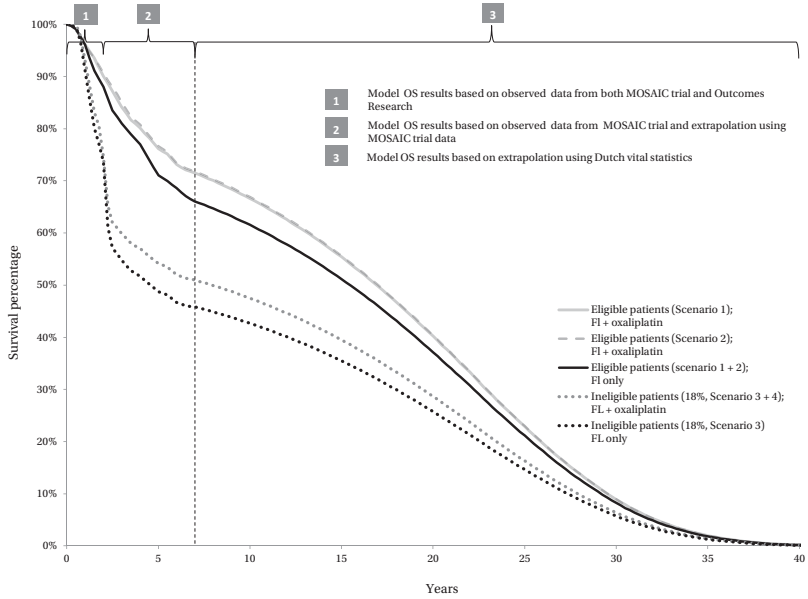


Figure 4.2 Model Validation; Undiscounted results from modelled overall survival, by scenario and treatment group

also included RW patients who would not have been eligible for the MOSAIC trial. The assumption of a similar treatment effect of oxaliplatin in this ineligible proportion (scenario 3) resulted in a QALY gain of 1.17, incremental cost of €9,814, and an incremental cost-effectiveness ratio of €8,388, which is beneficial when compared to the results from scenario 1 and 2. This can be explained by the worse life expectancy of ineligible patients. In scenario 4, where no effect of oxaliplatin was assumed in ineligible patients, we found a lower overall incremental QALY gain of 0.93 against incremental cost of €11,854, which resulted in an incremental cost-effectiveness ratio of €12,746. The ICERs are only marginally influenced by discounting.

Sensitivity Analysis

Table 4.6 presents the results of the one-way sensitivity analysis. The sensitivity analysis suggests that all alternative assumptions resulted in some variability in the cost-effectiveness estimates. Reducing the time horizon to 7 years resulted

Table 4.5 Summary of the cost-effectiveness results

Results discounted (costs 4% and benefits 1.5%) over a lifetime horizon										
Scenario	Treatment	Mean costs	Incremental costs	*95% CI	Mean QALYs (LYs)		Incremental QALYs (LYs)		*95% CI	Incremental costs per QALY gained
1	FL	€ 40,698	€ 9,961	-€ 10,245 - € 28,160	10.53	(12.86)	1.02	(0.86)	0.07 - 1.89	€ 9,766
	FL + oxaliplatin	€ 50,659			11.55	(13.72)				
2	FL	€ 40,698	€ 11,055	-€ 7,861 - € 28,619	10.53	(12.86)	1.13	(1.04)	0.22 - 2.00	€ 9,783
	FL + oxaliplatin	€ 51,753			11.67	(13.89)				
3	FL	€ 46,355	€ 9,814	-€ 13,958 - € 29,249	9.67	(12.23)	1.17	(1.00)	0.10 - 2.17	€ 8,388
	FL + oxaliplatin	€ 56,169			10.84	(13.24)				
4	FL	€ 44,315	€ 11,854	-€ 8,323 - € 30,876	9.91	(12.39)	0.93	(0.85)	-0.10 - 1.90	€ 12,746
	FL + oxaliplatin	€ 56,169			10.84	(13.24)				
Undiscounted results										
Scenario	Treatment	Mean costs	Incremental costs	*95% CI	Mean QALYs (LYs)		Incremental QALYs (LYs)		*95% CI	Incremental costs per QALY gained
1	FL	€ 56,390	€ 7,621	-€ 20,002 - € 27,715	12.30	(15.05)	1.22	(1.00)	0.10 - 2.28	€ 6,247
	FL + oxaliplatin	€ 64,011			13.52	(16.04)				
2	FL	€ 56,390	€ 9,308	-€ 15,951 - € 29,522	12.30	(15.05)	1.36	(1.20)	0.26 - 2.39	€ 6,844
	FL + oxaliplatin	€ 65,698			13.66	(16.25)				
3	FL	€ 65,263	€ 7,342	-€ 25,455 - € 31,381	11.27	(14.30)	1.40	(1.17)	0.14 - 2.59	€ 5,244
	FL + oxaliplatin	€ 72,605			12.67	(15.47)				
4	FL	€ 62,184	€ 10,421	-€ 19660 - € 33462	11.56	(14.48)	1.11	(0.99)	-0.14 - 2.33	€ 9,388
	FL + oxaliplatin	€ 72,605			12.67	(15.47)				

FL, fluoropyrimidines; CI, confidence interval; QALY, Quality adjusted life-year; LY, Life-year

Table 4.6 Univariate sensitivity analysis results

Parameter	value in base-case analysis	Value in sensitivity analysis	ICER per Scenario			
			1	2	3	4
Basecase	–	–	€ 9,766	€ 9,783	€ 8,388	€ 12,746
Hazard ratio 5-year DFS	0.78	0.65 (Lower limit of MOSAIC DFS 95% CI)	€ 4,579	€ 4,838	€ 3,821	€ 5,719
Hazard ratio 5-year DFS	0.78	0.93 (Upper limit of MOSAIC DFS 95% CI)	€ 32,280	€ 27,037	€ 25,432	€ 56,685
Utility of remaining disease-free	0.92	0.8[100]	€ 11,397	€ 11,338	€ 9,782	€ 14,840
Utility of being in relapse	0.24	0.58[190]	€ 11,230	€ 10,796	€ 9,526	€ 14,129
Treatment costs FL + oxaliplatin	€ 20,861 (48% FOLFOX and 52% CAPOX)	€ 27,446 based on 100% FOLFOX	€ 16,222	€ 15,556	€ 14,023	€ 19,844
	€ 20,861 (48% FOLFOX and 52% CAPOX)	€ 14,783 based on 100% CAPOX	€ 3,807	€ 4,389	€ 3,195	€ 6,216
Costs of relapse	€ 24,814	€ 24,814 plus € 18,457 extra costs for bevacizumab use[88]	€ 5,274	€ 6,389	€ 4,318	€ 9,395
Assumed relapse rates for ineligibles beyond 2 years	3 times higher than MOSAIC patients	Same rates as MOSAIC patients	€ 9,766	€ 9,783	€ 8,637	€ 12,755
Assumed duration over which relapses and death due to colon cancer may occur	5 years for relapses, 7 years for deaths	8 and 10 years, respectively	€ 9,685	€ 9,673	€ 8,393	€ 12,681
Time horizon	life-time (i.e. 42 years)	Within- trial analysis (7-year horizon)	€ 39,397	€ 36,299	€ 33,592	€ 46,020
Patients eligible for MOSAIC trial	82% eligible patients	50% eligible patients	€ 9,766	€ 9,783	€ 6,173	€ 23,493
	82% eligible patients	15% eligible patients	€ 9,766	€ 9,783	€ 3,992	€ 88,941

DFS, disease-free survival; ICER, incremental cost-effectiveness ratio; FL, fluoropyrimidines; CI, confidence interval

in a large increase in ICERs in all scenarios. The cost-effectiveness of oxaliplatin in all scenarios was also sensitive to the treatment cost of FL + oxaliplatin (FOLFOX versus CAPOX). Furthermore, the ICERs in all scenarios were sensitive to the parameters regarding the hazard ratio for disease free survival, while only in scenario 4 the increase in cost per QALY was particularly sensitive to the percentage of patients being eligible for the MOSAIC trial. The ICERs are only marginally influenced by the values of the other parameters, which included relapse rates beyond 2 years of ineligible patients, the utility of being disease-free and in relapse, the cost of relapse, and the assumed time period during which relapses and death due to colon cancer may occur.

Figure 4.3 presents cost-effectiveness planes for all scenarios, showing the marginal cost and QALYs associated with FL + oxaliplatin in comparison to FL alone. In all scenarios, the addition of oxaliplatin is expected to produce greater health gains than FL only albeit at a greater cost, in most of the 10,000 simulations. Considering all four scenarios, the 95% confidence intervals of the incremental cost (generated from the simulations) varied from - €13,958 to €30,876

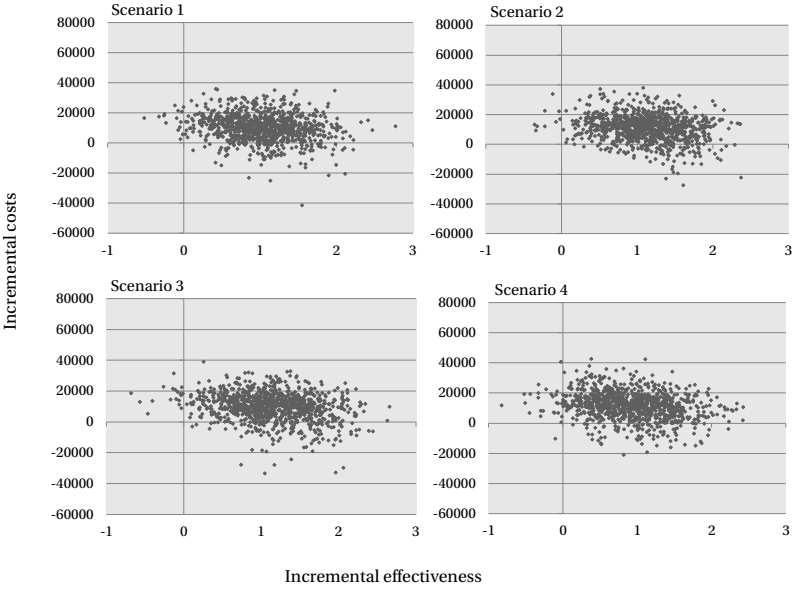


Figure 4.3 Cost-effectiveness planes for FL + oxaliplatin versus Fl alone, by scenario

(Table 5.5). In scenarios 1-3, the 95% confidence intervals of the incremental QALYs varied from 0.07 to 2.17. Scenario 4 is the only scenario which resulted in a minimum value that was less than 0 (-0.10). This is as expected because this scenario assumed no treatment benefit from oxaliplatin for the ineligible RW patients, which comprised 18% of the population.

Cost-effectiveness acceptability curves are shown per scenario in Figure 4.4 to facilitate conclusions about the cost-effectiveness of oxaliplatin at different willingness-to-pay thresholds. For example, at a willingness-to-pay threshold of €50,000, there was a probability in scenarios 1-3 of more than 90% that the addition of oxaliplatin would be cost-effective. This percentage was 86% in scenario 4.

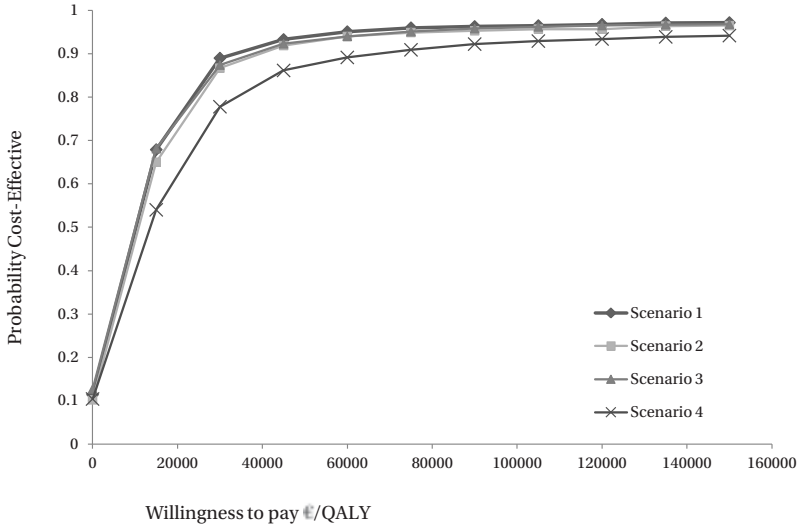


Figure 4.4 Cost-effectiveness acceptability curves for FL + oxaliplatin versus FL alone, by scenario

DISCUSSION

The increasing number of expensive drugs is making it extremely difficult for health care systems to strike an optimal balance between ensuring timely access to these 'promising' drugs and having sufficient evidence of their comparative benefits and risks. Dutch policy regulations allow early conditional access for these expensive drugs, but require the estimation of RW cost-effectiveness using RW data after four years.

In our analysis, the RW cost-effectiveness of FL + oxaliplatin versus FL alone in the adjuvant treatment of colon cancer was estimated by combining the MOSAIC clinical trial with data from a Dutch observational study in a Markov model. We found certain differences but also great similarities between trial and RW patients and incorporated these findings in different model scenarios. This chapter illustrates how one could design and implement a real-world cost-effectiveness study to yield internally valid results that could also be generalisable. The ICERs of the different scenarios all suggest that FL + oxaliplatin versus FL alone is cost-effective in the adjuvant treatment of stage III colon cancer.

In scenario 1, clinical effectiveness was modelled based on just the MOSAIC trial. In this scenario incremental QALYs gained with the addition of oxaliplatin were estimated at 1.02 at a discount rate of 1.5%. Five other cost-effectiveness analyses based on the MOSAIC trial have been published.[82, 83, 86, 99, 100] Comparing the QALY estimates, all were very similar to ours, with incremental differences between the two arms of 0.64 in the Canadian model, 0.75 in the US model, 0.76 in the Japanese model, and 0.68 and 1.33 in two UK models. The differences mainly resulted from the different discount rates in the analyses (5%, 3%, 3%, 3.5% and 1.5% in the Canadian, US, Japanese, and the two UK models, respectively).[82, 83, 86, 99, 100]

In our model, cost calculations were based on Dutch RW resource use and unit costs, resulting in incremental total cost of €9,961 at a discount rate of 4% in scenario 1. Other published models also used country specific unit cost, but resource utilisation estimates were mainly informed from the MOSAIC trial.[53] Reported incremental costs were €10,779, €13,370, €11,041, €4,969, and €6,763 in the Canadian, US, Japanese and two UK models, respectively (country specific currencies converted to Euro 2012). Differences resulted from relative differences in price and administration of FL + oxaliplatin versus FL only, and

differences between countries in costs post-relapse, which were also sensitive to differences in discount rates (5%, 3%, 3%, 3.5% and 6%, respectively).

In scenario 1, the resulting ICER was €9,766 per QALY gained. In the Canadian, US, Japanese and UK evaluations, the reported ICERS per QALY gained were €18,986, €17,898, €14,438, €7,309, and €4,518, respectively. All studies concluded that adding oxaliplatin to the adjuvant treatment with FL in patients with stage III colon cancer was cost-effective.

However, the effectiveness results of all these evaluations, including scenario 1 of our study, have been based on a clinical trial. The use of results from clinical trials always raises the question of external validity. Are the benefits seen by patients in the clinical trial (efficacy) applicable to patients treated in routine practice (effectiveness)? We made use of the internal validity provided by the randomised design of the MOSAIC trial, while taking external validity into consideration by incorporating effectiveness results from RW patients in scenarios 2-4 of our model. We examined RW patients included in a Dutch population based observational study, to see how many patients fulfilled the MOSAIC trial eligibility criteria. A total of 82% of the RW patients receiving oxaliplatin fulfilled these criteria and were added in scenario 2. The resulting base case ICER of €9,783 was very similar to the results of scenario 1. In scenarios 3 and 4, the 18% of the patient population that did not meet the MOSAIC trial eligibility criteria were added to the model. The resulting ICERs per QALY gained were €8,388 and €12,746 for scenario 3 and 4, respectively, which all suggest the cost-effectiveness of oxaliplatin in the real world.

The value of scenarios 3 and 4 is that they focus on estimating the cost-effectiveness of oxaliplatin in the most relevant population: all stage III colon cancer patients who are treated with oxaliplatin in Dutch RW practice. In the RW population-based observational study, all patients treated with FL + oxaliplatin in 2005 and 2006 in 19 representative hospitals were included (n = 281) (REF). Apart from 14% (n = 38) of patients with missing CEA values, which were considered to occur completely at random, these patients were all taken into account in scenarios 3 and 4. Based on this we believe that our results are generalisable to Dutch RW practice.

However, it is possible that physicians will be steadily less reluctant to prescribe oxaliplatin to older patients. If this happens, the target population (of patients in the future) will be broader than the population we considered (i.e.

only patients who received oxaliplatin in 2005 and 2006), which would reduce the generalisability of our present results. The probability of limited generalisability due to temporal changes in the patient population, is particularly present in case new medicines face a slow uptake. However, in our study we do not expect this since we know that most physicians began to use FL + oxaliplatin as standard treatment soon after the guideline changed in early 2005. This rapid adoption is most likely due to the extensive experience that physicians already had using oxaliplatin in advanced colorectal cancer.[60] Moreover, during 2005 and 2006 we did not observe any changes in the types of patients receiving oxaliplatin and even now there is no strong evidence to support the use of adjuvant treatment with oxaliplatin in the elderly.[120]

In the RW observational study, 82% of the patients receiving oxaliplatin and 63% of the patients receiving FL only were eligible for the MOSAIC trial. These percentages are similar to other expensive oncology medicines, where substantial percentages of patients treated with these new drugs would not have fulfilled the pivotal clinical registration trial eligibility criteria.[119] Therefore, cost-effectiveness analyses of new expensive medicines based only on trial data run the risk of not being applicable to the RW population that is actually treated with these medicines; they are in fact cost-efficacy analyses and not cost-effectiveness analyses.

In this cost-effectiveness analysis we took the relative treatment effect of oxaliplatin directly from the MOSAIC trial because of its randomised study design. Subsequently we concluded that the eligible RW oxaliplatin population was sufficiently comparable to the MOSAIC trial population, based on eligibility, baseline characteristics, the way treatments were given and the treatment outcomes (Kaplan-Meier curves), and therefore felt confident that the relative treatment effect of oxaliplatin observed in the MOSAIC trial would also be expected in the eligible RW oxaliplatin-treated patients. We did notice some differences in baseline prognostic factors between MOSAIC patients and RW eligible patients, such as depth of invasion, number of involved nodes and tumour differentiation. However, we expect that these differences do not translate into systematic differences in the underlying baseline prognosis and/or the relative treatment effect of oxaliplatin since subgroup analyses of the MOSAIC trial showed that the treatment effect (measured as reduced risk of relapse) was consistent in all subgroups defined using various baseline prognostic factors.[53]

Next to this, the full 95% confidence interval of the trial-based hazard ratio (HR = 0.78, 98% CI 0.65 – 0.93) was considered in the univariate and probabilistic sensitivity analyses, which resulted in maximal ICERs of €32,280 and €27,037 in scenarios 1 and 2, respectively.

Additional assumptions were made in scenarios 3 and 4 regarding the baseline prognosis and treatment effect of oxaliplatin for the ineligible RW oxaliplatin patients. The lower 2-year DFS and OS rates of ineligible RW oxaliplatin patients can largely be explained by the higher percentage of patients in the ineligible RW oxaliplatin with abnormal CEA values. Projections beyond two years were uncertain, but alternative assumptions did not impact the ICER results in the univariate sensitivity analysis. We varied the expected treatment effect of oxaliplatin among ineligible RW patients using a base-case (scenario 3) and a worst-case (scenario 4) scenario. In the base-case scenario, an equal relative effectiveness as in the MOSAIC trial was assumed. This is based on subgroup analyses of the MOSAIC trial that showed a positive but statistically non-significant effect of oxaliplatin in patients with elevated CEA levels.[53, 64] The worst-case scenario involved a zero treatment effect for patients who were ineligible for the MOSAIC trial. This was viewed as a worst-case scenario since it is very unlikely that oxaliplatin would be harmful for patients with elevated CEA levels as oxaliplatin is known to have be beneficial in the treatment of advanced colon carcinoma.[57]

There were several study limitations. Firstly, no comparable data from RW patients treated with FL only were available, due to the rapid adoption of oxaliplatin in early 2005 An effort was made to correct for the resulting differences between the patients treated with FL only versus FL + oxaliplatin, but this was hampered by the low number of eligible RW patients treated with FL only (n = 54) and a median follow-up time of only 2 years, which appeared to be the maximum achievable duration given the limited study timeframe of 4 years of conditional reimbursement .[119]

But even without power problems, it would have been uncertain whether correction led to results that were valid, since the potential for bias on unmeasured baseline characteristics cannot be ruled out. RW patients receiving FL only in 2004, just before oxaliplatin became indicated, i.e. historical control patients might be usable in this context. However, since the MOSAIC patients were very

similar to the eligible RW patients treated with FL + oxaliplatin, we don't expect this to alter the ICER results significantly.

Secondly, the availability of relevant utility values for the calculation of QALYs was limited. We believe that the evaluation presented here used the best available data; the results of the sensitivity analysis suggest that the effect of this parameter on long-term results is modest.

A third possible limitation is that we did not correct for possible differences in baseline prognostic factors when using cost data from RW patients for both treatment groups. This might have resulted in less valid cost comparisons. However, we found that the treatment costs were not associated with baseline characteristics such as age and comorbidities. Moreover, treatment costs were already incurred within the first six months and appeared to be independent of a patients' prognosis. For these reasons we don't expect this limitation to influence the ICER results.

Our effectiveness estimate was based on RW patients combined with only one trial. The MOSAIC trial was the clinical registration trial based on which treatment with FL + oxaliplatin became indicated for the treatment of stage III colon cancer in the Netherlands. When the cost-effectiveness model was built, only one other trial investigating FL + oxaliplatin versus FL only was available.[62] However, dosing schemes used in that trial are not seen in Dutch RW practice. We therefore decided not to use this trial in our model. Recently, a third trial was published.[71] Efficacy results of the MOSAIC trial were comparable to the results presented in the other two trials. Incorporating these trials into our analysis would have resulted in similar ICERs, which would only support the conclusion of its cost-effectiveness.

Lastly, modelling in cost-effectiveness analyses always requires various assumptions. In all scenarios, univariate and probabilistic sensitivity analyses indicated that the ICERs are acceptable and robust under a wide range of model assumptions.

Conclusions

The requirement in Dutch policy for RW cost-effectiveness estimations using patients treated in daily clinical practice will result in the application of different methods that aim to prove internally valid estimations, but also focus on the generalisability of the results in Dutch RW practice. This can result in better

evidence by addressing uncertainty in outcomes arising from the gap between clinical trials and everyday practice. In this study we combined data from a Dutch observational study with the pivotal clinical registration trial. We expect that this approach can be used in determining the RW cost-effectiveness of other expensive medicines. In conclusion, the ICERs of the different scenarios are all acceptable and support the use of oxaliplatin in the adjuvant treatment of stage III colon cancer in the Netherlands.

APPENDICES

Appendix 4A. Health utility scores

4A1 Utility parameters used in the economic model

Health status	*utility*	*SE*	*References*
On adjuvant chemotherapy (without significant side-effects)	0.70	0.036	Ness et al., 1999[114]
On adjuvant chemotherapy (with significant side-effects)	0.63	0.036	Ness et al.,1999[114]
In remission	0.92	0.05	Ramsay et al., 2000[115]
On palliative chemotherapy	0.24	0.041	Ness et al., 1999[114]
Patients experiencing significant side-effects	*percentage*		
FL only	56%		van Gils et al., 2010[110]
FL + oxaliplatin	65%		van Gils et al., 2010[110]
Utility input in model incorporating significant side-effects	*utility*	*SE*	
on adjuvant treatment with FL only	0.66	0.036	
on adjuvant treatment with FL + oxaliplatin	0.65	0.036	

FL, fluoropyrimidine; SE, standard error

In order to derive estimates of quality adjusted life-years (QALYs) for each adjuvant treatment, the survival benefits seen within the MOSAIC trial and RW observational study need to be weighted by patients QoL over that period. This was done by assigning health utilities to the various health states in which patients could be.

Since health utility scores were not collected in either MOSAIC or RW patients. Modelled survival estimates were adjusted to account for the patients'

health related quality of life using published colorectal cancer utility estimates. A similar approach was used in the model of Eggington et al.[86]

Utility associated with "receiving adjuvant chemotherapy" in stage III colon cancer

Utility estimates for patients on adjuvant treatment were taken from a study by Ness et al.[114] In this study a standard gamble approach was taken to elicit utility scores from 81 patients with colorectal cancer. The results report utilities for all stages, including those of patients with stage III treated with chemotherapy. Two separate utilities were reported for patients with and without significant side-effects, 0.63 and 0.70, respectively, reflecting a degree of utility loss associated with treatment-related adverse events.

