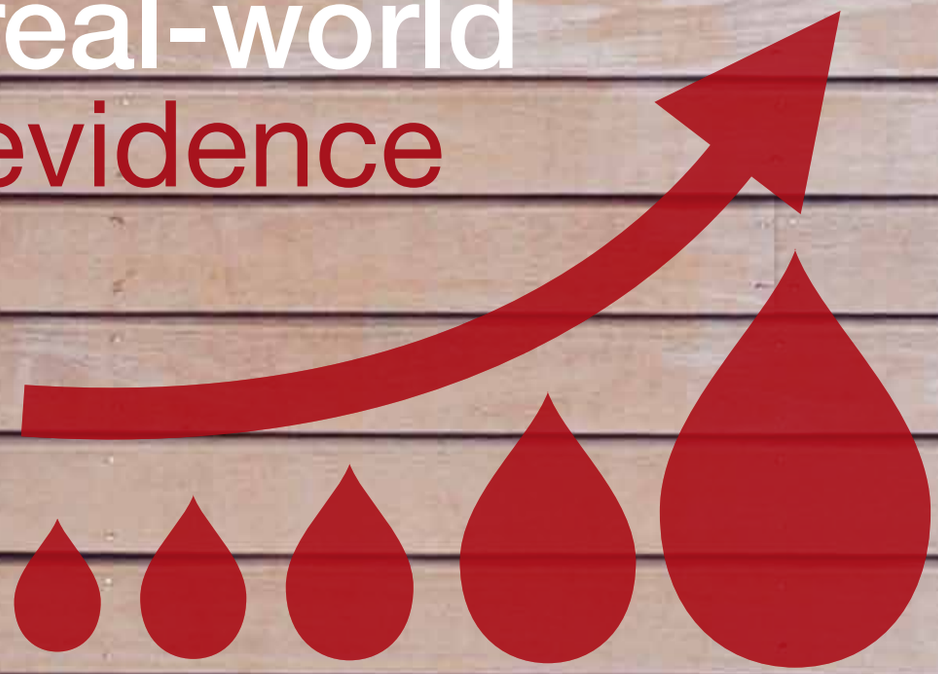
The added value of real-world evidence

Hedwig Blommestein

The added value of real-world evidence



Hedwig Blommestein

The added value of real-world evidence

De toegevoegde waarde van
gegevens uit de dagelijkse praktijk

Printing and lay out: Optima grafische communicatie Rotterdam, the Netherlands

ISBN: 978-94-6169-784-4

The Added Value of Real-World Evidence

De toegevoegde waarde van gegevens uit de dagelijkse praktijk

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

vrijdag 1 april 2016 om 13.30 uur

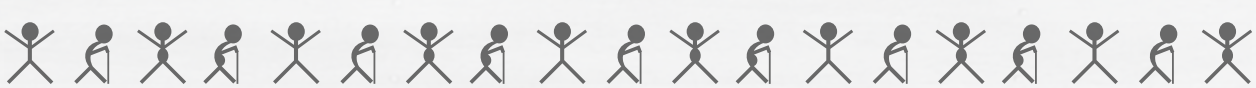
Hedwig Maria Sluimer-Blommestein
geboren te Zaanstad

Erasmus University Rotterdam

PROMOTIECOMMISSIE:

Promotoren:	Prof.dr. C.A. Uyl-de Groot
	Prof.dr. P.C. Huijgens
Overige leden:	Prof.dr. J.J. Cornelissen
	Prof.dr. M.J. IJzerman
	Prof.dr. J.L. Severens

CONTENTS



Chapter 1

General introduction

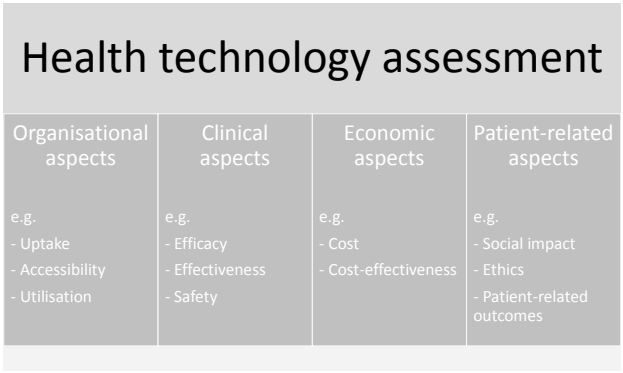
BACKGROUND

Average life expectancy at birth in the Organisation for Economic Co-operation and Development (OECD) countries increased from 77.1 in 2000 to 80.2 years in 2012[1]. While life expectancy is increasing, expenditures on health increased as well during the past decade and not only in absolute terms. In 2000, countries spent on health on average 7.7% of gross domestic product (GDP) and this increased to 9.3% of GDP in 2012. In OECD countries, most health expenditures are public expenditures (on average 72.2%). In 2012, the Netherlands spent 11.8% of GDP on health which was 70,514 billion, life expectancy was 81.2 years. Due to the development of new medical technologies and ageing of the population, health care expenditures in general[2] and for cancer in particular[3] are likely to increase further.