Data on toxicities were collected within the RW observational study. A toxicity was considered significant if it led to dose modifications. Distinctions were made between haematological toxicity, gastrointestinal toxicity, neurological toxicity and toxicity as a result of the hand-foot syndrome. Haematological toxicity and neurotoxicity were the most frequent reasons for dose adjustment and/or interrupting treatment with oxaliplatin.[110] In total 56% of the patients receiving treatment FL only treatment experienced significant side-effects, this percentage was 65% in patients receiving FL + oxaliplatin.

The utility scores reported by Ness et al., were multiplied by the percentage of patients that did and did not experience significant toxicities. This resulted in a utility of 0.66 and 0.65 for the FL only and FL + oxaliplatin treatments, respectively. These utilities were applied to the first two cycles of the model.

Utility associated with the "disease-free" health state in stage III colon cancer

Utility estimates associated with the disease-free health state were taken from a study conducted by Ramsey et al.[115] This study included 173 patients and reported a mean utility rate of 0.92 beyond 60 months, after which patients are assumed to no longer be at risk of relapse. This has been used as a proxy utility for patients in remission following adjuvant chemotherapy.

Utility associated with "being relapsed" in stage III colon cancer

The study of Ness et al reported a mean utility of 0.24 for relapsed patients. This single utility score is applied to the patients who relapsed for their remaining survival period following relapse.

Appendix 4B. Cost calculations

Cost of treatment and cost of follow-up (year 1 and 2)

Within the RW observational study, resource use data were drawn from a random subset of 206 individual patients included in this study. Mean costs per patient were calculated for the four most common treatment regimens seen in daily practice: 5-FU/LV, capecitabine, 5-FU/LV with oxaliplatin (FOLFOX) and

4B1 Cost in the treatment, follow-up phase (period 1 and 2 Euro 2012)

n = 206	FL only		FL + oxaliplatin	
	5FU/LV	Capecitabine	FOLFOX	CAPOX
	N = 15	N = 89	N = 37	N = 65
Total treatment costs				
mean	€ 6,163	€ 5,229	€ 27,446	€ 14,783
SD	€ 3,076	€ 3,356	€ 11,938	€ 5,944
Total costs of follow-up				
mean	€ 3,518	€ 4,507	€ 5,347	€ 3,578
SD	€ 3,863	€ 8,083	€ 7,511	€ 4,240
Share of regimen within RW population	15%	85%	48%	52%
Pooled total treatment costs*				
mean	€ 5,369		€ 20,861	
SD	€ 3,302		€ 8,555	
Total costs of follow-up				
mean	€ 4,293			
SD	€ 6,677			
Average Follow-up time	1.8 year			
Average cost of follow-up/ 3-month cycle, during year 1 and 2				
mean	€ 596			
SD	€ 927			

FL , fluoropyrimidines; FOLFOX, oxaliplatin combined with 5FU/LV; CAPOX, oxaliplatin combined with capecitabine SD, standard deviation; * calculated by multiplying costs per regimen by the actual share found in RW practice

4B2 Resource use & costs year 3 - 5

	Unit costs	Units/3 months	Costs/cycle
Laboratory tests	€ 38	0.50	€ 19
Ultrasound liver	€ 92	0.25	€ 23
X-ray of thorax	€ 55	0.25	€ 14
Colonoscopy	€ 473	0.08	€ 39
Monitoring visits	€ 104	0.50	€ 52
Totals/ 3- month cycle			*€ 147*

Fees as issued by the Dutch Healthcare Authority[4]

4B3 Total costs post-relapse (including costs in later treatment lines) in patients receiving FL alone, FL + oxaliplatin or FL + irinotecan as first-line treatment

	FL alone	FL + oxaliplatin	FL + irinotecan
	n = 198	n = 92	n = 26
Total cost post-relapse			
mean	€ 20,433	€ 29,435	€ 41,824
SD	€ 15,592	€ 20,681	€ 25,883
Total costs post-relapse over all treatments			
mean	€ 24,814		
SD	€ 16,967		
Mean survival		466 days	
Mean cost post-relapse/ 3-month cycle	€ 4,872		

van Gils et al., 2010[110]; Fl, fluoropyrimidine; SD, standard deviation.

capecitabine with oxaliplatin (CAPOX). Total costs for individual patients were determined by the identification of resource use and unit costs of the following cost components: inpatient hospital days, intensive care days, outpatient visits, consultations by telephone, daycare treatments, emergency room visits, radiotherapy, surgical procedures, laboratory services, medical imaging services, chemotherapy and concomitant medications[4, 95-97]

Resource use was divided into two time periods, a treatment phase (period 1) and a follow-up phase (Period 2). Period 1 began on day 1 of the first administration of adjuvant chemotherapy. To capture resource use resulting from treatment related toxicity, period 1 ended one month after the last administration of chemotherapy. Period 2 started one month after the last administration of chemotherapy and lasted until disease progression (or end of follow up).[110]

Chapter 5

Practical feasibility of outcomes research in oncology: Lessons learned in assessing drug use and cost-effectiveness in The Netherlands

European Journal of Cancer 2013; 49: 8– 16

Margreet Franken*
Chantal van Gils*
Jennifer Gaultney
Gepke Delwel
Wim Goettsch
Peter Huijgens
Adri Steenhoek
Miriam Koopman
Kees Punt
Ken Redekop
Carin Uyl- de Groot

*Both authors contributed equally to this manuscript

SUMMARY

Objective: To investigate the practical feasibility to develop evidence on drug use and cost-effectiveness in oncology practice. Patients and Methods: Feasibility was examined using three Dutch case studies. Each case study investigated the degree of appropriate drug use and its incremental cost-effectiveness. Detailed data were retrospectively collected from hospital records. In total, 391, 316 and 139 patients with stage III colon cancer, metastatic colorectal cancer and multiple myeloma were included in 19, 29, and 42 hospitals, respectively. Results: The methods used in the case studies were feasible to develop evidence on some aspects of drug use including types of treatments used, dosages, dose modifications, and healthcare costs. Aspects such as baseline patient characteristics, reasons to start or stop a treatment, and treatment effects were less feasible because of missing values. Despite difficulties to correct for confounding by indication, it was possible to estimate incremental cost-effectiveness by synthesising evidence in two of the three case studies. Conclusion: It is possible to generate evidence about drug use and cost-effectiveness in oncology practice to facilitate informed decision-making by both payers and physicians. This can improve quality of care and enhance the efficient allocation of resources. However, the optimal approach differs between drugs and their indications. Generating high-quality evidence requires active interdisciplinary collaboration. Patient registries can facilitate data collection but cannot resolve all issues. In most circumstances it is inevitable to use data-synthesis to obtain valid incremental cost-effectiveness estimates, but for some indications it will not be feasible to derive a valid and precise estimate.

INTRODUCTION

The increasing number of expensive oncology drugs is making it extremely difficult to strike an optimal balance between ensuring timely access to 'promising' drugs and having sufficient evidence of their comparative benefits and risks. In the last decade, governments have therefore introduced policies linking reimbursement to a requirement for additional data collection.[7-10] It has been claimed that the resulting 'schemes' or requirements, such as 'patient access schemes,'[121] 'managed entry agreements,'[122] 'access/coverage with evidence development,'[123, 124] comparative effectiveness research,[35, 37] or outcomes research,[11-13] can result in better evidence by addressing uncertainty arising from the gap between clinical trials and everyday practice.

The initiatives to promote and implement such requirements have initiated a stream of literature, touting its potential as well as highlighting the myriad of methodological challenges to its feasibility.[15, 35, 37-39, 125-128] In particular, one of the major concerns raised has been the lack of a randomised controlled setting, which results in problems with internal validity.[15-18] Moreover, other challenges have been debated such as the role of data synthesis and modelling,[17, 35, 36, 129] use of existing data sources,[35, 37-40] applicability of outcome measures,[8, 12] timeliness,[9, 41] and generalisability.[42] However, these issues have mainly been discussed on the basis of theoretical expectations and expert opinion; studies of the practical feasibility are scarce.

Experience with conducting outcomes research has already been gained in The Netherlands as a result of policy regulations for expensive inpatient drugs implemented in 2006. If a drug is included in this policy, hospitals receive an additional ear-marked budget of 80% of its acquisition costs.[109] However, this early access is linked with the obligation to gather data on appropriate drug use and incremental cost-effectiveness.[5, 109] In practice, this means that after four years of use, a reassessment will determine whether or not additional financing will continue to exist.

This Chapter describes our experiences in The Netherlands regarding the practical feasibility of different aspects of outcomes research in oncology. These experiences were based on three different outcomes research studies which examined the feasibility to gather evidence on appropriate drug use and estimate incremental cost-effectiveness of a particular drug.

METHODS

We conducted outcomes research of two expensive drugs for three indications in cancer: oxaliplatin as adjuvant treatment in stage III colon cancer, oxaliplatin as palliative treatment in metastatic colorectal cancer and bortezomib as palliative treatment in relapsed or refractory multiple myeloma. Each of the studies was used to investigate the feasibility to develop evidence on appropriate drug use and to estimate incremental cost-effectiveness. For appropriate drug use, we examined the feasibility to develop evidence in the following areas: types of treatments and regimes (*'treatments'*), dosages and dose modifications (*'dosages'*), baseline patient characteristics (*'patients'*), reasons for choosing a particular treatment and starting or stopping treatment (*'reasons'*), and treatment outcomes (*'effects' and 'costs'*). To investigate the feasibility of data collection, we examined which data were available through existing databases and which data required retrieval from hospital records. For incremental cost-effectiveness, we investigated the feasibility to obtain comparable patient groups, identify treatment comparators, obtain information from literature, and estimate incremental cost-effectiveness. We explored issues with internal validity, data synthesis and modelling, outcome measures and generalisability.

Description of the case studies

In the two oxaliplatin studies, patients were identified using the population-based registry of the Dutch Comprehensive Cancer Centres. This registry enabled the identification of all Dutch patients who received chemotherapy (2249 stage III colon cancer patients diagnosed in 2005 and 2006; 1957 metastatic colorectal cancer patients diagnosed in 2003 and 2004). Since this registry did not contain all of the required data, we had to contact individual hospitals. Most of the Dutch hospitals (72%) were approached to expedite the data collection and we continued to include hospitals in order of response until the desired number of patients had been reached (stage III colon cancer: n=391; metastatic colorectal cancer n=316). Using hospital records, additional data were retrospectively collected on baseline patient characteristics, known prognostic information, considerations for choosing a treatment, types of treatments, disease free survival, and overall survival. For a randomly selected subgroup (stage III colon cancer: n=206; metastatic colorectal cancer n=130), detailed data were

Table 5.1 Summary table oxaliplatin in stage III colon cancer

Study design	Retrospective observational study			
	Number of patients			
Population-based cancer registry	2249 (diagnosed in 2005 and 2006)			
Additional/ detailed data collection (hospital records)	391/ 206			
Number of hospitals visited (% of Dutch hospitals)	19 (17%)			
Treatments received in everyday practice	FL + oxaliplatin FOLFOX or CAPOX[a] (n=281)		FL alone 5FU/LV or capecitabine (n=110)	
5FU/LV	48%		15%	
Capecitabine	54%		85%	
Dosages[b]				
Oxaliplatin dose according to guidelines	1,020 mg/m2		n/a	
Percentage of planned dose given	81% for FOLFOX 71% for CAPOX		n/a	
Prognostic baseline characteristics		*% Missing*		*% Missing*
Eligible for pivotal registration trial	82%	14%	63%	23%
Age [median (range)]	61 (22–82)		73 (41–85)	
Co-morbid conditions ≥ 2	11%		25%	
Depth of invasion T2-T3 (T4)	85% (15%)		88% (12%)	
Nodes involved N1 (N2)	60% (40%)		66%	
Abnormal CEA levels[c]	18%	14%	8%	22%
Reasons for not prescribing oxaliplatin[b]	Hospital policy (18%), advanced age (21%), patient refusal (19%), poor health status (10%), combination of these factors (7%), specific contra-indication (2%), and unknown (23%)			
Regimes requiring dose modifications	56%		66%	
Toxicity requiring hospitalisation	7%		6%	
Effects: 2-year disease-free survival probability				
Eligible [Mean (95% CI)]	78.4% (72.5% – 84.3%)		82.8% (72.5% – 93.0%)	
Ineligible [Mean (95% CI)]	56.7% (41.4% – 72.0%)		83.7% (70.5% – 96.8%)	
Costs[b]				
Total costs [mean (median)]	€19,639 (€20,230)		€5,055 (€4,482)	
Minimum – maximum	€1708 – €60,149		€316 – €12,127	

[a] FL, Fluoropyrimidines; FOLFOX, oxaliplatin combined with 5FU/LV; CAPOX, oxaliplatin combined with capecitabine
[b] Based on a representative subsample of 206 patients
[c] CEA, serum carcinoembryonic antigen levels

Table 5.2 Summary table oxaliplatin in metastatic colorectal cancer

Study design	Retrospective observational study					
	Number of patients					
Population-based cancer registry	1957 (diagnosed in 2003 and 2004)					
Additional/ detailed data collection (hospital records)	316/130					
Number of hospitals visited (% of Dutch hospitals)	29 (26%)					
Treatments						
First-line treatments received in everyday practice	FL + oxaliplatin FOLFOX/CAPOX [a] (n=92)		FL alone 5FU/LV [a] or capecitabine (n=198)		(FL +) irinotecan alone/FOLFIRI/ CAPIRI [a] (n=26)	
Dosages[b]						
Mean total cumulative dose of oxaliplatin	1274 mg		n/a		n/a	
Patients receiving 2nd (3rd) line treatment	57% (22%)		52% (22%)		32% (20%)	
Prognostic baseline characteristics		*% Missing*		*% Missing*		*% Missing*
Eligible for clinical trial	85%		63%		73%	
Age [median (range)]	60(29–81)		64(30–92)		59(39–73)	
WHO performance status ≥ 2	13%	41%	23%	39%	6%	35%
Abnormal lactate dehydrogenase levels	46%	13%	52%	20%	47%	27%
Resection of primary tumour	65%	4%	61%	3%	58%	8%
Toxicity in first line[b]						
Requiring hospitalisation	6%		13%		6%	
Causing termination of treatment	20%		1%		20%	
Effects: Overall Survival in months						
Eligible [Mean (median)] 95% CI	18.6 (14.1) 15.5 – 21.8		14.9 (11.3) 12.8 – 17.0		27.1 (21.3) 18.0 – 36.3	
Ineligible [Mean (median)] 95% CI	18.4 (17.8) 12.3 – 24.5		11.0 (6.6) 8.2 – 13.7		9.2 (6.9) 3.6 – 14.9	
Costs[b]						
Total costs [mean (median)]	€27,711 (€23,172)		€19,236 (€16,208)		€39,375 (€38,754)	
Minimum – maximum	€2,200 –€95,118		€462 –€65,288		€10,258 – €109,139	

[a] FL, Fluoropyrimidines; FOLFOX, oxaliplatin combined with 5FU/LV; CAPOX, oxaliplatin combined with
capecitabine; FOLFIRI, irinotecan combined with 5FU/LV; CAPIRI, irinotecan combined with capecitabine

[b] Based on a representative subsample of 130 patients

Table 5.3 Summary table bortezomib in relapsed/ refractory multiple myeloma

Study design	Retrospective observational study			
	Number of patients			
HOVON trial database/ detailed data collection	543/ 139			
Number of hospitals visited (% of Dutch hospitals)	41 (38%)			
Treatments				
Treatments received in everyday practice	Ever bortezomib [a] (n=72)		Never bortezomib (n=67)	
Bortezomib combination therapy	71%		n/a	
One other drug (dexamethasone)	80% (58%)		n/a	
Two/ three or more other drugs	12%/ 7%		n/a	
Bortezomib total dose compared to pivotal registration trial	87%		n/a	
Number of bortezomib cycles in everyday practice (pivotal registration trial)	4 (6)		n/a	
Prognostic factors at start of 2nd line treatment		*% Missing*		*% Missing*
Age [mean (range)]	57 (34-69)		58 (35-68)	
WHO performance status 0/ 1/ ≥ 2	59%/ 33%/ 8%	8%	35%/ 42%/ 22%	7%
Present with neurotoxicity	49%	10%	26%	3%
Serum B2 (mg/l) [mean (range)]	4 (1.3–16.7)	64%	3 (1.1–5.7)	79%
Albumin (g/l) [mean (range)]	40 (27.0–59.0)	31%	38.4 (16.6–52.0)	37%
Haemoglobin (mmol/l) [mean (range)]	7.5 (5.0–9.5)	31%	7.1 (2.1–10.0)	31%
First line HOVON50 experimental TAD arm	40%		27%	
Received allogeneic stem cell transplantation	27%	1%	15%	
Maintenance treatment				
None/ IFNα	43%/ 21%		63%/ 24%	
Thalidomide	36%		13%	
Best response first line treatment		*6%*		
Complete response/ Partial & minor response	16%/ 80%		10%/ 76%	
No change/ Progressive disease	3%/ 1%		4%/ 9%	
Time until first progression in months[median (SD) {range}]	27.6 (12.8) {2.0-57.9}	1%	22.4 (15.8) {1.9-61.4}	1%
Reasons to start a treatment regime		[b]		[b]
Regimes requiring dose modifications	52.5%		n/a	
Due to toxicity	79%	14%	n/a	
Effects				
Overall Survival from start of relapsed/ refractory disease in months [mean (median)]	29.5 (33.2)		28 (21.6)	
Confidence interval (95%)	25.1 – 38.8		14.6 – 50.4	
Costs				
Total costs [mean (median)]	€81,626 (€72,182)		€52,760 (€36,882)	
Minimum – maximum	€17,793 – €229,783		€748 – €179,571	

[b] *Not part of data collection because this was generally not reported in medical records*

collected on dosage schemes, adverse effects and all hospital resource use. Table 5.1 and 5.2 present relevant findings of the oxaliplatin studies.

In the bortezomib study, patients (n=543) were identified using a trial database (HOVON50 study[130]) for first line treatment. We approached Dutch hospitals for data collection for outcomes research and continued to include hospitals until the desired number of patients who received off-protocol treatment for relapsed or refractory disease had been reached (n=139). Because many patients (49%) were treated in more than one hospital, we had to collect data in 42 hospitals. Using hospital records, detailed data were retrospectively collected on baseline patient characteristics, known prognostic information, types of treatments, dosage schemes, treatment response, time to progression, time till next treatment, adverse effects, survival and all hospital resource use. Table 5.3 presents relevant findings of the multiple myeloma study.

RESULTS

Feasibility to develop evidence on appropriate drug use

Table 5.4 summarises the results regarding the feasibility to develop evidence on different aspects of appropriate drug use.

Treatments (types of treatments and regimes)

In all three studies it was feasible to ascertain the types of treatments used and their regimes using data from hospital records. Both oxaliplatin studies showed that patients were treated in a way that was similar to the regimes used in clinical trials and described in professional guidelines.[92] In contrast, the bortezomib study revealed a high degree of treatment variation.[131] More importantly, treatments differed significantly from those described in both the pivotal registration trial and professional guidelines.

Dosages (treatment dosages and dose modifications)

Details on dosages and dose modifications were well reported in hospital records. However, retrieval of these details required a great deal of time, which significantly reduced the efficiency of data collection. Both oxaliplatin studies showed that the received dosages were comparable to those observed

Table 5.4 Feasibility to develop evidence on appropriate drug use

	Oxaliplatin in stage III colon cancer	Oxaliplatin in metastatic colorectal cancer	Bortezomib in multiple myeloma
Feasibility to use existing databases to identify patients	+	+	+
Feasibility to obtain a complete dataset using hospital records	+	+/–	+/–
Feasibility to develop evidence on:			
treatments	+	+	+
dosages	+	+	+
patients	+	+	+/–
reasons	+	+	–
effects: intermediate and final outcomes	+	+	+/–
effects: safety outcomes	+/–	+/–	+/–
costs	+	+	+

+ = good; +/– = moderate; – = poor

in clinical trials.[92] The bortezomib study showed that patients received lower dosages (13%) and fewer treatment cycles (4 versus 6) compared to patients in the pivotal registration trial.

Patients (baseline patient characteristics)

For the oxaliplatin studies, the cancer registry provided information on age, gender, date of diagnosis, disease stage, and tumour location. Additional data on prognostic baseline characteristics required data from hospital records. For the bortezomib study, the HOVON database provided information on age, gender, date of diagnosis, and disease stage. Other baseline characteristics required data from hospital records. In all three studies it was impossible to compile a complete dataset, including prognostic factors. For example, 13% of serum carcinoembryonic antigen levels (stage III colon cancer) and 40% of performance scores (metastatic colorectal cancer), and 71% of serum β2-microglobulin levels (multiple myeloma) were missing. Nevertheless, based on available baseline characteristics, it seemed like patients treated with oxaliplatin[92] and bortezomib[132] were comparable to trial patients.

Reasons (reasons for starting or stopping a treatment)
The rationale for choosing a particular treatment was often retrievable (74%) from the hospital records in stage III colon cancer. It was possible to determine the most frequent reasons for dose modifications or treatment interruptions in both oxaliplatin studies. In contrast, in the bortezomib study the reasons to start a treatment, reduce its dose or stop a treatment were often not reported.

Effects (health effects)
In both oxaliplatin studies, the cancer registry only provided survival data, whereas hospital records provided data on disease-free survival, adverse effects, and survival. Similarly, in the bortezomib study, the HOVON database provided survival data and hospital records data on treatment response, adverse effects, and survival. However, in all three studies, treatment responses and adverse effects were often not reported using standardised outcome measures (e.g., RECIST or EBMT response criteria, CTC toxicity grading scale). Although these results could probably be estimated using for instance laboratory test results, this lack of data severely limited a retrospective assessment using outcome measures as treatment response and time to progression often found in clinical trials. Lastly, hospital records did not provide any standardised data on quality of life (e.g., QLQ-C30, EQ-5D or SF36).

Costs (costs of a treatment)
In all three studies it was possible to collect data on hospital resource use of individual patients. However, due to feasibility constraints, unit costs for laboratory services were based on a detailed inventory of a subsample of patients. Similarly, in the bortezomib study detailed data collection on concomitant medication was extremely time-intensive. Therefore, detailed data were only collected for a subsample of 18 patients.

Feasibility to estimate incremental cost-effectiveness

Table 5.5 summarises the results regarding the feasibility to estimate incremental cost-effectiveness.

Table 5.5 Feasibility to estimate incremental cost-effectiveness

	Oxaliplatin in stage III colon cancer	Oxaliplatin in metastatic colorectal cancer	Bortezomib in multiple myeloma
Comparability of baseline characteristics between treatment arms	–	+/–	–
Feasibility of using data from everyday practice:			
to correct for bias	–	+/–	–
to identify treatment comparator	+	+	–
to estimate incremental cost-effectiveness	–	–	–
Comparability of eligible everyday practice patients (treated with oxaliplatin/ bortezomib) and clinical trial patients	+	+	+/–
Feasibility of data synthesis:	+	+	–
Feasibility to obtain additional data from the literature on:			
quality of life	+/–	+/–	–
efficacy	+	+	+/–
effectiveness	–	+/–	–
costs	+/–	+/–	–
Feasibility to estimate (using data synthesis):			
internally valid incremental cost-effectiveness	+	+	–
precise incremental cost-effectiveness	+/–	+/–	–
externally valid incremental cost-effectiveness	+	+	–

+ = good; +/– = moderate; – = poor

Case 1: Oxaliplatin in stage III colon cancer

Due to the strong preference of physicians to use oxaliplatin whenever indicated, patients receiving the comparator treatment were significantly different regarding important prognostic factors. To correct for the resulting confounding, different adjustment techniques were applied to the Cox multivariate regression model, such as average covariate adjustment, regression adjustment by propensity score matching, and survival analysis matched on propensity score matching. However, our sample size (n=391) was not powered for this purpose. This, in combination with missing data on prognostic factors, resulted in possibly biased estimates with wide confidence intervals. It was not feasible to estimate incremental cost-effectiveness using only everyday practice data. Therefore, we developed a Markov model to estimate incremental cost-effectiveness. In this model we synthesised effectiveness data from everyday practice

with efficacy data from the pivotal registration trial. Patients were categorised "eligible" or "ineligible", depending on whether the patients fulfilled the trial eligibility criteria. Ineligible patients (18%) had a worse prognosis compared to eligible patients (82%), but trial patients and eligible case study patients had similar two-year disease-free survivals (80% vs 78%). Effectiveness of the comparator was modelled using trial results. All costs were based on the case study. Applying scenario analyses, incremental cost-effectiveness ratios ranged from €8,247 to €12,289 per quality adjusted life year. Sensitivity analyses of input parameters and model assumptions produced little differences, supporting robustness of the results. Data synthesis resulted in internally valid incremental cost-effectiveness estimates generalisable to Dutch everyday practice.

Case 2: Oxaliplatin in metastatic colorectal cancer

As with case 1, patients receiving oxaliplatin were not comparable to patients not receiving oxaliplatin. In this case, the differences in baseline prognosis were less pronounced than in stage III colon cancer, but correction for confounding was hindered by missing values. Although not performed, modelling evidence from the literature with evidence from the case study would have been feasible.

Case 3: Bortezomib in relapsed or refractory multiple myeloma

Rapid developments in treatment for multiple myeloma resulted in great heterogeneity. Patients treated with bortezomib were not comparable to other patients regarding prognostic factors. It was impossible to identify a single treatment comparator; more than 10 drugs were given in more than 20 different combinations. Therefore, our comparator included any treatment besides bortezomib. Similar to the oxaliplatin cases, different adjustment techniques were applied to the Cox multivariate regression model to obtain a valid overall survival estimate. However, none succeeded in correcting for the observed confounding. New evidence from extended follow-up and other trials comparing different treatments and combinations became available. No information was published on treatment-related costs or quality of life. The great heterogeneity caused by many treatment arms made it impossible to develop a feasible model to estimate incremental cost-effectiveness.

DISCUSSION

We investigated the feasibility of different aspects of outcomes research in oncology. Our results show that the degree of feasibility depends on both the aspect and treatment indication. To our knowledge, this is the first feasibility study of outcomes research in oncology that is based on empirical evidence.

Based on theoretical expectations and expert opinion, the lack of a randomised controlled setting is one of the major concerns.[15, 17, 18, 133] As expected, our results confirm that heterogeneity resulted in incomparable patient groups and the inability to correct for confounding. Therefore, it was not possible to estimate incremental cost-effectiveness only using everyday practice data. However, our results also show that it may still be feasible to obtain internally valid and generalisable incremental estimates by synthesising everyday practice data with trial data, provided that everyday practice patients fulfil the eligibility criteria of trials in which these drugs were tested. Furthermore, our results confirm that current databases do not provide sufficient information.[35, 37-40] The need for additional data required the retrieval and scrutiny of hospital records. Regarding applicability of outcome measures,[8, 12] our results show that measures used in clinical trials are susceptible to bias due to missing data and the lack of standardisation in their reporting in hospital records. The choice of relevant outcome measures depends on the disease. For example, survival is often the primary outcome measure in oncology, but not in diseases such as rheumatoid arthritis or COPD where quality of life is more relevant. Moreover, timeliness[9, 11, 41, 124] can differ per drug and disease. While a three-year time frame was sufficient for the oxaliplatin studies, the bortezomib study revealed that treatment advances limited the relevance of the gathered evidence with such a time frame. The challenge of generalisability[42] might be of lesser concern. In our studies, the ability to select representative samples (e.g., by means of the cancer registry), was a key to ensuring generalisability.

The feasibility of outcomes research also depends on its study design. The main limitation of our case studies was the use of retrospective research designs. As a consequence, we faced a great deal of important missing information. A prospective design, using a registry, would offer greater control over data collection as well as the opportunity to collect data on quality of life. However, in many prospective designs, including registries, data are still retrospectively

collected and rely on information provided by others (e.g. physicians, research assistants). Moreover, such a design would still not solve the issue of randomisation. Although a pragmatic trial would be a solution for this, these trials are often impossible due to ethical or feasibility considerations.

Our study was only based on three case studies. However, we believe that our findings can be extended to other oncological diseases. We intentionally selected different indications in cancer reflecting different types of disease populations (small versus large), expectations regarding practice variation (small versus large), and relevant outcome measures (intermediate versus final endpoints).

Our study provides important insight into the implementation of evidence development schemes. We believe that data from everyday practice results in valuable evidence, addressing uncertainties arising from the gap between clinical trials and everyday practice. Above all, it is essential to have a comprehensive understanding of the disease and the treatment effect and this requires interdisciplinary collaboration.

Active interdisciplinary collaboration will result in an enhanced research design focusing on feasible objectives for a particular treatment in a specific indication. It will also reduce problems with missing information and lack of standardisation in reporting. Because current databases do not provide sufficient information, patient registries can offer an opportunity to build new research infrastructures. Although patient registries cannot resolve all issues, if they are used by an active interdisciplinary collaborative research group, they could increase efficiency of data collection and help to reduce issues of generalisability, incomparability of patient groups, missing information, and lack of standardisation in reporting. For orphan drugs, international registries may be the best means to obtain a sufficient amount of evidence (e.g., Pompe Registry.[134] Furthermore, registries can also be used to monitor and improve quality of care beyond outcomes research.

In conclusion, our results show that it is feasible to generate evidence about drug use in everyday oncology practice. For some aspects of appropriate drug use, this will require improvements in reporting in hospital records or compiling data in registries. The feasibility to estimate incremental cost-effectiveness depends on the drug and its indication. We believe that in most circumstances it is inevitable to synthesise data to obtain valid and precise estimates. However,

it is essential to realise that for some drugs and indications, it may sometimes be impossible to estimate sufficiently valid and precise incremental cost-effectiveness.

In the end, the generation of more evidence will improve the quality of decisions made by both payers and physicians. This, in turn, can improve quality of care and enhance the efficient allocation of resources and thereby help to ensure long-term sustainability of healthcare systems.

Chapter 6

A propensity to get it right: A Monte Carlo simulation study comparing statistical methods to obtain correct cost-effectiveness estimates in observational studies

Submitted

Chantal van Gils*
Lucas Goossens*
Maiwenn Al
Miriam Koopman
Carin Uyl- de Groot
Ken Redekop

*Both authors contributed equally to this manuscript

SUMMARY

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), weighting by the odds and double robustness, each with several specifications. Methods: Adjustment approaches were compared using Monte Carlo simulations. In each simulation, four differently confounded samples were drawn from a synthesised population. Incremental survival time and costs were calculated using the results of Weibull survival and generalised 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 accuracy (proportion of simulated results within acceptable distance from the true values) and bias (systematic deviation from the true effect). Results: In estimates of the average treatment effect in the treated (ATT), kernel PS matching and weighting by the odds led to similar results, whereas 1-to-1 PS matching performed somewhat less well. Regarding average treatment effects for the sample as a whole (ATE), double robustness, IPTW and regression without a PS method had the best accuracy and the least bias. 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 performed the 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.

INTRODUCTION

Randomized 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, who represent the population that will receive the therapy. As an alternative to RCTs, the application of observational data allows investigators to estimate the (cost)-effectiveness of treatment in daily practice, which should result in evidence that is more relevant to policy makers. However, this approach has its own challenges. One key concern in observational studies is that treatment is not assigned randomly; individual patients often receive a particular treatment for specific reasons. This phenomenon will likely 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 to remove confounding. The estimate of the treatment effect is adjusted by taking the effect of additional baseline covariates into account. More recently, methods based on a so-called propensity score (PS) have become increasingly popular. The PS is defined as a patient's 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,[21, 23, 135] 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 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 the PS method alone is insufficient and 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.[136, 137] Another approach in which PS and a fully specified regression model are combined is called double robustness.[20]

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.[21, 22, 24] Although examples of the application of PS methods also exist in cost-effectiveness analysis,[25-34] 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 extensively. Kreif et al. compared the ability of three methods to estimate subgroup effects in cost-effectiveness on observational data.[138]

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.[139, 140] 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 synthesised. 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.

METHODS

All methods described in this section are summarised in a flowchart (Figure 6.1).

Source population with potential outcomes

We performed a simulation study on a synthesised 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[110, 141] .

The synthesising process was based on the Neyman-Rubin-Holland causal model,[142-144] 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 synthesised 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.

Synthesis of survival time and healthcare expenditure

A detailed description of the synthesising process can be found in Appendix 6A. In short, survival time was assigned based on a Weibull survival regression of

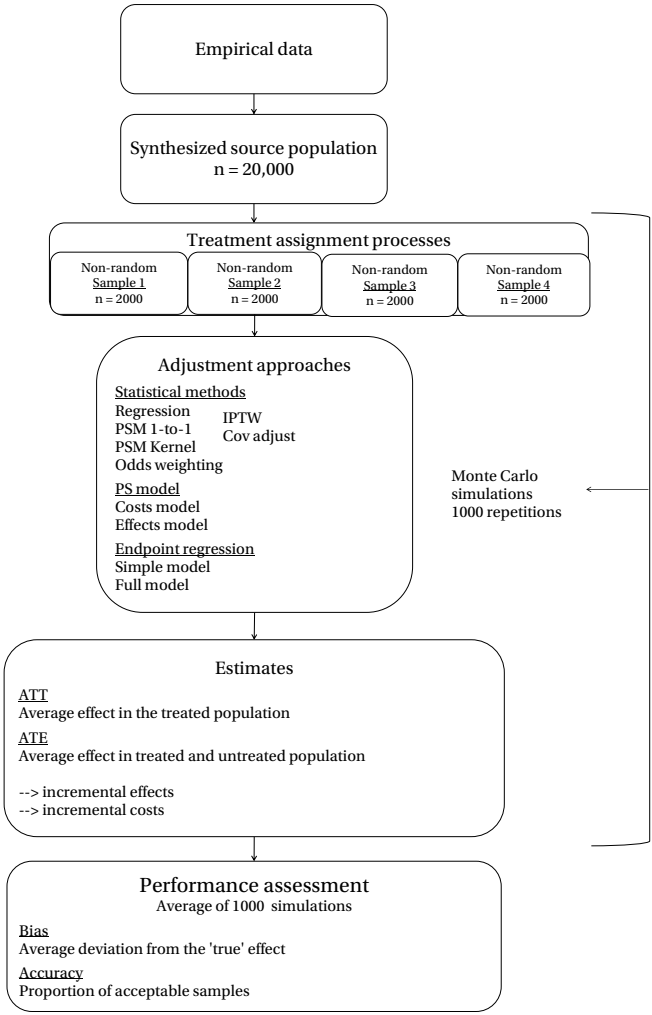


Figure 6.1 Flowchart of the method steps

the empirical trial data. The regression results were combined with the synthesised 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.

A generalised linear model with a power link was fit on the cost data in order to estimate a model with which data could be synthesised. 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 6B.

The mean distribution of the baseline characteristics and treatment effects for samples resulting from these processes are summarised in Table 6.1. The first

Table 6.1 Drawing of samples with non-random treatment assignment

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

LDH, lactate dehydrogenase level in blood; AF, alcalic phosphatase level in blood; WBC, white blood cell count

treatment assignment process was based on covariates that were only in the synthesising model for costs and not in the model 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.

Monte Carlo simulations

For each of the four treatment assignment processes as well as for randomization, 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.

Treatment effects

The 'true' individual treatment effect was defined as the difference between the potential outcomes for each treatment in each patient. Outside simulation studies, this difference is usually not observable. The average of these effects per treatment group can be estimated.

In this study, two conceptually different treatment effects were applied and estimated: the average treatment effect (ATE) and the average treatment effect in the treated (ATT). The ATE is defined as the mean difference in outcomes between the hypothetical situation in which the entire sample had been treated and in which the entire sample had not been treated. In contrast, the ATT is focused on patients who actually received the treatment. It is the average difference between the actual outcomes and the potential outcomes if these patients had not been treated. 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.

Statistical methods

Three methods to estimate an ATT were compared, as were six methods to estimate the ATE.

(1) *PSM 1-to-1* for ATT: PS matching (with replacement and common support requirement). Propensity scores were calculated by fitting a probit model with

being treated as the dependent variable. 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.

(2) *PSM kernel* for ATT: 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.[145] The Epanechnikov kernel was used as a weighting function.[146] After the matching procedure, analysis was similar to the previous method.

(3) *Weighing by the odds* for ATT. Treated patients were assigned a weight of 1, while the untreated patients were weighted by their odds of being treated. These odds were based on the estimated PS. After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

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

(5) *IPTW* for ATE: inverse probability of treatment weighting.[147, 148] 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). For treated patients, these were stabilised by multiplying them with the unconditional probability of receiving the treatment, which equals the proportion of patients that were treated. For untreated patients, the weights were multiplied by the unconditional probability of not being treated.[149] After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(6) *Double robustness* for ATE. This technique combines regression and weighting by the PS in one equation.[20] Results should be unbiased if either the regression model or the propensity-score model is correct.

(7) *PSM 1-to-1* for ATE. After the previous procedure for PSM- 1-to-1 for ATT, a new matching procedure was performed, which was based on untreated patient. For each of them, the closest matching treated patient was selected, based on their PS. Estimates were calculated for both matching procedures.

These preliminary estimates were then weighted by the proportion of treated and untreated patients in the sample in order to find the final estimates.

(8) *PSM kernel* for ATT. This method is equal to the previous one, except that Kernel matching instead of 1-to-1 matching was performed.

(9) *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 synthesising survival time. Costs were analysed in a generalised 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 synthesising cost data.

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 synthesising 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:

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

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.

Weibull and GLM models were also used to calculate unadjusted results, which involved not taking any 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 balance

For all PSM, IPTW and weighting by the odds methods, the post-matching or post-weighting balance of covariates was assessed. A covariate was considered balanced across treatment groups if the standardised difference between means or proportions was less than 0.25.The standardised difference was calculated by dividing the absolute difference by the pooled standard deviation of that variable. Results from a method on a certain sample were included in the final assessment of performance only if balance had been achieved on all covariates.

Assessment of performance

The performance of each method and specification was assessed by calculating the accuracy of its estimates of the cost-effectiveness ratio. This accuracy was defined as the proportion of samples with an effect estimate within an acceptable range around the true effect, which was set at +/- €5000. Next to this, accuracy was calculated by for effectiveness and costs estimates, with ranges of +/- 20%.

Additionally, the bias of the results was calculated for each method and specification. 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.

The accuracy and bias results were averaged over all non-random treatment assignment processes.

Table 6.2 Baseline characteristics, health, costs and cost-effectiveness outcomes

	Source population	Non-random treatment assignment process							
		1		2		3		4	
Untreated (U)/ Treated (T)	U=T	U	T	U	T	U	T	U	T
Baseline Characteristics									
Female	38%	46%	29%	32%	46%	39%	36%	42%	33%
Age, mean	61	66	56	65	55	61	61	66	56
Performance Score > 0	46%	50%	47%	52%	35%	39%	46%	53%	37%
Abnormal AF	63%	57%	65%	69%	55%	54%	68%	67%	62%
WBC, mean	89	90	91	92	86	84	89	92	89
Number of metastases > 1	55%	55%	55%	55%	58%	37%	60%	49%	61%
Unresected primary tumor	37%	37%	37%	37%	37%	23%	38%	33%	41%
Abnormal LDH	45%	41%	48%	47%	40%	20%	51%	38%	54%
Health outcomes									
Overall survival									
Mean (days)	531 664	532	664	496	724	634	630	541	653
Unadjusted incremental effect		132		228		-4		113	
True incremental effect ATT	133	133		145		126		130	
True incremental effect ATE	133	133		135		142		133	
Cost outcomes (euro)									
Total costs									
Mean	23,016 34,612	21,292	35,873	21,200	35,790	25,864	33,752	21,625	34,991
Unadjusted incremental costs		14,580		14,590		7,888		13,366	
True incremental costs ATT	11,596	11,040		10,182		11,460		10,742	
True incremental costs ATE	11,596	11,589		11,406		11,822		11,658	
Cost-effectiveness outcomes (Euro)									
Unadjusted mean ICER		41,570		23,595		-119,962		45,543	
Lower limit		26,779		18,088		-223,000,000		27,055	
Upper limit		78,185		34,798		115,000,000		153,863	
True mean ICER ATT	31,845	30,365		25,738		33,355		30,211	
Lower limit		22,522		18,332		25,012		22,104	
Upper limit		39,195		32,794		41,522		40,293	
True mean ICER ATE	31,845	31,790		30,978		30,420		32,144	
Lower limit		26,169		25,892		25,899		27,099	
Upper limit		38,058		36,022		36,587		37,741	

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

RESULTS

Baseline characteristics and outcomes

Table 6.2 summarises the distribution of the baseline characteristics and health, costs and cost-effectiveness outcomes in the synthesised source population and the mean results 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.

Table 6.3 shows the percentages of iterations where the baseline characteristics were balanced after adjustment via a propensity score model. Generally, a high percentage of the iterations was balanced, although up to 40% of the iterations was not balanced in treatment assignment process 4. Balance was more easily reached in the PS effects models compared to PS cost models. This was to be expected given the greater number of variables that needed to be balanced in the PS cost models.

Table 6.3 Percentage of Iterations where balance was achieved

		Non-random Treatment assignment process			
Method	PS model	1	2	3	4
PSM 1-to-1 -ATT	Effects	95%	95%	93%	75%
PSM 1-to-1 -ATT	Costs	97%	96%	96%	69%
PSM kernel -ATT	Effects	100%	100%	100%	87%
PSM kernel -ATT	Costs	100%	100%	100%	81%
Odds Weighting	Effects	100%	100%	100%	87%
Odds Weighting	Costs	88%	93%	100%	60%
IPTW	Effects	100%	100%	99%	98%
IPTW	Costs	99%	99%	99%	90%
PSM 1-to-1 -ATE	Effects	100%	100%	97%	94%
PSM 1-to-1 -ATE	Costs	99%	100%	95%	92%
PSM kernel -ATE	Effects	100%	100%	99%	99%
PSM kernel -ATE	Costs	100%	100%	99%	97%

ATT, average treatment effect in the treated patients; ATE, average treatment effect; IPTW,inverse probability of treatment weighting; PS, propensity score; PSM, propensity score matching

Overall performance

The performance of all methods for ATTs and ATEs over all treatment assignment processes is summarised in Table 6.4 and 6.5, respectively. Results are presented in more detail (i.e. per treatment assignment process) in Appendix 6C. Most methods succeeded in improving the accuracy in effectiveness estimates removing most the bias and compared to the unadjusted analyses. The proportion of accurate estimates of the ICER varied greatly (from 33% to 68%) across methods and treatment assignment processes. The same variation occurred for effectiveness and costs.

Table 6.4 ATT results

Method	Endpoint model	PS model	BIAS			ACCURACY		
			Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			41%	32%	95%	37%	10%	29%
PSM 1-to-1	Simple	Effects	5%	14%	34%	54%	62%	49%
PSM 1-to-1	Simple	Costs	4%	12%	15%	54%	65%	51%
PSM kernel	Simple	Effects	6%	15%	21%	67%	63%	51%
PSM kernel	Simple	Costs	4%	13%	18%	66%	67%	55%
Odds Weighting	Simple	Effects	4%	14%	23%	62%	64%	57%
Odds Weighting	Simple	Costs	4%	13%	20%	62%	66%	57%
PSM 1-to-1	Full	Effects	8%	9%	5%	58%	75%	63%
PSM 1-to-1	Full	Costs	10%	10%	6%	56%	74%	63%
PSM kernel	Full	Effects	5%	9%	6%	70%	80%	67%
PSM kernel	Full	Costs	6%	9%	6%	68%	80%	67%
Odds Weighting	Full	Effects	8%	9%	4%	65%	79%	68%
Odds Weighting	Full	Costs	8%	9%	5%	63%	79%	67%

ATT, average treatment effect in the treated patients; PS, propensity score; PSM, propensity score matching

Endpoint regression model

Fully specified endpoint regression generally produced more accurate estimates of the ICER than simple models. This was especially consistent for ATTs. The same patterns were found for costs and effects separately.

Table 6.5 ATE results

Method	Endpoint model	PS model	BIAS			ACCURACY		
			Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			47%	25%	148%	33%	29%	13%
Cov adjust	Simple	Effects	7%	6%	6%	43%	85%	39%
Cov adjust	Simple	Costs	6%	6%	8%	49%	86%	44%
IPTW	Simple	Effects	2%	5%	8%	66%	84%	54%
IPTW	Simple	Costs	1%	3%	4%	64%	84%	54%
PSM 1-to-1	Simple	Effects	2%	12%	18%	58%	63%	43%
PSM 1-to-1	Simple	Costs	1%	8%	132%	57%	65%	45%
PSM kernel	Simple	Effects	3%	12%	17%	66%	69%	45%
PSM kernel	Simple	Costs	3%	7%	11%	66%	73%	49%
Regression	Full		1%	8%	10%	78%	83%	65%
Cov adjust	Full	Effects	11%	10%	13%	48%	80%	33%
Cov adjust	Full	Costs	11%	10%	14%	54%	81%	35%
IPTW	Full	Effects	3%	7%	7%	70%	84%	65%
IPTW	Full	Costs	3%	7%	7%	67%	82%	64%
Double R	Full	Effects	2%	8%	11%	66%	86%	62%
Double R	Full	Costs	1%	1%	3%	62%	92%	66%
PSM 1-to-1	Full	Effects	2%	14%	11%	61%	67%	41%
PSM 1-to-1	Full	Costs	2%	15%	12%	63%	66%	41%
PSM kernel	Full	Effects	2%	14%	13%	71%	72%	46%
PSM kernel	Full	Costs	2%	14%	13%	70%	70%	44%

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

With regards to bias, the differences between full and simple endpoint models were less clear. However, in ATT estimates of the ICER, full models performed consistently better

Propensity-score model

Models that included the full combination of covariates (cost models for the propensity score) performed at least as well as the effect models at estimating accurate ICERs. This was most pronounced for some treatment assignment

processes 2 and 4 (see Appendix 6C), where they led to equally good estimates of incremental effects and less biased estimates of incremental costs.

Statistical method

For ATTs, PSM kernel and weighting by the odds led to substantially more accurate estimates than PSM 1-to-1. These differences were not observed for bias.

For ATEs, three methods (IPTW, double robustness and conventional regression) achieved markedly more accurate ICER estimates than the other three (PS covariate adjustment and the two matching methods). These differences were observed to a lesser extent for bias.

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 synthesised dataset were estimated using conventional regression, 1-to-1 PS matching, kernel PS matching, inverse probability-of-treatment weighting and double robustness, weighting by the odds and PS covariate adjustment. 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, all methods failed to achieve accurate estimates in a substantial proportion of samples, although calculations were based only on samples that were balanced after applying a PS method. For incremental cost-effectiveness ratios, this was even more pronounced than for estimates of incremental costs and incremental effects.

Second, estimates were better when propensity-score methods were followed by a fully specified regression adjustment.

Third, 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.[139, 140]

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'.[21]

Fourth, for estimating ATTs, kernel PS matching and weighting by the odds led to similar results, whereas 1-to-1 PS matching performed somewhat less well.

For ATEs, double robustness and IPTW and conventional regression had the best performance. The most widely used method, PS covariate adjustment achieved the least accurate results.

Regression versus PS methods

To further explain differences in PS methods versus conventional regression we want to emphasise that 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.[150] The problem of confounding 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, one can verify 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 is not possible at all in conventional regression.

The second advantage of PS methods is in its ability to reduce the dimensionality problem. The possible number of covariates in regression is not limitless and is restricted by the number of subjects or events. A ratio of at least 10 subjects or events per independent variable has been suggested.[151] The use of PS, which provide a scalar summary of the covariate information, does not have this limitation, since after successful propensity-score adjustment, the only required covariate in the final analysis model would be the treatment variable.

A third advantage, which limits the second advantage described above, is that propensity-score methods can be used to 'pre-process' 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.[136, 137] This approach, which does not equal the double robustness method, does require that several covariates are included in the model.

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 and cost-effectiveness.[138, 152, 153] 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,[154] 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.[24] 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 accuracy 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 synthesising as well as analysing the data. Instead, we chose the – rarely used - complementary log–log model for treatment assignment in the synthesising process, while probit models were applied to estimate the PS.

Another limitation is that we did not apply censoring in our synthesised 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,[21] 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.[152] In that case, kernel matching performed well when it was combined with regression. Without the added regression, Mahalanobis matching had the best results.

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, kernel PS matching methods and weighting by the odds are good candidates. If the outcome of interest is an ATE, inverse probability weighting, double robustness and conventional regression 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 accurate results in all circumstances, it is recommended that several methods are applied and compared in sensitivity analyses. If different methods lead to conflicting results, one should report this and thereby make this structural uncertainty transparent.

APPENDICES

6A Synthesis of source population

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 synthesised 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 synthesised 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.

Synthesising costs

A generalised linear model was fit on the cost data in order to estimate a model with which data could be synthesised. 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 synthesised 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.