Most health expenditures are publicly financed. However, health expenditures are not the only public expenditures and expenditures on health have to compete with spending on for example social security, education, and infrastructure. In addition, although patients demand access to all health care programmes available, the scarcity of resources necessitates that political choices have to be made[4]. These choices are not limited to choices between different areas of public spending. Also within health care, choices have to be made since, for example, money spent on patients with cancer cannot be spent on patients with cardiovascular diseases or elderly care.

Given that health is such an important matter for every human and that political decisions usually have far-reaching consequences for individuals, the allocation of resources requires careful consideration preferably based on transparent and unbiased data. Cost in relation to the expected outcomes are most important in health technology assessment (HTA)[5]. HTA is a policy approach that examines the short- and long-term social consequences of the application or use of a health technology[4,6]. Health technologies include drugs, procedures, devices, and health programmes[7,8]. The main dimensions of HTA are presented in Figure 1.1 and include organisational, clinical, economic, and patient-related aspects[6]. Organisational aspects of health technologies include the uptake (i.e. whether a technology is adopted in daily practice), accessibility (i.e. whether the technology is available for patients who may benefit), and utilisation (i.e. how many patients receive the technology, by which regimen and in what frequency). Clinical aspects include the efficacy, effectiveness and safety of a technology. Economic aspects include costs, including budget impact, and cost-effectiveness of health technologies. Finally, patient-related aspects involve social impact, ethics and patient related outcomes. For a long time, HTA was mainly based on evidence from randomised controlled trials (RCTs). However, more recently, there is growing interest in HTA based on real-world data[7].

Figure 1.1 The dimensions of health technology assessment



Adapted from Draborg et al.[6]

DATA FROM RANDOMISED CONTROLLED TRIALS VERSUS REAL-WORLD DATA

RCTs are considered the golden standard for establishing efficacy since these studies demonstrate whether the treatment works and is safe under optimal and highly controlled circumstances[7]. Although RCTs ensure internal validity by ensuring optimal and highly controlled circumstances, the results are not generalisable to the context of care in daily practice (i.e. external validity). Technologies may not be adopted, adopted under different circumstances or applied to patients who do not fulfil the inclusion criteria of studies[9]. As a consequence, efficacy as demonstrated in RCTs will most certainly differ from effectiveness in daily practice[10,11]. Since RCT data might not be sufficient for making decisions in daily practice, real-world evidence is increasingly requested[7].

Real-world data are not collected through conventional RCTs but from real-world practice (i.e. not under controlled circumstances)[7]. Population based disease registries may provide a convenient way to collect real-world data[12,13]. Real-world data allow studying the organisational, clinical, economic and patient-related aspects of HTA. For example, evidence can be obtained on the uptake, accessibility and utilisation under daily practice circumstances. Furthermore, real-world data provide generalisable effectiveness and cost-effectiveness estimates of technologies. While findings from real-world data are of more practical value to health care decision makers, using real-world data also imposes methodological challenges due to the absence of random treatment assignment and uncontrolled circumstances. As a consequence, the evidence base regarding the actual added value of real-world data compared to RCTs is currently inconclusive. This thesis evaluates the added value of real-world evidence for health care decision makers.

CANCER AND HAEMATOLOGICAL MALIGNANCIES

The study of the main dimensions of HTA is especially relevant for disease areas with high treatment costs, rapid introduction of innovative technologies and rising incidence rates. In developed countries, cancer is a disease area that fulfils these criteria and delivering high quality and accessible care for patients with cancer is a challenge for decision makers[3]. First, cancer is a severe, often fatal disease and has a major impact on many people. Decisions regarding reimbursement of cancer treatment have far reaching consequences and are constantly the subject of public debate[14]. Second, the economic burden of cancer is high and expected to increase. In the European Union the economic burden of cancer was calculated to be €126 billion in 2009, €51 billion of which were health care costs[15]. Spending on cancer care is expected to increase due to ageing, improved diagnostics and treatment advances. For example, conventional chemotherapy, introduced during the second half of the twentieth century, improved the outcomes of patients with cancer and a much greater step forward is expected from the introduction of immunotherapy (i.e. therapies that stimulate the immune system to destroy tumours) and targeted therapies (i.e. therapies that target critical molecular pathways of tumours)[16]. However, these treatment advances are also associated with higher cost. For example in the Netherlands, the costs for expensive inpatient cancer drugs were €376 million in 2011 and increased to €675 million in 2014[17].

Malignancies of blood, bone marrow and lymph nodes, the so called haematological malignancies are a likely target for HTA. The previously mentioned criteria show an enormous development in technology and survival. Many patients with haematological malignancies are cured but, treatment also aims for prolonging survival without cure, especially in older people. Incidence for most haematological malignancies increases with age; the median age at diagnosis is above 65 years[18]. Although haematological malignancies are more prevalent among the elderly, very few RCTs focus on older patients. For example in non-Hodgkin lymphoma, only 10% of the RCTs focused on older patients and in 25% of the RCTs, patients were excluded if they were older than 65 years[19]. As a consequence, and for reasons discussed earlier, efficacy evidence from RCTs and economic evaluations based on RCT data in haematology will not provide sufficiently relevant information for health care decision makers. Data from real-world studies are eagerly awaited for health care decision makers. Therefore, haematological malignancies was selected as case study for evaluating the added value of real-world evidence.

THE PHAROS-REGISTRY

Real-world data of patients with haematological malignancies were available from the Population-based Haematological Registry for Observational Studies (PHAROS-registry)[20,21]. This Dutch registry was initiated in 2010 for three haematological malignancies: chronic lymphocytic leukaemia, non-Hodgkin lymphoma and multiple myeloma. Newly diagnosed patients were included from three Dutch regions; these regions cover 40% of the Netherlands[21]. Detailed real-world data of the PHAROS-registry (i.e. patient and disease characteristics, diagnostics, treatments, response to treatment and health care utilisation) supplemented the data of the Netherlands Cancer Registry (i.e. date of birth, sex and date of incidence)[22] The PHAROS-registry was set up to measure and improve the quality of haematological care and to provide a basis for assessing the cost-effectiveness of new treatments in a real-world setting.

OBJECTIVES

This thesis evaluates the added value of real-world evidence for health care decision makers regarding organisational, clinical and economic aspects of HTA. As a start, the shortcomings of evidence derived from RCTs are explored. Further insights in the added value of real-world evidence are obtained by addressing the different aspects of HTA using real-world data. Practical guidance is provided for how to best use real-world data. The main potentials, including the development of a full disease model, as well as the methodological challenges are described in detail.

This thesis addresses the following research questions:

1) What are the shortcomings of evidence from randomised controlled trials for health care decision makers?
2) What is the added value of real-world evidence for health care decision makers regarding
   A) organisational aspects of health technology assessment?
   B) clinical aspects of health technology assessment?
   C) economic aspects of health technology assessment?
3) What are the main methodological challenges for using real-world data to inform health care decision makers?

OUTLINE

This thesis consists of three parts. Part 1 includes Chapter 2 and reports on the limitations of RCT evidence for health care decision makers. A cost-effectiveness analysis based on RCT data for a novel treatment is presented. This study illustrates the difficulties in determining value for money for health care decision makers based on RCT data.

Part 2 focuses on the organisational, clinical and economic aspects of HTA using real-world data and consist of the Chapters 3, 4, and 5. Chapter 3 reports on the uptake, accessibility and utilisation of an expensive drug (i.e. bortezomib) in the Netherlands. This chapter illustrates how organisational aspects of HTA can be investigated with real-world data. Chapter 4 illustrates how economic aspects of HTA can be studied with real-world data and describes a real-world study on the costs of treatment (i.e. stem cell transplantations). In this chapter, a comparison between the costs in daily clinical practice and reimbursement is made. Both clinical and economic aspects of HTA are addressed in Chapter 5. In this chapter, the real-world effectiveness and cost-effectiveness of an expensive drug is presented on both RCT and real-world data.

Part 3 of this thesis provides practical guidance on using real-world data and highlights the methodological challenges for obtaining real-world evidence. Chapter 6 illustrates how real-world data can be used for HTA and describes methodological challenges and possible solutions for using real-world data. One of the potentials of real-world data is to develop full disease models and to study treatment sequences. This is illustrated in Chapter 7 and Chapter 8.

Finally, Chapter 9 discusses the main findings regarding organisational, clinical and economic aspect of HTA and the added value of real-world evidence for health care decision makers.

Chapter 2

Lenalidomide for the treatment of low- or intermediate-1 risk myelodysplastic syndromes associated with deletion 5q cytogenetic abnormality: an evidence review of the NICE submission from Celgene.