6B 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 synthesised patients. In the first non-random treatment assignment process (see Table 6.1), 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 sites (>1). The fourth process contained the variables from the third process plus gender, age, elevated AF and elevated WBC.

6C ATT and ATE results

6C1 ATT results for non-random treatment assignment process 1

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			0%	32%	36%	78%	6%	13%
PSM 1-to-1	Simple	Effects	3%	14%	71%	56%	62%	44%
PSM 1-to-1	Simple	Costs	1%	8%	6%	52%	73%	51%
PSM kernel	Simple	Effects	3%	16%	24%	68%	63%	43%
PSM kernel	Simple	Costs	1%	9%	13%	64%	77%	59%
Odds Weighting	Simple	Effects	3%	14%	16%	64%	65%	54%
Odds Weighting	Simple	Costs	3%	9%	10%	63%	75%	62%
PSM 1-to-1	Full	Effects	5%	5%	4%	63%	82%	64%
PSM 1-to-1	Full	Costs	8%	7%	2%	58%	80%	65%
PSM kernel	Full	Effects	2%	5%	6%	76%	89%	70%
PSM kernel	Full	Costs	5%	6%	4%	70%	86%	69%
Odds Weighting	Full	Effects	6%	5%	1%	70%	87%	74%
Odds Weighting	Full	Costs	7%	6%	2%	67%	85%	69%

ATT, average treatment effect in the treated patients; PS, propensity score; PSM, propensity score matching

6C2 ATT results for non-random treatment assignment process 2

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			55%	43%	8%	1%	0%	93%
PSM 1-to-1	Simple	Effects	10%	23%	21%	47%	39%	57%
PSM 1-to-1	Simple	Costs	10%	26%	21%	49%	36%	51%
PSM kernel	Simple	Effects	11%	24%	18%	60%	39%	59%
PSM kernel	Simple	Costs	10%	28%	25%	60%	32%	52%
Odds Weighting	Simple	Effects	10%	22%	16%	54%	42%	61%
Odds Weighting	Simple	Costs	9%	27%	23%	51%	33%	53%
PSM 1-to-1	Full	Effects	9%	19%	15%	57%	51%	63%
PSM 1-to-1	Full	Costs	8%	18%	16%	57%	55%	62%
PSM kernel	Full	Effects	5%	18%	17%	72%	53%	59%
PSM kernel	Full	Costs	5%	17%	17%	72%	58%	60%
Odds Weighting	Full	Effects	7%	18%	14%	66%	54%	65%
Odds Weighting	Full	Costs	5%	16%	14%	67%	58%	65%

ATT, average treatment effect in the treated patients; PS, propensity score; PSM, propensity score matching

6C3 ATT results for non-random treatment assignment process 3

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			99%	30%	288%	0%	10%	0%
PSM 1-to-1	Simple	Effects	1%	8%	12%	66%	82%	54%
PSM 1-to-1	Simple	Costs	2%	8%	9%	65%	80%	57%
PSM kernel	Simple	Effects	1%	8%	12%	79%	84%	61%
PSM kernel	Simple	Costs	1%	8%	12%	80%	84%	61%
Odds Weighting	Simple	Effects	1%	8%	10%	78%	85%	63%
Odds Weighting	Simple	Costs	1%	8%	10%	78%	84%	63%
PSM 1-to-1	Full	Effects	10%	7%	0%	55%	88%	61%
PSM 1-to-1	Full	Costs	12%	6%	2%	55%	85%	63%
PSM kernel	Full	Effects	8%	6%	0%	67%	92%	72%
PSM kernel	Full	Costs	8%	6%	0%	68%	92%	71%
Odds Weighting	Full	Effects	8%	6%	0%	67%	93%	70%
Odds Weighting	Full	Costs	9%	7%	0%	64%	91%	70%

ATT, average treatment effect in the treated patients; PS, propensity score; PSM, propensity score matching

6C4 ATT results for non-random treatment assignment process 4

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			10%	25%	46%	68%	25%	10%
PSM 1-to-1	Simple	Effects	7%	11%	31%	49%	64%	43%
PSM 1-to-1	Simple	Costs	4%	8%	22%	50%	71%	47%
PSM kernel	Simple	Effects	8%	12%	29%	59%	67%	40%
PSM kernel	Simple	Costs	6%	8%	21%	59%	74%	50%
IPTW	Simple	Effects	1%	12%	48%	53%	63%	51%
IPTW	Simple	Costs	4%	8%	38%	57%	72%	51%
PSM 1-to-1	Full	Effects	9%	7%	2%	54%	78%	63%
PSM 1-to-1	Full	Costs	10%	8%	4%	56%	76%	62%
PSM kernel	Full	Effects	6%	6%	3%	65%	86%	69%
PSM kernel	Full	Costs	7%	7%	3%	64%	83%	68%
IPTW	Full	Effects	10%	7%	2%	56%	83%	65%
IPTW	Full	Costs	10%	8%	5%	55%	80%	66%

ATT, average treatment effect in the treated patients; PS, propensity score; PSM, propensity score matching

6C1 ATE results for non-random treatment assignment process 1

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			1%	26%	31%	73%	22%	23%
Cov adjust	Simple	Effects	2%	7%	11%	42%	84%	42%
Cov adjust	Simple	Costs	4%	9%	15%	48%	81%	36%
IPTW	Simple	Effects	3%	6%	7%	68%	84%	56%
IPTW	Simple	Costs	3%	3%	2%	64%	79%	52%
PSM 1-to-1	Simple	Effects	0%	9%	15%	56%	69%	49%
PSM 1-to-1	Simple	Costs	1%	2%	3%	56%	70%	51%
PSM kernel	Simple	Effects	0%	8%	13%	66%	77%	54%
PSM kernel	Simple	Costs	1%	2%	1%	67%	79%	56%
Regression	Full		0%	7%	9%	82%	84%	67%
Cov adjust	Full	Effects	8%	5%	4%	50%	89%	42%
Cov adjust	Full	Costs	10%	3%	8%	57%	91%	46%
IPTW	Full	Effects	3%	5%	5%	76%	88%	69%
IPTW	Full	Costs	2%	5%	5%	70%	85%	67%
Double R	Full	Effects	2%	15%	15%	69%	73%	58%
Double R	Full	Costs	2%	2%	0%	64%	93%	64%
PSM 1-to-1	Full	Effects	1%	14%	9%	66%	65%	45%
PSM 1-to-1	Full	Costs	0%	16%	12%	66%	61%	39%
PSM kernel	Full	Effects	0%	14%	11%	76%	71%	48%
PSM kernel	Full	Costs	1%	15%	13%	74%	66%	43%

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

6C2 ATE results for non-random treatment assignment process 2

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			69%	28%	24%	0%	13%	11%
Cov adjust	Simple	Effects	10%	7%	2%	43%	85%	34%
Cov adjust	Simple	Costs	9%	1%	2%	52%	91%	49%
IPTW	Simple	Effects	3%	7%	5%	64%	80%	52%
IPTW	Simple	Costs	1%	1%	3%	61%	85%	54%
PSM 1-to-1	Simple	Effects	2%	16%	12%	58%	57%	37%
PSM 1-to-1	Simple	Costs	3%	10%	10%	57%	67%	43%
PSM kernel	Simple	Effects	4%	14%	15%	66%	65%	42%
PSM kernel	Simple	Costs	4%	9%	9%	65%	76%	51%
Regression	Full		1%	5%	6%	79%	91%	71%
Cov adjust	Full	Effects	11%	12%	13%	48%	77%	33%
Cov adjust	Full	Costs	9%	12%	15%	55%	76%	35%
IPTW	Full	Effects	2%	6%	6%	67%	86%	64%
IPTW	Full	Costs	2%	7%	8%	63%	83%	63%
Double R	Full	Effects	3%	8%	7%	64%	86%	61%
Double R	Full	Costs	1%	0%	4%	61%	92%	68%
PSM 1-to-1	Full	Effects	1%	12%	8%	61%	70%	43%
PSM 1-to-1	Full	Costs	1%	12%	8%	59%	73%	46%
PSM kernel	Full	Effects	1%	11%	10%	69%	77%	52%
PSM kernel	Full	Costs	1%	11%	9%	68%	77%	52%

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

6C3 ATE results for non-random treatment assignment process 3

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			103%	33%	494%	0%	2%	0%
Cov adjust	Simple	Effects	6%	6%	7%	48%	89%	47%
Cov adjust	Simple	Costs	7%	6%	5%	54%	89%	56%
IPTW	Simple	Effects	0%	4%	7%	72%	93%	59%
IPTW	Simple	Costs	0%	4%	7%	72%	92%	60%
PSM 1-to-1	Simple	Effects	0%	18%	23%	66%	55%	43%
PSM 1-to-1	Simple	Costs	0%	18%	22%	64%	53%	42%
PSM kernel	Simple	Effects	3%	17%	25%	71%	56%	38%
PSM kernel	Simple	Costs	3%	17%	25%	71%	56%	38%
Regression	Full		2%	14%	15%	75%	70%	59%
Cov adjust	Full	Effects	14%	13%	20%	46%	77%	25%
Cov adjust	Full	Costs	13%	13%	19%	54%	78%	28%
IPTW	Full	Effects	3%	11%	10%	75%	79%	66%
IPTW	Full	Costs	3%	12%	10%	75%	79%	65%
Double R	Full	Effects	0%	1%	3%	70%	99%	75%
Double R	Full	Costs	0%	0%	3%	70%	99%	76%
PSM 1-to-1	Full	Effects	5%	15%	15%	63%	72%	39%
PSM 1-to-1	Full	Costs	5%	15%	16%	70%	70%	40%
PSM kernel	Full	Effects	4%	15%	16%	72%	75%	42%
PSM kernel	Full	Costs	4%	15%	16%	72%	73%	42%

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

6C4 ATE results for non-random treatment assignment process 4

			BIAS			ACCURACY		
Method	Endpoint model	PS model	Effectiveness results	Costs results	ICER results	Effectiveness results	Costs results	ICER results
Unadjusted			15%	15%	42%	59%	77%	18%
Cov adjust	Simple	Effects	8%	5%	3%	40%	84%	34%
Cov adjust	Simple	Costs	3%	6%	11%	43%	83%	35%
IPTW	Simple	Effects	1%	3%	11%	61%	81%	48%
IPTW	Simple	Costs	1%	3%	4%	57%	80%	48%
PSM 1-to-1	Simple	Effects	4%	6%	22%	53%	71%	44%
PSM 1-to-1	Simple	Costs	2%	1%	493%	50%	70%	45%
PSM kernel	Simple	Effects	4%	7%	17%	61%	77%	46%
PSM kernel	Simple	Costs	3%	1%	9%	60%	79%	51%
Regression	Full		0%	6%	8%	75%	89%	65%
Cov adjust	Full	Effects	12%	11%	13%	47%	78%	30%
Cov adjust	Full	Costs	12%	11%	15%	52%	78%	34%
IPTW	Full	Effects	4%	6%	6%	64%	83%	62%
IPTW	Full	Costs	4%	6%	6%	61%	81%	61%
Double R	Full	Effects	1%	8%	18%	60%	86%	55%
Double R	Full	Costs	1%	4%	5%	55%	85%	54%
PSM 1-to-1	Full	Effects	2%	16%	12%	57%	60%	36%
PSM 1-to-1	Full	Costs	3%	16%	13%	57%	61%	37%
PSM kernel	Full	Effects	2%	15%	14%	66%	67%	41%
PSM kernel	Full	Costs	3%	16%	14%	66%	65%	40%

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

Chapter 7

A propensity to be precise: A simulation study comparing statistical methods to obtain reliable cost-effectiveness estimates in observational studies

Submitted

Lucas Goossens*
Chantal van Gils*
Maiwenn Al
Miriam Koopman
Carin Uyl- de Groot
Ken Redekop

*Both authors contributed equally to this manuscript

SUMMARY

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), weighting by the odds and double robustness, each with several specifications. Methods: A synthesised 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 and weighting by the odds. 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.

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.[155, 156]

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.[157, 158] This has been shown to enhance the statistical power and precision of the estimates.[159] According to Senn, ignoring balanced prognostic factors in the analysis is 'a grave mistake'.[160]

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.[21, 22, 24, 139] 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,[161, 161] 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.[21, 139, 161, 162] 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[150]: 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.

METHODS

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

Synthesis of source population

We performed a simulation study on a synthesised source population of 20,000 patients. The synthesising 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.[110, 141, 163] The synthesised 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

Three methods for estimating the average treatment effect in the treated (ATT) were compared, as were four methods for estimating the average treatment effect (ATE). The latter is the mean difference in outcomes between the hypothetical situations in which the entire sample had been treated and in which the entire sample had not been treated. The ATT is focused on patients who actually had the treatment. The ATE and ATT may be different, since the treatment may not be as effective in all subgroups.

(1) *Conventional regression* for ATE. A fully specified Weibull regression model was used to estimate the survival gains. The covariates were the prognostic factors that were used for synthesising survival time. Costs were analysed in a generalised 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) *PS covariate adjustment* regression analysis for ATE. Propensity scores were used as covariates in the Weibull model and GLM for survival and costs.

(3) *IPTW* for ATE: inverse probability of treatment weighting.[147, 148] 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). For treated patients, the weights were stabilised by multiplying them with the unconditional probability of receiving the treatment, which equals the proportion of patients that were treated. For untreated patients, the weights were multiplied by the unconditional probability of not being treated.[164] After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

(4) *Double robustness* for ATE. This technique combined the Weibull regression for effects or GLM for costs with weighting by the PS in one equation.[20] Results should be unbiased if either the regression model or the propensity-score model is correct.

(5) *PSM 1-to-1* for ATT: 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.

(6) *PSM kernel* for ATT: 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.[145] The Epanechnikov kernel was used as a weighting function.[146] After the matching procedure, analysis was similar to the previous method.

(7) *Weighing by the odds* for ATT. Treated patients were assigned a weight of 1, while the untreated patients were weighted by their odds of being treated. These odds were based on the estimated PS. After weighting, a Weibull model was used to analyse survival while a GLM was used to analyse costs.

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 (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.[165, 166] 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:

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

RESULTS

The widths of the confidence intervals and the areas of the confidence ellipses are summarised in Table 7.1. They could be calculated for all samples with n=2000. Seventeen out of the 60 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 7.1.

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.

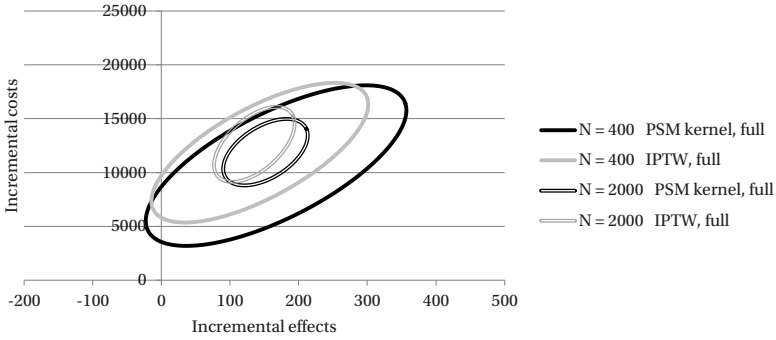


Figure 7.1 95% Confidence ellipses on the CE-plane: 400 versus 2000 (sample2)

Table 7.1 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											
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	44	78	57	96	-[c]	21	33	39	38	33
IPTW	Full	63	71	63	57	-[c]	15	28	23	27	21
Double robustness	Full	49	-[c]	38	69	-[c]	13	41	31	20	27
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
Odds weighting	Simple	83	-[c]	-[c]	69	-[c]	24	46	44	50	45
Odds weighting	Full	106	-[c]	82	53	-[c]	23	36	35	55	47
N = 2000											
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	31	22	23	26	4	9	4	5	7
IPTW	Full	19	30	19	19	17	3	8	3	4	5
Double robustness	Full	19	23	20	18	19	3	7	4	3	5
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
Odds weighting	Simple	32	7	21	34	7	6	1	8	6	1
Odds weighting	Full	28	10	20	31	10	5	1	9	7	1

[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.

Simple versus full endpoint models

In all cases except odds weighting, the confidence ellipse areas for methods with fully specified endpoint models were smaller than those for methods with simple endpoint models. In odd weighting, the confidence ellipse areas for fully specified endpoint models and simple endpoint models were comparable. The results for the widths of most all confidence intervals were consistent with this observation. An illustration is presented in Figure 7.2 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.

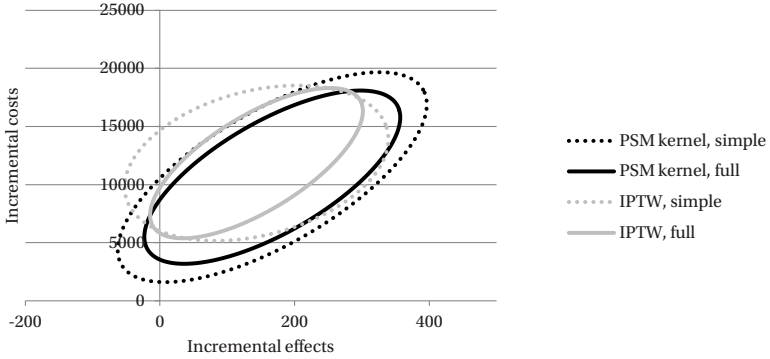


Figure 7.2 95% Confidence ellipses on the CE-plane: full versus simple model (sample 2, n = 400)

ATE methods compared

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. The intervals for IPTW and double robustness were mostly somewhat wider than for conventional regression, but consistently smaller than the intervals for covariate adjustment. These 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.

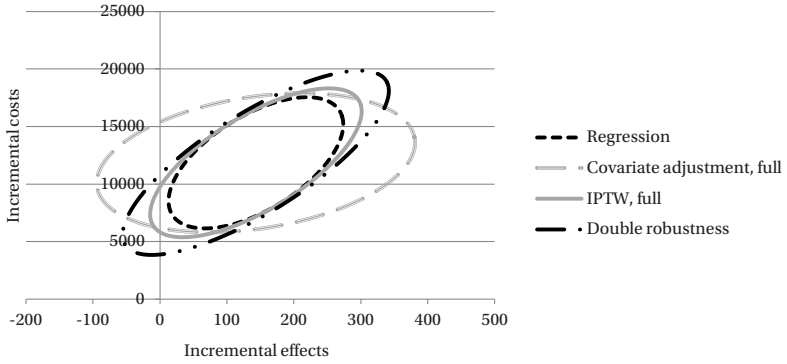


Figure 7.3 95% Confidence ellipses on the CE-plane: ATE methods compared (sample 2, n = 400)

Figure 7.3 shows the ellipses for conventional regression, double robustness covariate adjustment (with full model), IPTW (with full endpoint model) for sample 2 with n=400.

ATT methods compared

The PSM kernel and the odds weighting methods consistently led to smaller ellipse areas and interval widths than the PSM 1-to-1 method. Mostly PSM kernel performed also better than Odds weighting. Figure 7.4 illustrates this by

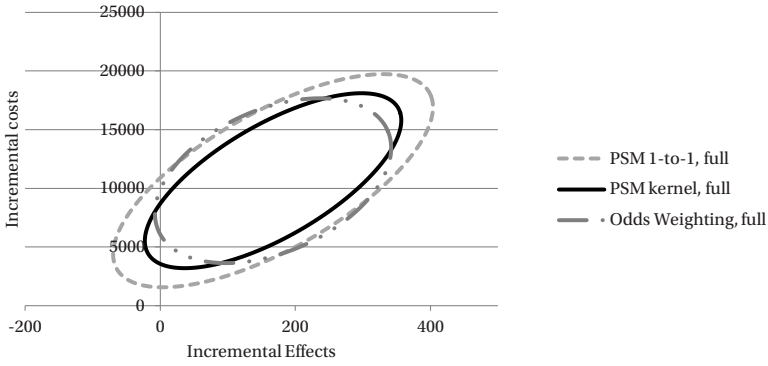


Figure 7.4 95% Confidence ellipses on the CE-plane: ATT methods compared (sample 2, n = 400)

showing the ellipses for both methods with fully specified endpoint models in sample 2 with n=400.

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 randomized 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. The other ATT method, odds weighting, did not perform better than kernel matching.

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.[161]

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.[136] 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.[167] 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 synthesising as well as analysing the data. Instead, we chose the – rarely used - complementary log–log model for treatment assignment in the synthesising process, while probit models were applied to estimate the PS.

Another limitation is that we did not apply censoring in our synthesised 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.[168] 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.[23] 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.

Chapter 8

General discussion

INTRODUCTION

In 2006, the Dutch Healthcare Insurance Board introduced the expensive drugs policy regulation. This policy aimed to ensure undelayed access of promising but expensive hospital medicines by reimbursing hospitals (most of) the drug costs of drugs placed on the 'expensive drug list'. This reimbursement, however, is conditional for four years, after which a re-assessment takes place. The decision about whether or not to continue reimbursement is mainly based on the appropriate use and the cost-effectiveness of the new medicine as measured in Dutch real world practice. The mandatory use of real-world data is new in Dutch policymaking. Debates regarding the feasibility of real-world studies and the added value are ongoing as experience in real-world cost-effectiveness studies is limited. This thesis demonstrates the added value (potential) of conducting real-world studies and addresses several related methodological challenges (pitfalls) in the field of expensive inpatient oncologic drugs.

QUESTIONS ADDRESSED IN THIS THESIS

1. How is oxaliplatin used in the adjuvant treatment of stage III colon cancer in real-world practice and what are the real-world costs, effects and cost-effectiveness?

Oxaliplatin as adjuvant treatment in stage III colon cancer was included in the 'expensive drug list' in early 2005. Uncertainty existed regarding overall survival benefit, real-world appropriate use, real-world (incremental) effectiveness and cost-effectiveness. A retrospective population-based study was performed to evaluate these elements in Dutch daily practice.

The results described in Chapter 2 indicated that there was a rapid adoption of oxaliplatin in the period shortly after it was reimbursed for the adjuvant treatment of colon cancer. In general, real-world practice seemed in good concordance with the RCT-based treatment recommendations. The dosages used in daily practice were comparable to those observed in clinical trials.[53] Consequently, oxaliplatin seems to be used appropriately in daily practice. Regarding the effectiveness in daily practice, patients receiving oxaliplatin in daily practice who fulfilled the MOSAIC study eligibility criteria (82%), had

comparable 2-year DFS and OS outcomes as the MOSAIC patients. Patients who did not fulfil the MOSAIC study eligibility criteria (18%) had lower DFS and OS outcomes. We can conclude that the RCT results are directly applicable to 82% of the real-world patients receiving oxaliplatin. Ineligible patients (18%) had a worse prognosis due to a worse prognostic profile these patients had.

The focus of Chapter 4 was the incremental cost-effectiveness of oxaliplatin in daily practice compared to FL. Real-world incremental effectiveness of oxaliplatin was difficult to determine since patients receiving the comparator treatment, FL, were significantly different from patients receiving FL + oxaliplatin regarding important prognostic factors, like age and comorbidities. These differences are explained by the strong preference of physicians to use oxaliplatin whenever indicated. It was not possible to adjust for these differences using propensity score matching techniques due to insufficient power because of limited overlap in the distribution of baseline prognostic variables between the two populations. Hence, the treatment effect of the RCT was used as proxy for the real-world treatment effect in the 82% that seemed similar to the RCT population. For the 18% patients that were ineligible, the treatment effect was uncertain, but it is expected that oxaliplatin would also be effective in this patient population based on the proven effectiveness of oxaliplatin in metastatic cancer. Next to this real-world evidence, a continuous survival benefit was demonstrated based on the 6-year extended follow-up results of the MOSAIC RCT.[64] This is essential extra information since only median follow-up of 2 years was available from the real-world study. The results from the real-world study supplemented with additional evidence from published literature suggest that the combination of FL + oxaliplatin has added value over the use of FL alone in the real-world adjuvant treatment of stage III colon cancer.

In Chapter 3 the costs of oxaliplatin in daily practice were discussed. The results showed that the mean total per patient ranged from €9,681 (5FU/LV) to €9,736 (capecitabine) for patients on FL only therapy. Adding oxaliplatin to FL significantly increased costs, although the mean total costs for patients receiving FL + oxaliplatin varied considerably depending on the type of FL chosen. For patients receiving CAPOX the costs amounted to €18,361, while for patients receiving FOLFOX these costs were €32,793. The higher costs of FOLFOX are predominantly explained by the high number of hospitalisation days used for the 48 hours administration of FOLFOX. Given the diversity found in trials

and in daily practice, it is questionable whether expensive hospitalisations are necessary for the administration of FOLOX. In this light it should be noted that Dutch hospitals have a financial incentive to hospitalise patients to administer FOLFOX since hospitals are relatively well financially compensated for these hospitalisation days. A potential for substantial cost savings exists if the 48 hour administration of 5FU/LV in the FOLFOX regimen were to take place at home or be replaced by the oral drug capecitabine like in the CAPOX regimen.