Blommestein HM, Armstrong N, Ryder S, Deshpande S, Worthy G, Noake C, Riemsma R, Kleijnen J, Severens JL, Al MJ

PharmacoEconomics. 2015 Aug 28. [Epub ahead of print]

ABSTRACT

The National Institute for Health and Care Excellence (NICE) invited the manufacturer of lenalidomide (Celgene) to submit evidence of the clinical and cost-effectiveness of the drug for treating adults with myelodysplastic syndromes (MDS) associated with deletion 5q cytogenetic abnormality, as part of the Institute's single technology appraisal (STA) process. Kleijnen Systematic Reviews Ltd (KSR), in collaboration with Erasmus University Rotterdam, was commissioned to act as the Evidence Review Group (ERG). This paper describes the company's submission, the ERG review, and the NICE's subsequent decisions.

The ERG reviewed the evidence for clinical and cost-effectiveness of the technology, as submitted by the manufacturer to the NICE. The ERG searched for relevant additional evidence and validated the manufacturer's decision analytic model to examine the robustness of the cost-effectiveness results.

Clinical effectiveness was obtained from a three-arm, European randomised phase III trial among red blood cell (RBC) transfusion-dependent patients with low-/intermediate-1 risk del5q31 MDS. The primary endpoint was RBC independence for ≥26 weeks, and was reached by a higher proportion of patients in the lenalidomide 10 and 5mg groups compared with placebo (56.1 and 42.6 vs 5.9%, respectively; both $p<0.001$). The option of dose adjustments after 16 weeks due to dose-limiting toxicities or lack of response made long-term effectiveness estimates, unreliable, e.g. overall survival (OS).

The de novo model of the manufacturer included a Markov state-transition cost-utility model implemented in Microsoft Excel. The base case incremental cost-effectiveness ratio (ICER) of the manufacturer was £56,965. The ERG assessment indicated that the modelling structure represented the course of the disease; however, a few errors were identified and some of the input parameters were challenged. In response to the appraisal documentation, the company revised the economic model, which increased the ICER to £68,125 per quality-adjusted life-year. The NICE Appraisal Committee (AC) did not recommend lenalidomide as a cost-effective treatment. Subsequently, the manufacturer submitted a Patient Access Scheme (PAS) that provided lenalidomide free of charge for patients who remained on treatment after 26 cycles. This PAS improved the ICER to £25,300, although the AC considered the proportion of patients who received treatment beyond 26 cycles, and hence the ICER to be uncertain.

Nevertheless, the AC accepted a commitment from the manufacturer to publish, once available, data on the proportion of patients eligible for the PAS and believed this provided reassurance that lenalidomide was a cost-effective treatment for low- or intermediate-1 risk MDS patients.

INTRODUCTION

The National Institute for Health and Care Excellence (NICE) is an independent organisation providing national guidance on promoting good health and preventing and treating ill health[23]. The single technology appraisal (STA) process is designed to provide recommendations and guidance on a single product, device or other technology with a single indication. The process covers new technologies and enables the NICE to produce guidance shortly after the technology is introduced in the United Kingdom (UK). The NICE Appraisal Committee (AC) obtains relevant evidence from several sources: the manufacturer's submission (MS), a report from the appointed independent Evidence Review Group (ERG) and advice from consultees (i.e. experts and other stakeholders). The MS includes a written report and a mathematical model that describe the clinical and cost-effectiveness of the technology under investigation. The ERG, an external organisation independent of the NICE, reviews the MS and produces a report. After consideration of all the relevant evidence, the AC formulates preliminary guidance in the form of the Appraisal Consultation Document (ACD) as to whether to recommend the intervention. The stakeholders are invited to comment on this ACD and the submitted evidence. A subsequent ACD may be produced or a Final Appraisal Determination (FAD) issued. The submission of a Patient Access Scheme (PAS) is allowed in order to allow the NICE to recommend treatments that would otherwise not have been found to be cost-effective. The PAS is a means of reducing the price of the drug by some means, e.g. simple discount or other formula, and has to be agreed by the Department of Health. This paper presents a summary of the ERG report and the development of NICE guidance based on the findings of the AC for the STA of lenalidomide for treating myelodysplastic syndromes (MDS) associated with deletion 5q (del5q) cytogenetic abnormality. Full details of all the relevant appraisal documents can be found on the NICE website[24]. This is one in a series of STA summaries being published in Pharmacoeconomics[25-30].