The costs presented in Chapter 3 were the sole input for the cost input parameters of the cost-effectiveness model described in Chapter 4. This may have resulted in some bias in the incremental cost estimates through confounding variables. This bias occurs when certain baseline characteristics, like for example age or comorbidities, are associated with both the total treatment costs and the probability to receive a certain treatment. In Chapter 2 it was found that younger age and fewer comorbidities were positively associated with the probability to receive oxaliplatin. Chapter 4 revealed differences in single cost components between older versus younger patients and patients with or without comorbidities. However, the mean total treatment costs were not associated with these baseline characteristics since the differences in single cost components cancelled out on a total cost level. Since the mean total costs are the measure of interest as cost input in the cost-effectiveness model, unadjusted incremental cost estimates are not likely to be biased.

To estimate the real-world cost-effectiveness (described in Chapter 4) we combined trial effectiveness data with evidence from daily practice to preserve both the validity of the RCT and the representativeness of daily practice data. Regarding cost input we showed that it was valid to use just data directly from Dutch practice. Via this approach we implemented a real-world cost-effectiveness study that yields internally valid results that are also generalisable to Dutch clinical practice. The ICERs of the different scenarios ranged from €8,388 to €12,746, were all considered acceptable and support the use of oxaliplatin in the adjuvant treatment of stage III colon cancer. In Chapter 3 we identified a potential for substantial cost-savings when using CAPOX instead of FOLFOX. Chapter 4 demonstrates its impact on cost-effectiveness in a sensitivity analysis. If all patients would receive CAPOX, ICERs of the different scenarios would range from €3,195 to €6,216, which is approximately 50% lower than current practice.

2. What is the feasibility of assessing appropriate drug use and real-world cost-effectiveness in oncologic drugs and what are the most important challenges in light of the Dutch expensive drug policy? How can these challenges best be addressed?

In Chapter 5 the feasibility of real-world cost-effectiveness studies and the most important methodological challenges were discussed. Three different case studies were used as empirical input: i) oxaliplatin in stage III colon cancer, ii) oxaliplatin in metastatic colorectal cancer and iii) bortezomib in patients with multiple myeloma. It was feasible in all case studies to develop evidence on appropriate drug use, including information on clinical effectiveness and costs, based on real-world data. It was also feasible to compare these real-world results to the results of RCTs. For example, the case study were oxaliplatin was studied in metastatic colorectal cancer brought to light that in The Netherlands patients were treated in daily practice with regimens according to results from the RCT in which these regimens were investigated.[141] However, in a substantial amount of patients, standard criteria that are required for the safe administration of these treatments were not fulfilled. These patients had a significantly worse outcome when compared to patients treated with these regimens in daily practice, who did meet these criteria.[163]

However, this feasibility only applied to the treatment with the new expensive drug and not to the comparator treatment. In the three case studies, real-world data did not provide suitable evidence on the comparator treatment, which is essential in the estimation of incremental benefits and costs of the new drug. Therefore, it was not feasible to determine the real-world comparative effectiveness and cost-effectiveness based on data from Dutch real-world practice alone.

The challenge of lacking comparator information played a major role in all case studies, but the reasons for this issue differed between case studies. Reasons could be identified at three levels: 1) treatment heterogeneity, 2) study design, and 3) missing information on baseline prognostic variables.

A high degree of treatment heterogeneity was present in the bortezomib case study. Many different chemical agents were given in different lines as monotherapy or in different combinations, and in different sequences. These treatments (or comparator treatments) also differed significantly from those described in clinical trials. Moreover, guidelines have changed over the years

and now recommend earlier use of bortezomib, leading to dynamic treatment patterns. This heterogeneity made it impossible to identify an appropriate comparator treatment in an appropriate patient group in this case study.

The study design of the adjuvant oxaliplatin case study (see part 1) led to a study population in which oxaliplatin patients differed greatly from the patients receiving the comparator treatment, with respect to their baseline prognostic variables. Patients receiving FL + oxaliplatin were younger and had less comorbidity than patients receiving FL. When the overlap in the distribution of baseline prognostic variables between the two populations is almost zero, no statistical adjustment method will be able to adequately adjust for confounding.[169] This study was designed in such a way that the start of study period corresponded with the addition of oxaliplatin to the expensive medicines list in early 2005. The uptake of oxaliplatin occurred faster than anticipated, which resulted in a low number of comparable control patients included in the study. It should be noted that, in contrast to the RCTs, all case studies involved relatively few patients and a limited duration of follow-up. Due to these design factors it was not possible to estimate the incremental effectiveness and cost-effectiveness of oxaliplatin based on real-world data alone.

Difficulties with missing values on important baseline prognostic values were present in all case studies. For example in the case study where oxaliplatin was evaluated in metastatic colorectal cancer, missing WHO performance scores, which is prognostic for both DFS and the likelihood to receive oxaliplatin, complicated an adequate correction of confounding.

Next to the challenge of lacking comparator information, the absence of a uniform measurement of the outcome of interest further challenged the ability to obtain valid incremental cost-effectiveness estimates. Besides, the collection of necessary data was very time-consuming, since it required the retrieval and detailed examination of hospital records.

Potential solutions for these challenges can be found in two directions: 1) optimise the design of the real-world study in such a way that it directly allows the estimation of real-world incremental effects and costs based on real-world data, or 2) estimate the real-world cost-effectiveness based on the combination of real-world data with other sources of evidence.

A number of comments need to be made regarding the choice to optimise the study design. First, using a prospective study design instead of a retrospec-

tive design can, to some extent, address the issue of missing variables and/or values through greater control over data collection. Second, it is important to identify an optimal timeframe when designing a real-world study. The choice of an optimal study timeframe should not only focus on the time needed to include information on the new drug, but should also be based on its comparator treatment. As we observed in the case study described in part 1, the comparability of the oxaliplatin and FL only treatment groups would have been much better if the starting date of the study would have been before 2005. This would have allowed the use of 'historical' controls treated in a time period prior to the introduction of oxaliplatin. As a consequence, the likelihood of a successful application of a confounding correction technique would increase. Moreover, to facilitate a successful synthesis of different sources of data (as we did by combining real-world data with data from the clinical registration trial) it is important to collect data on all relevant baseline prognostic variables and outcomes, and make sure they are measured in a similar way as is done in trials. It is also recommended to collect data on the in and exclusion criteria of the pivotal RCTs, which allows the selection of 'eligible' and 'ineligible' subgroups of real-world patients.

3. How should propensity-score based adjustment methods be applied in observational cost-effectiveness studies?

Chapters 6 and 7 of this thesis applied several methods to adjust for confounding in real-world cost-effectiveness studies. The methods were compared in their ability to produce valid (Chapter 6) and reliable (Chapter 7) estimates, while different model specifications were considered.

First, 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 patients who actually had the treatment. It is the average difference between the realised outcomes and the potential outcomes if these patients had not been treated The ATE is defined as the mean difference in outcomes between the hypothetical situations in which the entire sample had been treated and in which the entire sample had not been treated. In general, patient populations are heterogeneous. This means that patient characteristics, such as age sex, severity of disease, presence of comorbidities etc., vary be-

tween individuals. These varying patient characteristics can potentially modify the effect of a treatment on outcomes. If this occurs, ATT and ATE are different from each other.

In the context of the Dutch conditional reimbursement policy, the ATT is probably the relevant treatment effect: what is the incremental (cost-) effectiveness in the patient group that has been given the new treatment? In other situations, the ATE may be more relevant. This can be the case, when the treated sample is different from the population of interest.

The conceptual difference between ATT and ATE is often not acknowledged in cost-effectiveness studies. The most frequently 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.[23, 135] Inverse probability-of-treatment weighting and double robustness also estimate this ATE. Weighting by the odds, in contrast, results in an ATT estimate. Matching methods, although able to estimate both ATT and ATE, are usually used to estimate an ATT as well. It is important to be aware of these options and base the choice of method on the consideration which treatment effect reflects the population of interest.

In cost-effectiveness analyses, only one set of propensity scores should be used in the estimation of both cost and effects. This prevents that analyses of costs and effects are performed on different samples of patients. For instance, matching methods would select different sets of control patients depending on the propensity-score model specification. In the decision which parameters to include in the propensity score model, all potential confounders for either the cost or effectiveness endpoint should be carefully considered for inclusion in the propensity score model. The results indicated that 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. This suggests that using the costs PS model is preferred over the effects PS model, and errs on the side of including more rather than fewer.[21]

The findings of the simulation study show the appropriateness of applying a fully specified regression model, including all baseline variables, after the initial application of a PS method. In Chapter 6 it was demonstrated that using a full endpoint regression model after PS adjustment, via 1 to 1 matching,

kernel matching or IPTW, led to less biased estimates of effects, cost and cost-effectiveness than a simple model. This was as hypothesised since adding a full endpoint regression model offers an extra opportunity, on top of the PS model correction, 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. Moreover, in Chapter 6 we demonstrated that adjusting for prognostic covariates in the endpoint model does not only improve validity, but also the precision of the estimates.

Regarding methods for estimating an ATT, no large differences between the performance of 1-to-1 matching with replacement, kernel matching or weighting by the odds, with regards to the validity of ATT estimates. Kernel matching seemed to perform somewhat better. The application of kernel matching or weighting by the odds led to more precise estimates compared to 1 to 1 matching.

Regarding methods for estimating an ATE, both Chapter 6 and 7 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. Although it is the most frequently used PS method, no support was found for applying PS-as-covariate adjustment.[23] Similarly, the results indicate no reason to use matching methods (ATE version) to estimate an ATE. We observed that IPTW with a full endpoint model and double robustness performed best regarding validity, but at the cost of loss of precision. Conventional regression performed well with regards to validity and left the least uncertainty. However, both conventional regression methods as well as covariate adjustment methods have a serious disadvantage since they do not explicitly lead to an RCT-like design of the dataset. As a consequence, the researcher cannot verify whether important covariates are balanced across treatment groups; in other words, whether adjustment has been successful. Next to this, the possible number of covariates that can be included in a conventional 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.[151] PS, which provide a scalar summary of the covariate information, do not have this limita-

tion. After successful propensity-score adjustment, the only required covariate in the final analysis model would be the treatment variable.

In conclusion, 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 IPTW with full endpoint regression or double robustness. If these methods lead to similar results, the superior precision of conventional regression would prevail. If the results are different, then IPTW or DR results are likely to be the most accurate.

Limitations

Some limitations of the studies described in this thesis need to be noted. These limitations reflect both theoretical and practical choices made during the process of writing this thesis.

To start, the main limitation of the case studies in this thesis was the use of retrospective study designs. Retrospective studies look back by using information that has usually already been collected for reasons other than research, such as data in medical records. The major disadvantage of this design is that it is not possible to influence the type of data that are collected. This resulted in gaps in the real-world data collected in the case studies.

For instance, we were unable to collect data on Health Related Quality of Life (HRQOL) and thus unable to draw empirical conclusions about the HRQOL of the patients included in the case studies. Quality of life is an important health outcome and its use in real-world studies to assess real-world cost-effectiveness is strongly recommended. Given the lack of good HRQOL data in patients receiving oxaliplatin in stage III colon cancer, it would have been valuable to collect real-world HRQOL data. We were able to integrate HRQOL data from the literature in our analysis and also found that HRQOL was of very limited importance in the sensitivity analysis. Moreover, despite the Dutch pharmacoeconomic guidelines prescribing the use of a societal perspective[170] we conducted the analyses from a health care sector perspective. This was done for pragmatic reasons; given the dependency on retrospective data, we were unable to collect data on, for instance, productivity losses related to absenteeism, presenteeism and the diminished ability to perform unpaid labour. Given that productivity costs can have a strong impact on cost (-effectiveness) outcomes,[171] this has likely influenced the outcomes of the case studies.

Although retrospective studies do have advantages, especially regarding efficiency of data collection, it is preferable to use prospective designs when setting up a real-world study in the context of conditional reimbursement. Due to the pilot nature of the case studies this was not possible in this thesis. A related limitation of the case studies is that the chosen study designs were not updated based on knowledge obtained during the study period. The use of 'historical controls' in the stage III colon cancer case study would most likely have contributed to a further reduction of the uncertainty of the real-world cost-effectiveness of oxaliplatin. It should be noted that often only retrospective designs will be suitable for the data collection of 'historical controls'.

Another limitation in this thesis was the use of a simulation study (described in Chapter 6 and 7) instead of an empirical study. In this simulation study, a real-world dataset was synthesised. With it being only a simulation, results may vary greatly in the real world due to unforeseen factors. A little error during simulation can alter the results and sometimes it is difficult to interpret the simulation results. In order to maximise and check the generalisability, the simulation data applied in this study were based on empirical data from a randomised controlled trial and an observational study of chemotherapy in patients with colorectal carcinoma, which is one of the case studies discussed in Chapter 5. Although generally empirical data are preferred, the important advantage of using simulated data in the comparison of confounding correction methods is that the 'true' treatment effect is known.

In the simulation study, we assumed that there was no unmeasured confounding. However, when treatments are assigned to patients, something like the intuition of the treating physician may play a role. Since this cannot be explicitly expressed in a variable which can be adjusted for statistically using a regression or PS adjustment method, it could cause a biased estimate of the ICER as well as the uncertainty surrounding the ICER. It is expected that not taking unmeasured confounding into account in the simulation study led to an underestimation of this uncertainty. In practice, the assumption of no unobserved confounding cannot be tested. However, real-world studies should assess whether this assumption is plausible by drawing on external evidence or expert opinion of the potential influence of observed and unobserved baseline covariates on treatment assignment and endpoints.[125, 172] It is also advised to

perform sensitivity analyses that assess the sensitivity of study results to an unobserved confounder.[173, 174]

Our simulation study focussed on optimising the validity and precision of outcomes by means of optimising the use of propensity score methods. However, also other factors can have a substantial impact on the validity and precision of the outcomes, such as structural model uncertainty, parameter uncertainty, heterogeneity and defining the target population.

A final limitation of this thesis is that it only addresses issues regarding real-world cost-effectiveness studies in oncology. Obviously, this limits the generalisability of the findings of this thesis. Nevertheless, many of the potentials and pitfalls addressed in this thesis would equally apply to real-life cost-effectiveness studies of other types of interventions. It is, however, important to note that we did not investigate the feasibility of collecting real-life data on HRQOL and productivity losses in oncology. This feasibility needs to be addressed in future research.

Implications

The number of new innovative, but very expensive drugs will continue to rise in the near future. Moreover, different expensive drugs are often combined with each other, especially in oncology, which puts an even greater financial burden on health care systems. For this reason, a careful consideration of the therapeutic value and cost-effectiveness of these (combinations of) drugs is important. Although real-world studies have important limitations, they provide valuable information that is not available in RCTs, such as appropriate drug use in daily practice. This kind of data contributes to the estimation of the real-world benefits and costs during the years of conditional reimbursement since usually at the start of this period, there is no information from daily clinical practice available. The addition of real-world observational studies to the total evidence package of a drug results in better evidence by addressing uncertainty in outcomes arising from the gap between clinical trials and everyday clinical practice. Since 2006, decision makers require the collection of real-world data and aim to use this kind of information in their decision-making process. Next to this, real-world data can also provide valuable information for treating physicians and patients. Detailed data on the use of new drugs give physicians insight into their prescription behaviour and allow them to reflect their choices

and its consequences. Physicians can compare their approach with that of others. For example, this type of "mirror" information is currently being linked to a prospective observational melanoma registry in The Netherlands.[175] In this way, physicians can learn from each other, which can enhance quality of care. Moreover, we observed that clinical choices made in oncology based on the efficacy of therapy have economic repercussions and that the appropriateness of treatment must necessarily take this aspect into account considering, above all, the efficacy and safety of the treatments analysed. The careful balancing, as here facilitated by detailed real-world data, of drug use, its effectiveness, and the associated costs of treatment should help physicians and decision makers in their goals to offer optimal treatments to patients within the context of finite healthcare resources. It is advisable to also take evidence from real-world observational studies into consideration in the development of treatment guidelines.

In this thesis, diverse challenges that complicate the evaluation of the appropriate use, relative effectiveness and cost-effectiveness of new drugs were identified. The extent to which these challenges can be overcome within the set timeframe depends on the type of drug, number of patients, heterogeneity and natural history of disease. Whether it will be necessary to overcome these challenges also depends on these things, plus the total package of current evidence being available, and the costs of the new drugs. As a consequence, there is not just one formula for evidence building when it comes to drugs for different diseases and indications; a tailor made approach is necessary.

Regarding the application of real-world data for reimbursement, the study plan prepared before conditional reimbursement starts should include a clear description of how the data collection will reduce uncertainty for decision makers at the reappraisal time. Comprehensive knowledge of the disease and the treatment, early modelling and/or a value of information (VOI) analysis can assist in the identification of important knowledge gaps. A VOI analysis can provide information on the parameters for which additional research is most useful.[176] A recent study illustrated the potential value of applying a VOI analysis prior to the start of a real-life cost-effectiveness study.[177] However, to conduct a VOI analysis it is necessary to have a valid T = 0 model. Such models may not always be available at T = 0.

Therefore, next to a clear T=o study plan, one should strive to design real-world observational studies in an optimal way with regard to the issues discussed above. Prospective population based disease registries can help to obtain sufficient numbers of similarly treated patients, monitor patients over a long time period, support data collection, and can include the collection of patient reported outcomes. Furthermore, patient registries offer the opportunity to build new research infrastructures. Although patient registries cannot resolve all issues, if they are used by an active interdisciplinary collaborative research group, they could increase efficiency of data collection and help to reduce issues of generalisability, incomparability of patient groups, missing information and lack of standardisation in reporting.

Over the past few years, many prospective observational studies have been set up as oncologic disease registries in The Netherlands. Examples include the PHAROS registry for hematolo-oncologic diseases,[178, 179] the PERCEPTION registry for renal cell carcinoma,[180] and the DREAM registry in rheumatology.[181] From experience it is known that setting up a registry is a large and costly project that takes time for its design and implementation. In the context of the quality of the data collection process and it sustainability, it is important to first work out the goals of the registry and then carefully consider which set of parameters to include. Also a continuous involvement of physicians is important to guide these choices.

In the context of conditional reimbursement, real-world observational studies are best used to evaluate the real-world applicability of evidence derived largely through randomised trials, to study patients and conditions not typically included or studied in randomised trials, to better understand current treatment practices, and to obtain insights regarding resource use and costs.[126] Real-world data based on observational studies should never aim to replace evidence from randomised controlled trials. They complement each other.

We suggest that a real-world incremental cost-effectiveness estimate is based on a synthesis of the evidence drawn from different sources, including real-world experiences with the drug, the pivotal clinical registration trial (including long-term follow-up) and other relevant evidence available from the literature. Combining these data sources can be done in different ways, depending on the quantity and quality of the real-world data (which depends on the type of drug, number of patients, heterogeneity and natural history of disease and quality of

study design) and availability of other sources of evidence. In this light, some practical recommendations can be made based on the results of the studies described in this thesis, which should already be considered before the start of the conditional reimbursement phase.

First, when a substantial uncertainty regarding the real-world (cost-) effectiveness exists and when it seems impossible to collect data on a suitable comparator treatment, nor are applicable RCTs is available, it is recommended that other options for data collection are explored. Good studies in this case are pragmatic clinical trials which are especially designed to answer policy questions.[182, 183]

Second, when one does not expect to be able to obtain a valid estimate of treatment effect based on real-world observational study data alone, but applicable trials are available, it is advised to combine the studies via modelling and scenario analyses, such as those performed in Chapter 4. In these specific cases, designing a single-arm observational study including only patients treated with the new expensive drug would be an efficient alternative to the two-arm observational study designs we used in the case studies. However, these kind of single-armed studies can only be applied to test the predictive validity of the T = 0 models. Moreover, the outcomes will only be valid if the assumption holds that the predictive ability of the treatment and control arms is equally large. Consequently, the use of a single-arm study design may be limited in practice. Third, one should only consider calculating comparative (cost-) effectiveness based on real-world data alone when all earlier mentioned challenges can be overcome and when the study can be adequately powered for the estimation of a predefined treatment effect outcome of interest. In these cases, statistical adjustment methods can be used to correct for confounding. It should be noted that applying adjustment methods will only lead to meaningful results when the treatment groups that are the subject of the comparison in the observational dataset, do sufficiently overlap with respect to their baseline characteristics to allow balanced samples after application of propensity-score based adjustment methods, and when information on all relevant baseline characteristics is appropriately measured to minimise the potential for unmeasured confounding.

Based on Chapter 6 and 7, 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.

Since none of the methods and specifications consistently performs best in all circumstances, it is recommended that several methods be 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. The use and presentation of results of different methods make the extent of this structural uncertainty transparent. In addition, 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.[169, 172] 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, although evidence of agreement is not the same as evidence of validity. Subsequently, it is advised to conduct sensitivity analysis to explore the magnitude and possible impact of residual unmeasured confounding.[173] Lastly, the results should always be considered in the context of all available evidence from other studies (RCTs) and if possible be combined via meta-analysis.[184,183]

Concluding remarks

In the last decade, governments have introduced policies linking reimbursement to a requirement for additional data collection.[7-10] It has been claimed that the resulting 'schemes' or requirements, such as 'patient access schemes,'[121] 'managed entry agreements,'[122] 'access/coverage with evidence development,'[123, 124] comparative effectiveness research[35, 37] or outcomes research,[11-13] can result in better evidence by reducing uncertainty arising from the gap between clinical trials and daily clinical practice.

This thesis demonstrated that additional data collection from the 'real-world' is valuable and leads, when combined with other available evidence, to cost-effectiveness estimates that are applicable to real-world clinical practice. In the context of the Dutch conditional reimbursement policy this means that after four years the uncertainty of the relative effects and costs of an expensive drug

is reduced and/or adequately reflects uncertainties faced in real-world clinical practice, which is relevant for decision makers.

An important question that remains to be addressed is how this additional evidence facilitates the decision maker in making decisions on continued reimbursement of a particular drug. Carbonneil et al. (2009) identified four critical success factors for access with evidence generation: coordination between decision-makers, medical and health technology assessment (HTA) agencies; methodological guidance; funding; and an implemented regulatory framework.[7] This thesis aimed to contribute to the methodological guidance, but another aspect that continues to be the centre of attention in The Netherlands is the implementation of a solid regulatory framework.

According to the current framework, specific expensive drugs require to be re-assessed regarding their actual budget impact, real-world therapeutic added value, appropriate use and cost-effectiveness, after a maximum of four years of conditional reimbursement. A final reimbursement advice is based on the updated knowledge regarding the four package principles, necessity, effectiveness, cost-effectiveness and feasibility. Updated knowledge on appropriate drug use and cost-effectiveness are required to be investigated in real-world practice. According to current guidelines, the decision about whether or not to continue reimbursement will be also based on the real-world cost-effectiveness.[185]

Recent experiences in The Netherlands have taught us that the role of cost-effectiveness and real-world evidence might have been underestimated by manufacturers. After four years of conditional reimbursement, the first reassessments[186-188] showed a lack of sufficient real-world data provided by the manufacturer to substantiate the real-world cost-effectiveness, partly because of suboptimal real-world study designs. It is important to realise that after a period of real-world data gathering, robust evidence does not appear spontaneously and will require a careful consideration of the evidence gaps and, subsequently, a careful selection of an appropriate study design to fill these gaps, at the start of the conditional reimbursement phase.

On the other hand, the Dutch Health Care Insurance Board (CVZ) might have been overestimating the potential of the 'stand-alone' use of Dutch real-world data in the estimation of the real-world cost-effectiveness. It is important that the Dutch Health Care Insurance Board sets realistic and relevant requirements upfront and challenges and approves study designs to match these realistic

and relevant requirements. This thesis showed that it was very challenging to estimate unbiased comparative treatment effects of a new drug, due to inabilities to adequately correct for confounding by indication. If a need exists to demonstrate a certain unbiased comparative treatment effect as the primary objective of a real-world study, a (pragmatic) trial should be considered as first option, rather than a prospective registry.

Given existing treatment dynamics,[5] it is advised to have interim discussions (for example after one or two years) between CVZ and manufacturers, to check whether important deviations from the original study plan exist and/or are desirable. Also the potential consequences in case requirements are not met should be well documented. In this light, it should also be noted that legitimate decision making on whether or not drugs provide sufficient societal value for money requires a transparent cost-effectiveness threshold at the start of conditional reimbursement, as well as when a final reimbursement decision needs to be made.

Although cost-effectiveness plays an important role in the current Dutch regulatory framework, it seems doubtful whether it is feasible and/or desirable to stop continuing the reimbursement of a medicine that is positively judged regarding effectiveness (including appropriate use), necessity and feasibility, but is not considered cost-effective because of its high costs. This doubt is reflected in the developments of the Dutch regulatory framework where financial options are expanded rather than cost-ineffective drugs are excluded. In 2011, the Dutch minister has proposed the implementation of financial arrangements from 2013 onwards[189] and today we can find some examples where financial arrangements are used as an effective tool to reduce prices towards an acceptable real-world cost-effectiveness.[187]

References

REFERENCES

1. Pronk MH, Bonsel GJ. Out-patient drug policy by clinical assessment rather than financial constraints? the gate-keeping function of the out-patient drug reimbursement system in the netherlands. Eur J Health Econ 2004 Oct;5(3):274-7.
2. A background study on the 'cost-effectiveness' package principle for the benefit of the appraisal phase in package management [Internet]; c2010 [cited 2010 . Available from: http://www.cvz.nl/binaries/live/cvzinternet/hst_content/en/documents/assessments/asm1011-cost-effectiveness-principle.pdf.
3. Niezen MG, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. Eur J Cancer 2006 Nov;42(17):2887-92.
4. NZA tarieven [Internet]; c2009. Available from: http://ctg.bit-ic.nl/Nzatarieven/top.do.
5. Delwel G. Guidance for outcomes research [leidraad voor uitkomstenonderzoek]. diemen: College voor zorgverzekeringen; 2008. report no.: 270.
6. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005 Jan 1-7;365(9453):82-93.
7. Carbonneil C, Quentin F, Lee-Robin SH, European network for Health Technology Assessment (EUnetHTA). A common policy framework for evidence generation on promising health technologies. Int J Technol Assess Health Care 2009 Dec;25 Suppl 2:56-67.
8. Garrison LP,Jr, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: The ISPOR real-world data task force report. Value Health 2007 Sep-Oct;10(5):326-35.
9. Hutton J, Trueman P, Henshall C. Coverage with evidence development: An examination of conceptual and policy issues. Int J Technol Assess Health Care 2007 Fall;23(4):425-32.
10. Stafinski T, McCabe CJ, Menon D. Funding the unfundable: Mechanisms for managing uncertainty in decisions on the introduction of new and innovative technologies into healthcare systems. Pharmacoeconomics 2010;28(2):113-42.
11. Krumholz HM. Outcomes research: Generating evidence for best practice and policies. Circulation 2008 Jul 15;118(3):309-18.
12. Clancy CM, Eisenberg JM. Outcomes research: Measuring the end results of health care. Science 1998 Oct 9;282(5387):245-6.
13. Lipscomb J, Donaldson MS, Hiatt RA. Cancer outcomes research and the arenas of application. J Natl Cancer Inst Monogr 2004;(33)(33):1-7.
14. Silverman SL. From randomized controlled trials to observational studies. Am J Med 2009 Feb;122(2):114-20.

15. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: Approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The international society for pharmacoeconomics and outcomes research good research practices for retrospective database analysis task force report—part II. Value Health 2009 Nov-Dec;12(8):1053-61.
16. Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. Am J Med 2010 Dec;123(12 Suppl 1):e16-23.
17. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: Analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report—part III. Value Health 2009 Nov-Dec;12(8):1062-73.
18. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: The importance of restriction. JAMA 2010 Aug 25;304(8):897-8.
19. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55.
20. Funk MJ, Westreich D, Wiesen C, Sturmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol 2011 Apr 1;173(7):761-7.
21. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci 2010 Feb 1;25(1):1-21.
22. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3): 399-424.
23. 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 Jun;58(6):550-9.
24. 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 Feb;51(1):171-84.
25. Mitra N, Indurkhya A. A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data. Health Econ 2005 Aug;14(8):805-15.
26. Armstrong K. Methods in comparative effectiveness research. J Clin Oncol 2012 Dec 1;30(34):4208-14.
27. Hammond GC, Croudace TJ, Radhakrishnan M, Lafortune L, Watson A, McMillan-Shields F, Jones PB. Comparative effectiveness of cognitive therapies delivered

face-to-face or over the telephone: An observational study using propensity methods. PLoS One 2012;7(9):e42916.
28. Baumler M, Stargardt T, Schreyogg J, Busse R. Cost effectiveness of drug-eluting stents in acute myocardial infarction patients in germany: Results from administrative data using a propensity score-matching approach. Appl Health Econ Health Policy 2012 Jul 1;10(4):235-48.
29. Indurkhya A, Mitra N, Schrag D. Using propensity scores to estimate the cost-effectiveness of medical therapies. Stat Med 2006 May 15;25(9):1561-76.
30. Sekhon JS, Grieve RD. A matching method for improving covariate balance in cost-effectiveness analyses. Health Econ 2012 Jun;21(6):695-714.
31. Gruschkus SK, Lairson D, Dunn JK, Risser J, Du XL. 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 Mar-Apr;14(2):253-62.
32. Dhainaut JF, Payet S, Vallet B, Franca LR, Annane D, Bollaert PE, Le Tulzo Y, Runge I, Malledant Y, Guidet B, et al. Cost-effectiveness of activated protein C in real-life clinical practice. Crit Care 2007;11(5):R99.
33. Coleman CI, McKay RG, Boden WE, Mather JF, White CM. 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 Jul;28(7):1054-62.
34. 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 Oct 8.
35. Gatsonis C. The promise and realities of comparative effectiveness research. Stat Med 2010 Aug 30;29(19):1977,81; discussion 1996-7.
36. Sorenson C, Naci H, Cylus J, Mossialos E. Evidence of comparative efficacy should have a formal role in european drug approvals. BMJ 2011 Sep 6;343:d4849.
37. Sullivan P, Goldmann D. The promise of comparative effectiveness research. JAMA 2011 Jan 26;305(4):400-1.
38. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. Stat Med 2010 Aug 30;29(19):1963-76.
39. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: Defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report—part I. Value Health 2009 Nov-Dec;12(8):1044-52.

40. Krumholz HM. Real-world imperative of outcomes research. JAMA 2011 Aug 17; 306(7):754-5.
41. Tunis SR, Chalkidou K. Coverage with evidence development: A very good beginning, but much to be done. commentary to hutton et al. Int J Technol Assess Health Care 2007 Fall;23(4):432-5.
42. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. effectiveness versus efficacy studies. Pharmacoeconomics 1999 May;15(5):423-34.
43. Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, Starling N. Colorectal cancer. Lancet 2010 Mar 20;375(9719):1030-47.
44. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW. Recent trends of cancer in europe: A combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. Eur J Cancer 2008 Jul;44(10):1345-89.
45. Lemmens VEPP, Coebergh JWW. Epidemiologie van colorectale tumouren. IKR Bulletin 2006;30.
46. Obrand DI, Gordon PH. Incidence and patterns of recurrence following curative resection for colorectal carcinoma. Dis Colon Rectum 1997 Jan;40(1):15-24.
47. Thirion P, Michiels S, Pignon JP, Buyse M, Braud AC, Carlson RW, O'Connell M, Sargent P, Piedbois P, Meta-Analysis Group in Cancer. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: An updated meta-analysis. J Clin Oncol 2004 Sep 15;22(18):3766-75.
48. Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, et al. End points for colon cancer adjuvant trials: Observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT group. J Clin Oncol 2007 Oct 10;25(29):4569-74.
49. Labianca R, Marsoni S, Pancera G, Torri V, Zaniboni A, Erlichman C, Pater J, Shepherd L, Zee B, Seitz JF, et al. Efficacy of adjuvant fluorouracil and folinic acid in colon cancer. international multicentre pooled analysis of colon cancer trials (IMPACT) investigators. Lancet 1995 Apr 15;345(8955):939-44.
50. O'Connell MJ, Mailliard JA, Kahn MJ, Macdonald JS, Haller DG, Mayer RJ, Wieand HS. Controlled trial of fluorouracil and low-dose leucovorin given for 6 months as postoperative adjuvant therapy for colon cancer. J Clin Oncol 1997 Jan;15(1): 246-50.
51. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: A final report. Ann Intern Med 1995 Mar 1;122(5):321-6.
52. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Goodman PJ, Ungerleider JS, Emerson WA, Tormey DC, Glick JH. Levamisole and fluorouracil for

adjuvant therapy of resected colon carcinoma. N Engl J Med 1990 Feb 8;322(6): 352-8.

53. Andre T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, Hickish T, Topham C, Zaninelli M, Clingan P, Bridgewater J, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. N Engl J Med 2004 Jun 3;350(23): 2343-51.
54. Cassidy J, Tabernero J, Twelves C, Brunet R, Butts C, Conroy T, Debraud F, Figer A, Grossmann J, Sawada N, et al. XELOX (capecitabine plus oxaliplatin): Active first-line therapy for patients with metastatic colorectal cancer. J Clin Oncol 2004 Jun 1;22(11):2084-91.
55. Shields AF, Zalupski MM, Marshall JL, Meropol NJ. Treatment of advanced colorectal carcinoma with oxaliplatin and capecitabine: A phase II trial. Cancer 2004 Feb 1;100(3):531-7.
56. Twelves CJ. Xeloda in adjuvant colon cancer therapy (X-ACT) trial: Overview of efficacy, safety, and cost-effectiveness. Clin Colorectal Cancer 2006 Nov;6(4):278-87.
57. Punt CJ. Chemotherapy of patients with colorectal carcinoma. Ned Tijdschr Geneeskd 2005 Jun 25;149(26):1441-7.
58. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005 Mar-Apr;55(2):74-108.
59. Twelves C, Wong A, Nowacki MP, Abt M, Burris H,3rd, Carrato A, Cassidy J, Cervantes A, Fagerberg J, Georgoulias V, et al. Capecitabine as adjuvant treatment for stage III colon cancer. N Engl J Med 2005 Jun 30;352(26):2696-704.
60. Punt CJ. New options and old dilemmas in the treatment of patients with advanced colorectal cancer. Ann Oncol 2004 Oct;15(10):1453-9.
61. Cassidy J, Twelves C, Van Cutsem E, Hoff P, Bajetta E, Boyer M, Bugat R, Burger U, Garin A, Graeven U, et al. First-line oral capecitabine therapy in metastatic colorectal cancer: A favorable safety profile compared with intravenous 5-fluorouracil/leucovorin. Ann Oncol 2002 Apr;13(4):566-75.
62. Kuebler JP, Wieand HS, O'Connell MJ, Smith RE, Colangelo LH, Yothers G, Petrelli NJ, Findlay MP, Seay TE, Atkins JN, et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: Results from NSABP C-07. J Clin Oncol 2007 Jun 1;25(16):2198-204.
63. Schmoll HJ, Cartwright T, Tabernero J, Nowacki MP, Figer A, Maroun J, Price T, Lim R, Van Cutsem E, Park YS, et al. Phase III trial of capecitabine plus oxaliplatin as adjuvant therapy for stage III colon cancer: A planned safety analysis in 1,864 patients. J Clin Oncol 2007 Jan 1;25(1):102-9.
64. Andre T, Boni C, Navarro M, Tabernero J, Hickish T, Topham C, Bonetti A, Clingan P, Bridgewater J, Rivera F, et al. Improved overall survival with oxaliplatin, fluoro-

uracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. J Clin Oncol 2009 Jul 1;27(19):3109-16.

65. Labianca R, Nordlinger B, Beretta GD, Brouquet A, Cervantes A, ESMO Guidelines Working Group. Primary colon cancer: ESMO clinical practice guidelines for diagnosis, adjuvant treatment and follow-up. Ann Oncol 2010 May;21 Suppl 5: v70-7.
66. 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(5):373-83.
67. Lemmens VE, van Halteren AH, Janssen-Heijnen ML, Vreugdenhil G, Repelaer van Driel OJ, Coebergh JW. Adjuvant treatment for elderly patients with stage III colon cancer in the southern netherlands is affected by socioeconomic status, gender, and comorbidity. Ann Oncol 2005 May;16(5):767-72.
68. Kahn KL, Adams JL, Weeks JC, Chrischilles EA, Schrag D, Ayanian JZ, Kiefe CI, Ganz PA, Bhoopalam N, Potosky AL, et al. Adjuvant chemotherapy use and adverse events among older patients with stage III colon cancer. JAMA 2010 Mar 17; 303(11):1037-45.
69. Goldberg RM, Tabah-Fisch I, Bleiberg H, de Gramont A, Tournigand C, Andre T, Rothenberg ML, Green E, Sargent DJ. Pooled analysis of safety and efficacy of oxaliplatin plus fluorouracil/leucovorin administered bimonthly in elderly patients with colorectal cancer. J Clin Oncol 2006 Sep 1;24(25):4085-91.
70. McCleary JN, Meyerhardt J, Green E, et al. Impact of older age on the efficacy of newer adjuvant therapies in >12,500 patients (pts) with stage II/III colon cancer: Findings from the ACCENT database (abstract). J Clin Oncol 2009;27(170s).
71. Haller DG, Cassidy J, Tabernero J, Et al. Efficacy findings from a randomized phase III trial of capecitabine plus oxaliplatin verus bolus 5-FU/LV for stage III colon cancer (NO16968): No impact of age on disease-free survival (DFS). (abstract). presented at the 2010 ASCO gastrointestinal cancers symposium, orlando, FL, january 22. 2010.
72. Etzioni DA, El-Khoueiry AB, Beart RW,Jr. Rates and predictors of chemotherapy use for stage III colon cancer: A systematic review. Cancer 2008 Dec 15;113(12): 3279-89.
73. van Steenbergen LN, Rutten HJ, Creemers GJ, Pruijt JF, Coebergh JW, Lemmens VE. Large age and hospital-dependent variation in administration of adjuvant chemotherapy for stage III colon cancer in southern netherlands. Ann Oncol 2010 Jun;21(6):1273-8.
74. Arkenau HT, Arnold D, Cassidy J, Diaz-Rubio E, Douillard JY, Hochster H, Martoni A, Grothey A, Hinke A, Schmiegel W, et al. Efficacy of oxaliplatin plus capecitabine or infusional fluorouracil/leucovorin in patients with metastatic

colorectal cancer: A pooled analysis of randomized trials. J Clin Oncol 2008 Dec 20;26(36):5910-7.

75. Haller D, Tabernero J, Maroun J, Et al. First efficacy findings from a randomized phase III trial of capecitabine + oxaliplatin vs. bolus 5-FU/LV for stage III colon cancer (NO16968/XELODA study). presented at the joint ECCO/ESMO multidisciplinary congress, (abstr 5LBA), berlin, 20-24 september. 2009.
76. Mol L, Punt CJA, van Gils CWM, Ottevanger PB, Koopman M. Trial participation in a multicentre phase III trial (CAIRO) in advanced colorectal cancer patients in the netherlands, and a comparison of outcome between trial and non-trial patients. (abstract 581PD). presented at the 35th ESMO congress milan, italy 8-12 october 2010.
77. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in europe: Estimates for 40 countries in 2012. Eur J Cancer 2013 Feb 25.
78. Haller DG, Tabernero J, Maroun J, de Braud F, Price T, Van Cutsem E, Hill M, Gilberg F, Rittweger K, Schmoll HJ. Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer. J Clin Oncol 2011 Apr 10;29(11):1465-71.
79. Poos MJJC, Smit JM, Groen J, Kommer GJ, Slobbe LCJ. Costs of ilness in the netherlands. RIVM 2005.
80. Ferro SA, Myer BS, Wolff DA, Poniewierski MS, Culakova E, Cosler LE, Scarpace SL, Khorana AA, Lyman GH. Variation in the cost of medications for the treatment of colorectal cancer. Am J Manag Care 2008 Nov;14(11):717-25.
81. Krol M, Koopman M, Uyl-de Groot C, Punt CJ. A systematic review of economic analyses of pharmaceutical therapies for advanced colorectal cancer. Expert Opin Pharmacother 2007 Jun;8(9):1313-28.
82. Aballea S, Boler A, Craig A, Wasan H. An economic evaluation of oxaliplatin for the adjuvant treatment of colon cancer in the united kingdom (UK). Eur J Cancer 2007 Jul;43(11):1687-93.
83. Aballea S, Chancellor JV, Raikou M, Drummond MF, Weinstein MC, Jourdan S, Bridgewater J. Cost-effectiveness analysis of oxaliplatin compared with 5-fluorouracil/leucovorin in adjuvant treatment of stage III colon cancer in the US. Cancer 2007 Mar 15;109(6):1082-9.
84. Cassidy J, Douillard JY, Twelves C, McKendrick JJ, Scheithauer W, Bustova I, Johnston PG, Lesniewski-Kmak K, Jelic S, Fountzilas G, et al. Pharmacoeconomic analysis of adjuvant oral capecitabine vs intravenous 5-FU/LV in dukes' C colon cancer: The X-ACT trial. Br J Cancer 2006 Apr 24;94(8):1122-9.
85. Douillard JY, Tilleul P, Ychou M, Dufour P, Perrocheau G, Seitz JF, Maes P, Lafuma A, Husseini F. Cost consequences of adjuvant capecitabine, mayo clinic and de

gramont regimens for stage III colon cancer in the french setting. Oncology 2007; 72(3-4):248-54.

86. Eggington S, Tappenden P, Pandor A, Paisley S, Saunders M, Seymour M, Sutcliffe P, Chilcott J. Cost-effectiveness of oxaliplatin and capecitabine in the adjuvant treatment of stage III colon cancer. Br J Cancer 2006 Nov 6;95(9):1195-201.
87. Maniadakis N, Fragoulakis V, Pectasides D, Fountzilas G. XELOX versus FOLFOX6 as an adjuvant treatment in colorectal cancer: An economic analysis. Curr Med Res Opin 2009 Feb 12.
88. Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P. The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: Systematic review and economic evaluation. Health Technol Assess 2006 Nov;10(41):iii,iv, xi-xiv, 1-185.
89. Hsiao FY, Mullins CD, Onukwugha E, Pandya N, Hanna N. Comparative effectiveness of different chemotherapeutic regimens on survival of people aged 66 and older with stage III colon cancer: A "real world" analysis using surveillance, epidemiology, and end results-medicare data. J Am Geriatr Soc 2011 Sep;59(9): 1717-23.
90. Dane F, Akif Ozturk M, Gumus M, Guven A, Aliustaoglu M, Cabuk D, Teomete M, Basaran G, Fulden Yumuk P, Serdar Turhal N. Efficacy, safety and prognostic features of resected colon carcinoma treated in "real world" practice: A retrospective cohort-study. J BUON 2011 Apr-Jun;16(2):257-64.
91. van Steenbergen LN, Lemmens VE, Rutten HJ, Wymenga AN, Nortier JW, Janssen-Heijnen ML. Increased adjuvant treatment and improved survival in elderly stage III colon cancer patients in the netherlands. Ann Oncol 2012 Nov;23(11):2805-11.
92. van Gils CW, Koopman M, Mol L, Redekop WK, Uyl-de Groot CA, Punt CJ. Adjuvant chemotherapy in stage III colon cancer: Guideline implementation, patterns of use and outcomes in daily practice in the netherlands. Acta Oncol 2012 Jan; 51(1):57-64.
93. Xie Q, Wen F, Wei YQ, Deng HX, Li Q. Cost analysis of adjuvant therapy with XELOX or FOLFOX4 for colon cancer. Colorectal Dis 2013 Aug;15(8):958-62.
94. Aitini E, Rossi A, Morselli P, Vivorio B, Bruschi A, Bottura C, Colombo GL. Economic comparison of capecitabine + oxaliplatin and 5-fluorouracil + oxaliplatin in the adjuvant treatment of colon cancer. Cancer Manag Res 2012;4:99-103.
95. Tan SS, Hakkaart-van Roijen L, Al MJ, Bouwmans CA, Hoogendoorn ME, Spronk PE, Bakker J. A microcosting study of intensive care unit stay in the netherlands. J Intensive Care Med 2008 Jul-Aug;23(4):250-7.
96. Tan SS, Van Gils CW, Franken MG, Hakkaart-van Roijen L, Uyl-de Groot CA. The unit costs of inpatient hospital days, outpatient visits, and daycare treatments in the fields of oncology and hematology. Value Health 2010 Sep-Oct;13(6):712-9.

97. Farmacotherapeutisch Kompas[Internet] cited 2009 . Available from: www.fk.cvz.nl.
98. Centraal Bureau voor de Statistiek [Internet]. Available from www.cbs.nl.
99. Attard CL, Maroun JA, Alloul K, Grima DT, Bernard LM. Cost-effectiveness of oxaliplatin in the adjuvant treatment of colon cancer in canada. Curr Oncol 2010 Feb;17(1):17-24.
100. Shiroiwa T, Takeuchi T, Fukuda T, Shimozuma K, Ohashi Y. Cost-effectiveness of adjuvant FOLFOX therapy for stage III colon cancer in japan based on the MOSAIC trial. Value Health 2012 Mar-Apr;15(2):255-60.
101. Cassidy J, Saltz L, Twelves C, Van Cutsem E, Hoff P, Kang Y, Saini JP, Gilberg F, Cunningham D. Efficacy of capecitabine versus 5-fluorouracil in colorectal and gastric cancers: A meta-analysis of individual data from 6171 patients. Ann Oncol 2011 Dec;22(12):2604-9.
102. Twelves C, Scheithauer W, McKendrick J, Seitz JF, Van Hazel G, Wong A, Diaz-Rubio E, Gilberg F, Cassidy J. Capecitabine versus 5-fluorouracil/folinic acid as adjuvant therapy for stage III colon cancer: Final results from the X-ACT trial with analysis by age and preliminary evidence of a pharmacodynamic marker of efficacy. Ann Oncol 2012 May;23(5):1190-7.
103. Twelves C, Gollins S, Grieve R, Samuel L. A randomised cross-over trial comparing patient preference for oral capecitabine and 5-fluorouracil/leucovorin regimens in patients with advanced colorectal cancer. Ann Oncol 2006 Feb;17(2):239-45.
104. Di Costanzo F, Ravasio R, Sobrero A, Bertetto O, Vinante O, Luppi G, Labianca R, Amadori D, Barone C, Carlo Merlano M, et al. Capecitabine versus bolus fluorouracil plus leucovorin (folinic acid) as adjuvant chemotherapy for patients with dukes' C colon cancer: Economic evaluation in an italian NHS setting. Clin Drug Investig 2008;28(10):645-55.
105. Limat S, Bracco-Nolin CH, Legat-Fagnoni C, Chaigneau L, Stein U, Huchet B, Pivot X, Woronoff-Lemsi MC. Economic impact of simplified de gramont regimen in first-line therapy in metastatic colorectal cancer. Eur J Health Econ 2006 Jun;7(2):107-13.
106. Ward SE, Kaltenthaler E, Cowan J, Marples M, Orr B, Seymour MT. The clinical and economic benefits of capecitabine and tegafur with uracil in metastatic colorectal cancer. Br J Cancer 2006 Jul 3;95(1):27-34.
107. van Gils CW, de Groot S, Redekop WK, Koopman M, Punt CJ, Uyl-de Groot CA. Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: A synthesis of clinical trial and daily practice evidence. Pharmacoeconomics 2013 May 9.
108. Ligtenberg G, Staal PC, Goettsch WG, Knies S. Besluitvorming in de zorg op grond van kosteneffectiviteit: Consultatiedocument. ; 2013. Report nr 2013047960.

109. Kuijpers M, Toenders W. Assessment procedure for inpatient drugs [procedure beoordeling intramurale geneesmiddelen]. Diemen: College voor zorgverzekeringen; 2006. report no.: 26022597.
110. Pilot outcomes research: costs and effects of oxaliplatin in stage III and metastatic colorectal cancer [Internet]: Dutch Health Care Insurance Board; c2010. Available from: http://www.cvz.nl/binaries/live/cvzinternet/hst_content/nl/documenten/rubriek+zorgpakket/imta-onderzoek-dure-geneesmiddelen-1207.pdf.
111. Thirunavukarasu P, Sukumar S, Sathaiah M, Mahan M, Pragatheeshwar KD, Pingpank JF, Zeh H,3rd, Bartels CJ, Lee KK, Bartlett DL. C-stage in colon cancer: Implications of carcinoembryonic antigen biomarker in staging, prognosis, and management. J Natl Cancer Inst 2011 Apr 20;103(8):689-97.
112. Rodenburg-van Dieten HEM. Richtlijnen voor farmaco-economisch onderzoek; evaluatie en actualisatie. 2005.
113. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons ltd. 2009.
114. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. Am J Gastroenterol 1999 Jun;94(6):1650-7.
115. Ramsey SD, Andersen MR, Etzioni R, Moinpour C, Peacock S, Potosky A, Urban N. Quality of life in survivors of colorectal carcinoma. Cancer 2000 Mar 15;88(6):1294-303.
116. Centraal Bureau voor de Statistiek[Internet]; c2009. Available from: www.cbs.nl.
117. Richtlijnen oncologische zorg [Internet]. Available from: www.oncoline.nl.
118. Tol J, Koopman M, Cats A, Rodenburg CJ, Creemers GJ, Schrama JG, Erdkamp FL, Vos AH, van Groeningen CJ, Sinnige HA, et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. N Engl J Med 2009 Feb 5;360(6):563-72.
119. Franken MG, van Gils CW, Gaultney JG, Delwel GO, Goettsch W, Huijgens PC, Steenhoek A, Punt CJ, Koopman M, Redekop WK, et al. Practical feasibility of outcomes research in oncology: Lessons learned in assessing drug use and cost-effectiveness in the netherlands. Eur J Cancer 2013 Jan;49(1):8-16.
120. Tournigand C, Andre T, Bonnetain F, Chibaudel B, Lledo G, Hickish T, Tabernero J, Boni C, Bachet JB, Teixeira L, et al. Adjuvant therapy with fluorouracil and oxaliplatin in stage II and elderly patients (between ages 70 and 75 years) with colon cancer: Subgroup analyses of the multicenter international study of oxaliplatin, fluorouracil, and leucovorin in the adjuvant treatment of colon cancer trial. J Clin Oncol 2012 Sep 20;30(27):3353-60.
121. Briggs A, Ritchie K, Fenwick E, Chalkidou K, Littlejohns P. Access with evidence development in the UK: Past experience, current initiatives and future potential. Pharmacoeconomics 2010;28(2):163-70.

122. Klemp M, Fronsdal KB, Facey K, HTAi Policy Forum. What principles should govern the use of managed entry agreements? Int J Technol Assess Health Care 2011 Jan;27(1):77-83.
123. Mohr PE, Tunis SR. Access with evidence development: The US experience. Pharmacoeconomics 2010;28(2):153-62.
124. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's ‚coverage with evidence development'. Health Aff (Millwood) 2006 Sep-Oct;25(5): 1218-30.
125. Rubin DB. On the limitations of comparative effectiveness research. Stat Med 2010 Aug 30;29(19):1991,5; discussion 1996-7.
126. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. Health Aff (Millwood) 2010 Oct;29(10):1818-25.
127. Chalkidou K, Whicher D, Kary W, Tunis S. Comparative effectiveness research priorities: Identifying critical gaps in evidence for clinical and health policy decision making. Int J Technol Assess Health Care 2009 Jul;25(3):241-8.
128. Kupersmith J, Ommaya AK. The past, present, and future of comparative effectiveness research in the US department of veterans affairs. Am J Med 2010 Dec;123(12 Suppl 1):e3-7.
129. Garrison LP,Jr, Mansley EC, Abbott TA,3rd, Bresnahan BW, Hay JW, Smeeding J. Good research practices for measuring drug costs in cost-effectiveness analyses: A societal perspective: The ISPOR drug cost task force report—part II. Value Health 2010 Jan-Feb;13(1):8-13.
130. Lokhorst HM, Schmidt-Wolf I, Sonneveld P, van der Holt B, Martin H, Barge R, Bertsch U, Schlenzka J, Bos GM, Croockewit S, et al. Thalidomide in induction treatment increases the very good partial response rate before and after high-dose therapy in previously untreated multiple myeloma. Haematologica 2008 Jan;93(1): 124-7.
131. Franken MG, Gaultney JG, Huijgens PC, Redekop WK, Uyl-de Groot CA. Treatment variation complicates real-world pharmacoeconomics: Daily clinical practice of bortezomib in relapsed or refractory multiple myeloma. Value Health 2009;12(7):A226.
132. Gaultney JG, Franken MG, van Gils CWM, Huijgens PC, Redekop WK, Uyl-de Groot CA. Outcomes research of bortezomib indicated for multiple myeloma in the context of the dutch reimbursement policy for expensive medicines: Threats to the internal validity of the incremental effectiveness estimate. Value Health 2009;12(7):A259.
133. Concato J, Peduzzi P, Huang GD, O'Leary TJ, Kupersmith J. Comparative effectiveness research: What kind of studies do we need? J Investig Med 2010 Aug;58(6): 764-9.

134. Byrne BJ, Kishnani PS, Case LE, Merlini L, Muller-Felber W, Prasad S, van der Ploeg A. Pompe disease: Design, methodology, and early findings from the pompe registry. Mol Genet Metab 2011 May;103(1):1-11.
135. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. 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 May;59(5):437-47.
136. Rubin DB, Thomas N. Combining propensity score matching with additional covariate adjustments for prognostic covariates. J Am Stat Assoc 2000(450):573-85.
137. Rubin DB. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics 1973;29(1):185-203.
138. Kreif N, Grieve R, Radice R, Sadique Z, Ramsahai R, Sekhon JS. Methods for estimating subgroup effects in cost-effectiveness analyses that use observational data. Med Decis Making 2012 Nov-Dec;32(6):750-63.
139. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol 2006 Jun 15;163(12):1149-56.
140. 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 Feb 20;26(4):734-53.
141. Koopman M, Antonini NF, Douma J, Wals J, Honkoop AH, Erdkamp FL, de Jong RS, Rodenburg CJ, Vreugdenhil G, Loosveld OJ, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): A phase III randomised controlled trial. Lancet 2007 Jul 14;370(9582):135-42.
142. Holland PW. Statistics and causal inference. J Am Stat Assoc;81(396):945-60.
143. Neyman J. On the application of probability theory to agricultural experiments. essay on principles. section 9.". Stat Sci 1923 [1990];5((4)):464-72.
144. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66(5):688-701.
145. Heckman JJ, Ichimura H, Todd P. Matching as an econometric evaluation estimator. Rev Econ Stud 1998;65(223):261-94.
146. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. New York: Springer-Verlag; 2008. ID: 784.
147. Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc 1987;82:387-94.
148. 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(259):278.
149. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008 Sep 15;168(6):656-64.

150. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. ID: 32.
151. 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 Dec;48(12):1503-10.
152. Baser O. Too much ado about propensity score models? comparing methods of propensity score matching. Value Health 2006 Nov-Dec;9(6):377-85.
153. 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 Jul 1;11(1-2):1-26.
154. Martens EP, de Boer A, Pestman WR, Belitser SV, Stricker BH, Klungel OH. Comparing treatment effects after adjustment with multivariable cox proportional hazards regression and propensity score methods. Pharmacoepidemiol Drug Saf 2008 Jan;17(1):1-8.
155. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: The consequences of ignoring decision uncertainty. Health Econ 2011 Feb;20(2):212-24.
156. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999 Jun;18(3): 341-64.
157. 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 Oct 15;21(19):2917-30.
158. Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating. New York: Springer 2009.
159. 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 Jan;16(1):41-8.
160. Senn S. Controversies concerning randomization and additivity in clinical trials. Stat Med 2004 Dec 30;23(24):3729-53.
161. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics 1992;48:479 - 495.
162. 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 Sep 10;29(20):2137-48.
163. Mol L, Koopman M, van Gils CW, Ottevanger PB, Punt CJ. Comparison of treatment outcome in metastatic colorectal cancer patients included in a clinical trial versus daily practice in the netherlands. Acta Oncol 2013 Jun;52(5):950-5.
164. Cole SR, Hernan MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed 2004 Jul;75(1):45-9.

165. 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 Jul-Aug;6(4):327-40.
166. DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statist Sci 1996(3):189-228.
167. Briggs A, Fenn P. Confidence intervals or surfaces? uncertainty on the cost-effectiveness plane. Health Econ 1998 Dec;7(8):723-40.
168. 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 Mar;61(3):247-55.
169. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009 Nov 10;28(25):3083-107.
170. Guidelines for pharmacoeconomic research [Internet]; c2006 [cited 2013 June, 15]. Available from: http://www.cvz.nl/binaries/content/documents/cvzinternet/nl/ documenten/rubriek+zorgpakket/cfh/richtlijnen+farmacoeconomisch+onderzoek+-+engels.pdf.
171. Krol M, Papenburg J, Koopmanschap M, Brouwer W. Do productivity costs matter?: The impact of including productivity costs on the incremental costs of interventions targeted at depressive disorders. Pharmacoeconomics 2011 Jul;29(7):601-19.
172. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: A critical appraisal tool and review of current practice. Health Econ 2013 Apr;22(4):486-500.
173. Liu W, Kuramoto SJ, Stuart EA. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prev Sci 2013 Feb 14.
174. Groenwold RH, Nelson DB, Nichol KL, Hoes AW, Hak E. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. Int J Epidemiol 2010 Feb;39(1):107-17.
175. DICA registratie melanoom van start [Internet]; c2013 [cited 2013 June, 15]. Available from: http://www.clinicalaudit.nl/node/89.
176. Oostenbrink JB, Al MJ, Oppe M, Rutten-van Molken MP. Expected value of perfect information: An empirical example of reducing decision uncertainty by conducting additional research. Value Health 2008 Dec;11(7):1070-80.
177. Corro Ramos I, Rutten-van Molken MP, Al MJ. The role of value-of-information analysis in a health care research priority setting: A theoretical case study. Med Decis Making 2013 May;33(4):472-89.
178. Blommestein HM, Franken M, Verelst S, Gaultney JG, Huijgens P, Sonneveld P, Redekop WK, Uyl- de Groot CA. Can a population-based patient registry improve the feasibility of outcomes research in multiple myeloma? 2012;Vol. 15(7):A432.

179. Huijgens PC, Posthuma EFM, Coebergh JWW, van de Pol-Franse LV, Uyl-de Groot CA, Sonneveld P. A population-based registry for hemato-oncology. [een 'population-based' registry voor hemato-oncologie.]. Nederlands Tijdschrift Voor Hematologie 2010;7:321-5.
180. de Groot S, Redekop WK, Kiemeney L, Uyl-de Groot CA. A population-based registry for the evaluation of new treatment options for patients with metastatic renal cell carcinoma. Value Health 2011;14(7):A467-8.
181. Dutch RhEumatoid Arthritis Monitoring (DREAM) [Internet]; c2013 [cited 2013 June 15]. Available from: http://www.dreamregistry.nl/en-US/.
182. Brass EP. The gap between clinical trials and clinical practice: The use of pragmatic clinical trials to inform regulatory decision making. Clin Pharmacol Ther 2010 Mar;87(3):351-5.
183. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003 Sep 24;290(12):1624-32.
184. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: The ISPOR good research practices task force report. Value Health 2012 Mar-Apr;15(2):217-30.
185. Ligtenberg G. Voorwaardelijke toelating/financiering van zorg. Diemen, the Netherlands, CVZ 2012.
186. Dutch Health Care Insurance Board [College voor Zorgverenigingen (CVZ)]. Conditially authorised specialist medicine, update october 2012 [voorwaardelijk toegelaten specialistische geneesmiddelen, geactualiseerd oktober 2012].
187. Dutch Health Care Insurance Board [College voor Zorgverenigingen (CVZ)]. Pay-for-perfomance agreement for omalizumab (xolair®) for severe asthma [pay-for-performance overeenkomst omalizumab (xolair®) bij ernstige astma] 2012.
188. Dutch Health Care Insurance Board [College voor Zorgverenigingen (CVZ)]. Reimbursement advice for ranibizumab (lucentis®) for age-related macula degeneration [pakketadvies ranibizumab (lucentis®) bij leeftijdsgebonden maculadegeneratie] 2012.
189. Schippers EI. Proposals for provisional and conditional inclusion of new drugs in the drug reimbursement scheme [voorstellen voor voorlopige en voorwaardelijke opname van nieuwe geneesmiddelen in het pakket] 2011.
190. Petrou S, Campbell N. Stabilisation in colorectal cancer. Int J palliat nurs 1997; 3: 275 - 280.

Summary

SUMMARY

The development of new and expensive health care technologies has increased pressure on national health care budgets as well as hospital budgets, leading to difficult questions about the affordability of new medicines. In order to strike an optimal balance between ensuring timely access to new drugs and having sufficient evidence of their relative benefits and risks, the Dutch Healthcare Insurance Board introduced a new expensive drugs policy regulation in 2006. This policy aimed to ensure undelayed access of promising but expensive hospital medicines by reimbursing hospitals (most of) the drug costs of drugs placed on the 'expensive drug list'. This reimbursement, however, is conditional for four years, after which a re-assessment takes place. The decision about whether or not to continue reimbursement is mainly based on the appropriate use and the cost-effectiveness of the new medicine as measured in Dutch real-world practice. The mandatory use of real-world data is a recent development in Dutch policymaking. Debates regarding the feasibility and added value of real-world studies are ongoing as experience is limited. The advantage of real-world data is that it provides policy makers with more relevant information on costs and effects in daily practice compared to clinical trials. However, conducting observational studies aimed to determine the relative cost-effectiveness of new treatments involves important methodological challenges. One of the main concerns is not being able to randomly assign treatments to patients.

This thesis examines the added value (potential) of conducting real-world studies and addresses several related methodological challenges (pitfalls) in the field of expensive inpatient oncologic drugs. This is done in three linked parts, each addressing a specific question. Part 1 consists of a detailed description of a case study where the use, effectiveness and cost-effectiveness of the drug oxaliplatin in the adjuvant treatment of stage III colon cancer are studied. Part 2 addresses the feasibility and several related methodological challenges in the use of real-world data. Lastly, part 3 describes a simulation study that was conducted to explore one particular methodological challenge in more detail; namely, how to correctly adjust for confounding by indication.

1. How is oxaliplatin used in the adjuvant treatment of stage III colon cancer in real-world practice and what are the real-world costs, effects and cost-effectiveness?

For stage III colon cancer patients, chemotherapy adjuvant to initial surgery has clearly been shown to lengthen (disease-free) survival. Until 2005, treatment with intravenous 5-fluorouracil and leucovorin (5FU/LV) was the only available effective adjuvant chemotherapy for patients with stage III colon carcinoma. In 2004, the drug oxaliplatin was officially registered for the adjuvant treatment of stage III colon cancer, based on a large phase III clinical trial. This MOSAIC trial had demonstrated that the addition of oxaliplatin to 5FU/LV improved disease-free survival compared to 5FU/LV alone and showed acceptable toxicity. Based on the trial results, 6 months of oxaliplatin combined with 5FU/LV (together referred to as FOLFOX) became the standard adjuvant treatment in the Netherlands for stage III colon cancer patients as of early 2005. Treatment with 5FU/LV alone remained indicated for patients who were not eligible or refused treatment with oxaliplatin. At that time the oral fluoropyrimidine capecitabine, alone or in combination with oxaliplatin (CAPOX), also became available as an alternative to 5FU/LV or FOLFOX, as these treatments were found to be equally effective in the treatment of stage III colon cancer or metastatic colorectal cancer.

Oxaliplatin as adjuvant treatment in stage III colon cancer was included in the 'expensive drug list' in early 2005. At that time uncertainty existed regarding its appropriate use (i.e., whether or not it would be used as expected in daily practice) and real-world (cost-) effectiveness. To evaluate these elements in Dutch daily practice, a retrospective population-based study was performed, which included 391 patients treated with adjuvant chemotherapy for stage III colon cancer in 2005-2006. Data were gathered from the Dutch Cancer Registry and medical records of 19 hospitals. The following chapters are based on the results of this study.

Chapter 2 evaluates the current guideline for the treatment of stage III colon cancer by examining guideline implementation, treatment patterns and disease-free survival. The results indicated that there was a rapid adoption of oxaliplatin, in the period shortly after it was reimbursed for the adjuvant treatment of colon cancer. In our study, 281 patients received oxaliplatin and 110 did not. Patients receiving oxaliplatin were younger and had less comorbidity than

other patients. The dosages used in daily practice were comparable to those observed in clinical trials. Patients receiving oxaliplatin in daily practice who fulfilled the MOSAIC study eligibility criteria (82%) had 2-year disease-free survival and overall survival outcomes that were comparable to those of the MOSAIC oxaliplatin patients. Patients who did not fulfil the MOSAIC study eligibility criteria (18%) had poorer disease-free and overall survival outcomes, which can be mainly explained by their poorer baseline prognosis (based on tumour marker values). In general, real-world practice seemed in good concordance with the RCT-based treatment recommendations. However, uncertainty remains regarding the optimal treatment of elderly patients and patients with comorbidities, which underscores the need for practical clinical trials that include these patients.

Chapter 3 describes the real-world resource use and costs of the adjuvant treatment for stage III colon cancer. The results showed that the mean total costs per patient ranged from €9,681 (5FU/LV) to €9,736 (capecitabine) for patients on fluoropyrimidine (FL) only therapy. Adding oxaliplatin to FL significantly increased costs, although the mean total costs for patients receiving FL + oxaliplatin varied considerably depending on the type of FL chosen. For patients receiving CAPOX the costs amounted to €18,361, while for patients receiving FOLFOX these costs were €32,793. The higher costs of FOLFOX are predominantly explained by the high number of hospitalisation days used for the 48-hour administration of FOLFOX. Given the diversity found in trials and in daily practice, it is questionable whether expensive hospitalisations are necessary for the administration of FOLFOX. A potential for substantial cost savings exists if the 48 hour administration of 5FU/LV in the FOLFOX regimen were to take place at home or be replaced by the oral drug capecitabine like in the CAPOX regimen. This analysis based on detailed real-life data clearly indicates that clinical choices made in oncology based on efficacy of therapy have economic consequences. Considering today's reality of finite healthcare resources, these economic consequences deserve a formal role in clinical decision making, for instance in guideline development.

Chapter 4 focuses on the incremental cost-effectiveness of oxaliplatin in daily practice compared to FL. A Markov model was developed to estimate lifetime cost and quality-adjusted life-years (QALYs) from a hospital perspective. Real world incremental effectiveness was estimated by combining MOSAIC trial ef-

fectiveness data with evidence from daily practice to preserve both the validity of the RCT and the representativeness of daily practice data. All cost inputs were based on Dutch daily practice. Cost-effectiveness analyses were performed for four different scenarios: (1) cost-effectiveness analyses based only on MOSAIC trial patients; (2) cost-effectiveness analyses using a combination of MOSAIC and eligible RW patients; (3) cost-effectiveness analyses using a combination of MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had an equal relative effect in ineligible and eligible patients;(4) cost-effectiveness analyses using MOSAIC and both eligible and ineligible RW patients, assuming oxaliplatin had no effect amongst ineligibles. The ICERs using the different scenarios, which ranged from €8,388 to €12,746, were all considered acceptable under a wide variety of model assumptions and therefore support the use of oxaliplatin in the adjuvant treatment of stage III colon cancer. This chapter illustrates how one could design and implement a real-world cost-effectiveness study to yield internally valid results that could also be generalisable.

2. What is the feasibility of assessing appropriate drug use and real-world cost-effectiveness in oncologic drugs and what are the most important challenges in light of the Dutch expensive drug policy? How can these challenges best be addressed?

In Chapter 5 the feasibility of real-world cost-effectiveness studies and the most important methodological challenges were discussed. Three different case studies were used as empirical input: i) oxaliplatin in stage III colon cancer, ii) oxaliplatin in metastatic colorectal cancer and iii) bortezomib in patients with multiple myeloma. It was feasible in all case studies to develop evidence on appropriate drug use, including information on clinical effectiveness and costs, based on real-world data. It was also feasible to compare these real-world results to the results of RCTs. However, this feasibility only applied to treatment with the new expensive drug and not to the comparator treatment. This was however not the case for the comparator treatment. The patient group receiving the comparator in daily practice was neither comparable to the patient group receiving the new expensive drug in daily practice nor the patient groups in the RCTs. This prohibited the feasibility to determine the real-world comparative effectiveness and cost-effectiveness using only data from Dutch

real-world practice. In the bortezomib case study a high degree of treatment heterogeneity was observed. Many different chemical agents were given in different lines as monotherapy or in different combinations, and in different sequences. The study design of the adjuvant oxaliplatin case study (see part 1) led to a study population in which oxaliplatin-treated patients differed greatly from the patients receiving the comparator treatment, with respect to their baseline prognostic variables. This was further complicated by missing values of important prognostic characteristics, a problem that arose in all case studies. As a consequence, no statistical method was able to adequately adjust for confounding. Potential solutions for these challenges can be found in two directions: 1) optimise the design of the real-world study in such a way that it allows the estimation of real-world incremental effects and costs based on real-world data by making adequate adjustment for confounding possible, or 2) estimate the real-world cost-effectiveness based on the combination of real-world data with other sources of evidence, like the example of chapter 4.

3. How should propensity-score based adjustment methods be applied in observational cost-effectiveness studies?

Chapter 6 and 7 of this thesis are based on a simulation study, where several confounding adjustment methods are compared in the context of observational cost-effectiveness analyses in a hypothetical cohort of patients with metastatic colorectal cancer. These methods include regression and propensity score (PS) methods. The propensity score is the conditional probability of receiving the treatment under study, given a patients characteristics. The PS methods were: PS matching (kernel and one-to-one), covariate adjustment using PS, inverse probability-of-treatment weighting, weighting by the odds and double robustness. These methods were compared in their ability to produce accurate (Chapter 6) and precise (Chapter 7) estimates, while different model specifications were considered.

Based on Chapters 6 and 7, certain suggestions can be made regarding cost-effectiveness analysis using observational time-to-event data. First, 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). It is important to

be aware of these options and choose the method based on what the population of interest is. If an ATT is required, matching methods are the best candidates. However, if the outcome of interest is an ATE, inverse probability -of-treatment weighting and double robustness are most likely to lead to the most acceptable results. Covariate adjustment using the propensity score performed the worst. 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. Since no method is always superior, it is advised that sensitivity analyses with different techniques be performed.

Chapter 8 draws together the results of previous chapters, provides a discussion of the results and explores the implications and the limitations of this thesis. In the context of conditional reimbursement, real-world observational studies are best used to evaluate the real-world applicability of evidence derived largely through randomised trials, to study patients and conditions not typically included or studied in randomised trials, to better understand current treatment practices, and to obtain insights regarding resource use and costs. In this thesis, diverse challenges that complicate the evaluation of the appropriate use, relative effectiveness and cost-effectiveness of new drugs were identified. The extent to which these challenges can be overcome within the set timeframe depends on the type of drug, number of patients, heterogeneity and natural history of disease. As a consequence, there is not just one formula for evidence building when it comes to drugs for different diseases and indications; a tailor-made approach is necessary. We suggest that a real-world incremental cost-effectiveness estimate should be based on a synthesis of the evidence drawn from different sources, including real-world experiences with the drug, evidence from RCTs (including long-term follow-up) and other relevant evidence available from the literature.

Samenvatting

SAMENVATTING

De ontwikkeling van nieuwe en duurdere gezondheidszorg zorgt voor een stijgende druk op de nationale gezondheidszorg- en ziekenhuisbudgetten. Dit leidt tot moeilijke vraagstukken ten aanzien van de betaalbaarheid van nieuwe geneesmiddelen. Om tot een optimale balans te kunnen komen tussen een snelle toegang tot nieuwe geneesmiddelen en voldoende bewijs ten aanzien van zijn relatieve voordelen en risico's, heeft het College Voor Zorgverzekeraars in 2006 een nieuwe beleidsregel geïntroduceerd. Deze beleidsregel had als doel om onvertraagde toegang van bepaalde veelbelovende dure specialistische geneesmiddelen te realiseren door ziekenhuizen (het merendeel van) de geneesmiddelkosten te vergoeden. Onder deze beleidsregel worden deze geneesmiddelen echter voorwaardelijk vergoed voor een periode van vier jaar, waarna een herbeoordeling plaatsvindt. Het besluit om deze vergoeding wel of niet te continueren is voornamelijk gebaseerd op het passend gebruik en de kosteneffectiviteit van het nieuwe geneesmiddel zoals gemeten in de Nederlandse dagelijkse praktijk. Aangezien de ervaring ten aanzien van het gebruik van dit soort data in de Nederlandse beleidsvoering nog relatief beperkt is, zijn er tot op de dag van vandaag discussies betreffende zijn haalbaarheid en toegevoegde waarde. Het grote voordeel van het gebruik van data uit de dagelijkse praktijk is dat het de beleidsmaker relevantere inzichten geeft ten aanzien van kosten effecten in de dagelijkse praktijk dan klinische trials. De uitvoering van observationele studies in de dagelijkse praktijk met als doel de relatieve kosteneffectiviteit te bepalen van nieuwe behandelingen gaat echter gepaard met belangrijke methodologische uitdagingen. Hierbij is het onvermogen om behandelingen in de dagelijkse praktijk gerandomiseerd toe te kunnen kennen aan patiënten een grote zorg.

Dit proefschrift onderzoekt de toegevoegde waarde (het potentieel) van de uitvoering van observationele studies in de dagelijkse praktijk en adresseert diverse gerelateerde methodologische uitdagingen (valkuilen) op het gebied van dure specialistische oncologische geneesmiddelen. Dit is beschreven in drie samenhangende delen die ieder een specifieke onderzoeksvraag behandelen. In deel 1 wordt een voorbeeld studie beschreven waarin het gebruik, de effectiviteit en de kosteneffectiviteit van het geneesmiddel oxaliplatin bestudeerd zijn in de adjuvante behandeling van het stadium III colon carcinoom. Deel

2 adresseert de haalbaarheid en verscheidenen gerelateerde methodologische uitdagingen in het gebruik van data uit de dagelijkse praktijk. Als laatste beschrijft deel 3 een simulatie studie, die uitgevoerd is om 1 specifieke methodologische uitdaging meer gedetailleerd te onderzoeken: hoe op een correcte manier te corrigeren voor het fenomeen *confounding by indication*.

1. Hoe wordt oxaliplatin in de dagelijkse praktijk gebruikt als adjuvante behandeling bij stadium III colon carcinoom en wat zijn de kosten, effecten en kosteneffectiviteit in de dagelijkse praktijk.

Een adjuvante behandeling met chemotherapie na een initiële chirurgische behandeling heeft duidelijk aangetoond de (ziektevrije) overleving van patiënten met stadium III colon carcinoom te verlengen. Tot 2005 was een behandeling met intraveneus 5-fluorouracil en leucovorine (5FU/LV) de enige beschikbare effectieve adjuvante chemotherapie voor patiënten met dit stadium colon carcinoom. Maar gebaseerd op een grote fase III klinische studie werd in 2004 het geneesmiddel oxaliplatin officieel geregistreerd voor de indicatie van adjuvante behandeling in patiënten met een stadium III colon carcinoom. Deze klinische studie, de MOSAIC trial, toonde aan dat de toevoeging van oxaliplatin aan 5FU/LV, wanneer vergeleken met alleen 5FU/LV, leidde tot een verbetering van de ziektevrije overleving met acceptabele toxiciteit. Gebaseerd op deze trial resultaten werd een behandeling van 6 maanden oxaliplatin met 5FU/LV (Samen aangeduid als FOLFOX) in Nederland begin 2005 de standaard adjuvante behandeling bij een stadium III colon carcinoom. Een behandeling met alleen 5FU/LV bleef geïndiceerd voor die patiënten die niet geschikt bevonden werden voor een behandeling met oxaliplatin of deze behandeling weigerden. In die tijd kwam ook het orale fluoropyrimidine (FL) capecitabine beschikbaar, alleen of in combinatie met oxaliplatin (CAPOX). Dit kon als alternatief voor 5FU/LV of FOLFOX ingezet worden, omdat deze behandelingen even effectief bevonden waren in de behandeling van stadium III of gemetastaseerd colorectaal carcinoom. Ook werd oxaliplatin als adjuvante behandeling begin 2005 opgenomen op de beleidsregel dure geneesmiddelen. Op dat moment was er geen informatie beschikbaar ten aanzien van gepast gebruik (of oxaliplatin in de dagelijkse praktijk gebruikt zou worden zoals verwacht) en de relatieve kosteneffectiviteit in de dagelijkse praktijk. Om deze elementen te kunnen evalueren

werd een retrospectieve observationele studie uitgevoerd, representatief voor de gehele Nederlandse populatie. Deze studie includeerde 391 patiënten die in 2005 of 2006 adjuvante chemotherapie voor het stadium III colon carcinoom ontvingen. Data werden verkregen via de Nederlandse Kanker Registratie en op basis van patiëntendossiers in 19 ziekenhuizen. De volgende hoofstukken van dit proefschrift zijn gebaseerd op de resultaten van deze studie.

In hoofdstuk 2 wordt de huidige richtlijn voor de behandeling van het stadium III colon carcinoom geëvalueerd door middel van het onderzoeken van de implementatie van deze richtlijn, geobserveerde behandelpatronen in de praktijk en de ziektevrije overleving. De resultaten brachten naar voren dat er een snelle opname van oxaliplatin als standaardbehandeling van het stadium III colon carcinoom plaatsvond in de periode kort nadat oxaliplatin voor deze indicatie op de beleidsregel werd geplaatst. In totaal 281 patiënten in onze studie ontvingen oxaliplatin en 110 niet. Patiënten die oxaliplatin voorgeschreven kregen hadden waren gemiddeld jonger en hadden minder comorbiditeiten dan patiënten zonder oxaliplatin. De doseringen zoals gebruikt in de dagelijkse praktijk kwamen overeen met de doseringen in klinische studies. De patiënten die oxaliplatin in de dagelijkse praktijk ontvingen en die ook voldeden aan de MOSAIC studie inclusie criteria (82%), hadden na 2 jaar een vergelijkbare ziektevrije en algehele overleving als MOSAIC studie patiënten in de oxaliplatin arm. Patiënten die niet voldeden aan deze MOSAIC inclusie criteria (18%) hadden ongunstigere ziektevrije en algehele overlevingsuitkomsten. Dit kan voornamelijk verklaard worden door de ongunstigere uitgangswaarden die deze patiënten hadden (gebaseerd op prognostisch ongunstige tumor marker waarden). In het algemeen kan geconcludeerd worden dat de dagelijkse praktijk goed in overeenstemming bleek met de aanbevelingen die gebaseerd op de MOSAIC trial geformuleerd waren. Er bleef echter onduidelijkheid bestaan ten aanzien van de optimale behandeling van oudere patiënten met comorbiditeiten. Dit onderstreept de noodzaak voor praktische gerandomiseerde studies die ook deze patiënten includeren.

Hoofdstuk 3 beschrijft het zorggebruik en kosten van de adjuvante behandeling van het stadium III colon carcinoom in de dagelijkse praktijk. De resultaten laten zien dat, voor patiënten behandeld met alleen fluoropyrimidinen (FL), de gemiddelde totale kosten per patiënt varieerden van €9.681 (5FU/LV) tot €9.736 (capecitabine). De toevoeging van oxaliplatin leidde altijd tot een substantiële

verhoging van de kosten, hoewel de gemiddelde totale kosten sterk afhingen van het type FL waarmee het gecombineerd werd. Voor patiënten die CAPOX toegediend kregen bedroegen de kosten €18.361 en voor patiënten met FOLFOX waren deze €32.793. De hogere kosten van FOLFOX worden voornamelijk verklaard door het hogere aantal ziekenhuisopnamen dat gebruikt werd voor de 48-uur durende infusie van FOLFOX. Gegeven de gevonden diversiteit in studies en dagelijkse praktijk, is het de vraag of de relatief dure ziekenhuisopnamen noodzakelijk zijn voor de toediening van FOLFOX. Er bestaat een potentieel voor substantiële kostenbesparingen wanneer deze 48-uur durende toediening van 5FU/LV in FOLFOX thuis zou plaatsvinden of vervangen kan worden door het orale capecitabine zoals in CAPOX. Deze analyse welke gebaseerd is op gedetailleerde data uit de dagelijkse praktijk brengt duidelijk naar voren dat de gemaakte klinische keuzen gebaseerd op de effectiviteit van behandelingen, economische consequenties hebben. Gegeven de huidige beperkte gezondheidszorgbudgetten verdienen dergelijke economische consequenties een formele rol in de besluitvorming, bijvoorbeeld bij de ontwikkeling van richtlijnen.

Hoofdstuk 4 is gericht op de incrementele kosteneffectiviteit van oxaliplatin versus FL in de dagelijkse praktijk. Een Markov model werd ontwikkeld om over de totale levensduur van een patiënt alle kosten en voor kwaliteit van leven gecorrigeerde levensjaren (QALYs) te kunnen berekenen vanuit een ziekenhuisperspectief. De incrementele effectiviteit in de dagelijkse praktijk werd berekend door MOSAIC trial effectiviteit te combineren met gegevens uit de dagelijkse praktijk. Dit werd gedaan om zowel de validiteit van een gerandomiseerde studie als de representativiteit van de dagelijkse praktijk te kunnen waarborgen. Alle kosteninputs werden gebaseerd op de Nederlandse dagelijkse praktijk. De kosteneffectiviteitsanalyses werden uitgevoerd in vier verschillende scenario's: (1) een kosteneffectiviteitsanalyse gebaseerd op alleen MOSAIC trial patiënten; (2) een kosteneffectiviteitsanalyse gebaseerd op MOSAIC en patiënten uit de dagelijkse praktijk die voldeden aan de MOSAIC inclusie criteria; (3) een kosteneffectiviteitsanalyse gebaseerd op een combinatie van MOSAIC patiënten en alle oxaliplatin patiënten uit de dagelijkse praktijk, waarbij aangenomen werd dat oxaliplatin een gelijke relatief effect had in patiënten die wel en niet aan de MOSAIC studie criteria voldeden; (4) een kosteneffectiviteitsanalyse gebaseerd op een combinatie van MOSAIC patiënten en alle oxaliplatin pati-

enten uit de dagelijkse praktijk, waarbij aangenomen werd dat oxaliplatin geen effect had onder patiënten die niet voldeden aan de MOSAIC inclusie criteria. De incrementele kosteneffectiviteitsratio's in de verschillende scenario's, welke varieerden van €8.388 tot €12.746, werden allen acceptabel bevonden onder een groot aantal model assumpties en ondersteunen het gebruik van oxaliplatin in de adjuvante behandeling van het stadium III colon carcinoom. Dit hoofdstuk illustreert hoe een kosteneffectiviteitsstudie op een zodanige manier ontworpen kan worden dat er intern valide resultaten gegenereerd worden die ook nog generaliseerbaar zijn naar de dagelijkse praktijk.

2. Wat is de haalbaarheid van onderzoek naar gepast gebruik en de kosteneffectiviteit van oncologische geneesmiddelen in de dagelijkse praktijk en wat zijn de meest belangrijke uitdagingen in het kader van de beleidsregel dure geneesmiddelen? Hoe kunnen deze uitdagingen het best geadresseerd worden?

In hoofdstuk 5 stonden de haalbaarheid van kosteneffectiviteitsstudies in de dagelijkse praktijk en de meest belangrijke methodologische uitdagingen ter discussie. Hiervoor werden drie verschillende empirische voorbeeld studies gebruikt: i) oxaliplatin in stadium III colon carcinoom, ii) oxaliplatin in gemetastaseerd colon carcinoom en iii) bortezomib in patiënten met een multipel myeloom. Het bleek in alle voorbeeld studies haalbaar om op basis van data uit de dagelijkse praktijk bewijs te verzamelen ten aanzien van gepast gebruik, inclusief klinische effectiviteit en kosten. Het bleek ook haalbaar om deze resultaten uit de dagelijkse praktijk te vergelijken met resultaten van klinische studies. Deze haalbaarheid was echter alleen van toepassing op informatie ten aanzien van het nieuwe geneesmiddel en niet op de vergelijkende behandeling. De patiëntengroep die de vergelijkende behandeling in de dagelijkse praktijk ontving, bleek zowel niet vergelijkbaar met de patiëntengroep behandeld met het nieuwe geneesmiddel, als niet met de patiëntengroep geïncludeerd in klinische studies. Dit verhinderde de mogelijkheid om de vergelijkende effectiviteit en kosteneffectiviteit van de nieuwe geneesmiddelen vast te kunnen stellen op basis van alleen data uit de Nederlandse dagelijkse praktijk. In de bortezomib studie werd een hoge mate van behandel heterogeniteit geobserveerd. Vele verschillende middelen werden gegeven in verschillende behandellijnen,

hetzij als monotherapie, hetzij in verschillende combinatie behandelingen en in verschillende volgordes. De studieopzet van de studie naar oxaliplatin als adjuvante behandeling (zie deel 1) leidde tot een studiepopulatie waarin patiënten die behandeld waren met oxaliplatin enorm verschilden ten aanzien van hun prognostische variabelen, van patiënten die de vergelijkende behandeling ontvingen. Dit werd verder gecompliceerd door ontbrekende waarden bij belangrijke prognostische variabelen, hetgeen een probleem bij alle voorbeeld studies was. Dit had als consequentie dat geen enkele statistische methode in staat bleek om adequaat te kunnen corrigeren voor de verschillen in prognostische factoren (*confounding*). Mogelijke oplossingen voor deze uitdagingen kunnen gezocht worden in twee richtingen: 1) het optimaliseren van de studieopzet van de observationele studie in de dagelijkse praktijk, op een zodanige manier dat het mogelijk wordt om adequaat te corrigeren voor confounding, of 2) het berekenen van de kosteneffectiviteit gebaseerd op een combinatie van data uit de dagelijkse praktijk en andere bronnen van bewijs (klinische studies), zoals in het voorbeeld van hoofdstuk 4.

3. Hoe zouden propensity score methoden toegepast dienen te worden in observationele kosteneffectiviteitsstudies?

Hoofdstuk 6 en 7 van dit proefschrift zijn gebaseerd op een simulatie studie in een hypothetisch cohort van patiënten met gemetastaseerd colorectaal carcinoom waarin verscheidene correctiemethoden vergeleken werden in de context van observationele kosteneffectiviteitsstudies. Deze methoden waren conventionele regressie en *propensity score* methoden. De *propensity score* van een patiënt is de berekende kans dat iemand met dezelfde kenmerken als deze patiënt de onderzochte behandeling krijgt. De *propensity score* methoden waren: *propensity score matching* (met kernel en 1-op-1), regressie met de *propensity score* als covariaat, *inverse probability-of- treatment weighting* (IPTW), *weighting by the odds* en *double robustness*. Deze methoden werden vergelijken in hun vermogen om accurate (Hoofdstuk 6) en precieze (Hoofdstuk 7) schattingen te produceren, terwijl verschillende model specificaties beschouwd werden. Gebaseerd op hoofdstuk 6 en 7 kunnen zekere aanbevelingen gemaakt worden ten aanzien van het toepassen van kosteneffectiviteitsanalysen waarbij gebruik gemaakt wordt van observationele overlevingsdata. Allereerst is

het conceptueel belangrijk om onderscheid te maken tussen methoden die een gemiddeld behandel effect in behandelde patiënten schatten (ATT) en methoden die een gemiddeld behandel effect in de totale populatie schatten (ATE). Het is belangrijk om de passende methode te kiezen die aansluit op de populatie waarin men geïnteresseerd is. Wanneer een ATT nodig is zijn *matching methoden* de beste kandidaten. Is er echter een ATE nodig, dan leiden inverse probability- of – treatment weighting en double robustness het meest waarschijnlijk naar acceptabele resultaten. Regressie met de *propensity score* als covariaat presteerde het slechts. De voorbehandelingsbenadering, waarbij een volledig regressiemodel wordt toegepast nadat eerst matching of weging heeft plaatsgevonden, combineerde accuratesse met relatieve precisie. Aangezien geen enkele methode altijd superieur is, wordt aanbevolen om een gevoeligheidsanalyse met verschillende technieken uit te voeren.

In hoofdstuk 8 worden de resultaten van de voorgaande hoofdstukken samengebracht en bediscussieerd, en de implicaties en limitaties van dit proefschrift verkend. In de context van het voorwaardelijke vergoedingssysteem, kunnen studies die uitgevoerd worden in de dagelijkse praktijk het best ingezet worden om 1) de toepasbaarheid in de dagelijkse praktijk te testen van onderzoeksbewijs zoals gevonden in gerandomiseerde klinische studies, 2) patiënten en condities te bestuderen die gewoonlijk niet geïncludeerd worden in gerandomiseerde klinische studies, 3) een beter inzicht te krijgen in huidige behandelpraktijken en 4) inzichten te verkrijgen ten aanzien van zorggebruik en kosten. In dit proefschrift werden diverse uitdagingen op het gebied van de evaluatie van passend gebruik, incrementele effectiviteit en kosteneffectiviteit geïdentificeerd. In hoeverre deze uitdagingen overwonnen kunnen worden binnen de vastgestelde tijdsperiode van vier jaar, hangt af van het type geneesmiddel, het aantal patiënten, heterogeniteit van behandelpatronen en het natuurlijk ziektebeloop. Als gevolg hiervan bestaat er geen standaard formule voor het uitvoeren van onderzoek naar passend gebruik en kosteneffectiviteit in de dagelijkse praktijk. Bij voorkeur schat men de incrementele kosteneffectiviteit in de dagelijkse praktijk op basis van een combinatie van informatie van verschillende bronnen, waaronder ervaringen uit de dagelijkse praktijk, bewijs van gerandomiseerde studies (inclusief eindresultaten op de lange termijn) en alle andere relevante bronnen beschikbaar uit de literatuur.

About the author

ABOUT THE AUTHOR

Chantal Wilhelmina Martina was born on August 2, 1979, in Raamsdonk, The Netherlands. She graduated from secondary school at the Gymnasium St. Willibrord in Deurne in 1997. In that year she started her higher education at the University of Utrecht School of Medicine. After obtaining her medical degree in 2004, she worked as a resident (ANIOS) in the Intensive Care Unit at the Gelderse Vallei Hospital in Ede. From September 2005, she worked for six years as Duty Physician at the Clinical Pharmacology Unit of Kendle International B.V. in Utrecht. In 2007, she received her Master's degree in Health Economics from the Erasmus University Rotterdam, after which she started working at the institute for Medical Technology Assessment and at the institute of Health Policy & Management. Her research focuses mainly on methodological issues of real-world economic evaluations (involving issues on costs as well as effects of health care interventions). In 2009 she graduated with a Master of Science in Clinical Epidemiology from the Netherlands Institute for Health Sciences in Rotterdam. She has published papers in peer-reviewed journals such as Pharmacoeconomics and European Journal of Cancer. As of January 2012, she is working as a Manager Health Outcomes at GlaxoSmithKline.

List of publications

LIST OF PUBLICATIONS

van der Linden N, van Gils CW, Pescott CP, Buter J, Uyl-de Groot CA. Cetuximab in locally advanced squamous cell carcinoma of the head and neck: generalizability of EMR 062202-006 trial results. Eur Arch Otorhinolaryngol. 2013 Aug 2. [Epub ahead of print]

van Gils CW, de Groot S, Redekop WK, Koopman M, Punt CJ, Uyl-de Groot CA. Real-world cost-effectiveness of oxaliplatin in stage III colon cancer: a synthesis of clinical trial and daily practice evidence. Pharmacoeconomics. 2013 Aug; 31(8):703-18.

Mol L, Koopman M, van Gils CW, Ottevanger PB, Punt CJ. Comparison of treatment outcome in metastatic colorectal cancer patients included in a clinical trial versus daily practice in The Netherlands. Acta Oncol. 2013 Jun; 52(5):950-5.

Franken MG*, van Gils CW*, Gaultney JG, Delwel GO, Goettsch W, Huijgens PC, Steenhoek A, Punt CJ, Koopman M, Redekop WK, Uyl-de Groot CA. Practical feasibility of outcomes research in oncology: lessons learned in assessing drug use and cost-effectiveness in The Netherlands. Eur J Cancer. 2013 Jan; 49(1):8-16.

van Gils CW, Koopman M, Mol L, Redekop WK, Uyl-de Groot CA, Punt CJ. Adjuvant chemotherapy in stage III colon cancer: guideline implementation, patterns of use and outcomes in daily practice in The Netherlands. Acta Oncol. 2012 Jan; 51(1):57-64.

Tan SS, van Gils CW, Franken MG, Hakkaart-van Roijen L, Uyl-de Groot CA. The unit costs of inpatient hospital days, outpatient visits, and daycare treatments in the fields of oncology and hematology. Value Health. 2010 Sep-Oct; 13(6):712-9.

Submitted manuscripts

van Gils CW, de Groot S, Tan SS, Redekop WK, Koopman M, Punt CJ, Uyl- de Groot CA.

Real-world resource use and costs of adjuvant treatment for stage III colon cancer.

van Gils CW*, Goossens L*, Al MJ, Koopman M, Uyl- de Groot CA, Redekop WK. A propensity to get it right: A Monte Carlo simulation study comparing statistical methods to obtain correct cost-effectiveness estimates in observational studies.

Goossens L*, van Gils CW*, Al MJ, Koopman M, Uyl- de Groot CA, Redekop WK. A propensity to be precise: A simulation study comparing statistical methods to obtain reliable cost-effectiveness estimates in observational studies.

Goossens L, van Gils CW, Redekop WK. Collapsibility and censoring: what's the bias in estimating time-to-event?

van der Linden N, van Gils CW, Pescott CP, Buter J, Vergeer MR, Uyl- de Groot CA. Real-world cost-effectiveness of cetuximab in locally advanced squamous cell carcinoma of the head and neck.

Mohseninejad L, van Gils CW, Uyl- de Groot CA, Buskens E, Feenstra TL. Evaluation of patient registries supporting reimbursement decisions: The case of oxaliplatin for treatment of stage III colon cancer.

Dankwoord

DANKWOORD

Een proefschrift schrijf je gelukkig niet alleen. Ik voel me bijzonder bevoorrecht dat ik ben begeleid door mijn co-promotor Ken Redekop en mijn promotoren Carin Uyl-de Groot en Kees Punt. Ken, van jou heb ik leren denken als een wetenschapper. Jouw vermogen om, ook als werk al bijna af lijkt, kritisch te blijven reflecteren is een groot voorbeeld. Bedankt dat de deur altijd voor me open stond. Carin, dank voor het vertrouwen en voor de ruimte die je me gaf. Je bent bevlogen om de toegankelijkheid van de juiste behandeling voor patiënten te bevorderen. Dit is inspirerend. Jouw doelgerichtheid heeft zeer bijgedragen aan het afronden van dit proefschrift. Kees, jouw klinische kennis op het gebied van het colorectaal carcinoom is van onmisbare waarde geweest. Bedankt voor de fijne samenwerking.

De promotiecommissie dank ik van harte voor het lezen en beoordelen van dit proefschrift en voor het opponeren bij de verdediging.

De hoofdstukken van dit proefschrift zijn ontstaan uit de samenwerking met gedreven collega-wetenschappers. Hun enthousiasme, kennis en inzicht waren een grote verrijking. In het bijzonder wil ik Lucas Goossens, Jennifer Gaultney en Margreet Franken bedanken. Lucas, wat was het hard werken om ons onderzoek tot goede uitvoer te brengen, maar het was super om dit samen met jou te kunnen doen. Bedankt voor je vriendschap. Margreet, ik ben blij dat we nog steeds samenwerken, ook al heb ik de universiteit verlaten. Jen, het is bijzonder om zoveel met je te delen op zowel persoonlijk als professioneel vlak.

Collega's van iBMG, ik kijk terug op fijne jaren waaraan jullie allemaal hebben bijgedragen. Collega's van GSK, het viel niet mee een proefschrift af te ronden en ondertussen te beginnen aan een nieuwe, uitdagende baan. Veel dank voor jullie hulp hierbij.

Ik kan in dit dankwoord met geen mogelijkheid recht doen aan de steun, uitdaging, troost en afleiding die ik bij mijn familie en vriendenkring vond en vind. Mijn dank aan jullie is groot. Ik kijk er naar uit meer tijd met jullie te kunnen doorbrengen.

Lieve ouders, Hanny en Jan, bedankt voor jullie onvoorwaardelijke liefde, het grenzeloze vertrouwen, de steun en de vrijheid die jullie me altijd gegeven hebben. Maarten en Ronald, mijn lieve broers, we hebben elkaar mooi kunnen

stimuleren in ons harde werken het afgelopen jaar. Ik ben ontzettend trots dat jullie onlangs allebei ingenieur geworden zijn.

Mijn paranimfen Miranda en Marieke, jullie zijn niet alleen mijn schoonzussen, maar ook dierbare vriendinnen. Jullie aandeel in het afronden van dit proefschrift is absoluut groot. Miranda, terwijl ik aan dit proefschrift schreef, heb jij veel tijd met Norah doorgebracht. Ik ben ontzettend dankbaar voor jullie bijzondere band. Marieke, vanaf de eerste dag dat we een kamer gingen delen op iMTA was duidelijk dat we eenzelfde Coca Cola light verslaving hadden en hebben. Immers, volgens ons komen alle goede ideeën uit een zilveren blikje met rode letters. Bedankt voor je vele hulp bij de laatste loodjes.

Lieve George, het afronden van dit proefschrift heeft de afgelopen twee jaar centraal gestaan in ons leven. Hoewel dit vaak ten koste ging van onze tijd samen, ben je me altijd blijven motiveren en steunen. Dank voor je liefde, humor en geduld. Het leven is mooi, jij maakt het nog mooier. Norah, jij maakt het compleet.