THE DECISION PROBLEM

MDS are a heterogeneous group of haematological disorders in which the bone marrow functions abnormally, causing peripheral blood cytopenia due to insufficient production of mature blood cells[31]. MDS can affect red blood cells (RBCs), white blood cells (WBCs) and platelets, resulting in anaemia, increase in bleeding, infection and disease transformation to acute myeloid leukaemia (AML)[32]. The quality of life of patients with MDS is impaired due to symptoms such as fatigue and dyspnoea as well as treatments involving hospitalisations with drug administration and blood transfusions. As reported in 2003, the incidence is approximately 4 per 100,000 population but rises to >30 per

100,000 in the over 70 years age group[33]. In the UK, there are approximately 11,200 patients diagnosed with MDS[34], a condition that is mainly caused by cytogenetic abnormalities found in marrow cells. The most common cytogenetic abnormality, present in approximately 15% of patients with MDS, is del5q[35].

Currently, there is no active treatment available for patients with MDS del5q since stem cell transplantations or treatment with azacitidine are not recommended for this patient group[33]. Patients receive best supportive care (BSC), which includes blood transfusions to control symptoms associated with bone marrow failure and antibiotics to treat or prevent infection. In addition, growth factors such as erythropoietin and/or granulocyte colony-stimulating factors to stimulate the production of RBCs and WBCs are prescribed.

Lenalidomide was already available in the UK for the treatment of relapsed refractory multiple myeloma. In 2013 the European Medicines Agency extended the market authorisation of lenalidomide to include patients with transfusion-dependent anaemia due to low- or intermediate-1 risk MDS associated with del5q when other therapeutic options were insufficient or inadequate[36]. Lenalidomide is an oral therapy that aims to reverse transfusion dependence.

NICE developed a scope for the assessment of lenalidomide, which specified that the clinical and cost-effectiveness of this drug should be established, relative to BSC for the treatment of patients with transfusion-dependent anaemia due to low- or intermediate-1 risk MDS associated with del5q cytogenetic abnormality with or without other cytogenetic abnormalities.

THE INDEPENDENT ERG REVIEW

Kleijnen Systematic Reviews Ltd (KSR), in collaboration with Erasmus University Rotterdam, acted as the ERG. The ERG reviewed the evidence on the product's clinical and cost-effectiveness among low- or intermediate-1 risk MDS del5q patients as submitted by the manufacturer (Celgene). The review embodied three aims:

- to assess whether the MS conformed to the methodological guidelines issued by the NICE[23]
- to assess whether the manufacturer's interpretation and analysis of the evidence was appropriate
- to indicate the presence of other sources of evidence or alternative interpretations of the evidence that could help to inform NICE guidance

The ERG critically reviewed the MS, conducted additional searches, explored the impact of assumptions on the incremental cost-effectiveness ratio (ICER), revised the economic model and explored additional scenario analyses. The ERG review detailed here relates to the evidence contained in the original MS and additional information submitted by the manufacturer in response to the clarification questions and ACD which included a PAS.

Clinical evidence

The MS included a systematic review of the literature on the clinical effectiveness of lenalidomide. Evidence on the efficacy of lenalidomide was extracted from the MDS-004 trial, a phase III, multicentre, randomised, double blind, placebo controlled study[37]. Adult patients with low- or intermediate-1 risk MDS with del5q, with or without additional cytogenetic abnormalities and RBC transfusion-dependent anaemia (N=205) were randomly assigned to three arms: lenalidomide 10 mg on days 1–21, lenalidomide 5 mg on days 1–28, or placebo on days 1–28 for each 4-week cycle. BSC included blood transfusions that were provided to all transfusion-dependent patients as required. If dose-limiting toxicities occurred, the dose of lenalidomide was reduced. Crossover was allowed at 16 weeks if at least a minor erythroid response (i.e. a 50% decrease in transfusion requirements) was not achieved, and all but 11 patients on the placebo arm crossed over to lenalidomide 5 mg. Before crossover at 16 weeks, two patients (3%) in the placebo group, two (2.9%) in the lenalidomide 5 mg group and none in the lenalidomide 10 mg group progressed to AML. The primary endpoint was RBC transfusion independence for ≥26 weeks, which was reached in 56.1, 42.6, and 5.9% of patients in the lenalidomide 10 mg, lenalidomide 5 mg and placebo groups, respectively. Transfusion-independent rates in both lenalidomide groups were different compared with placebo ($p< 0.001$). Median duration of transfusion independence was not reached in either lenalidomide group after a median follow up of 1.55 years. Of the patients who initially received placebo and crossed over to lenalidomide 5 mg, 30.4% progressed to AML, compared with 23.2% in the 5 mg group and 21.7% in the 10 mg group. Median overall survival (OS) was not statistically significantly different between the groups, and ranged from 35.5 to 44.5 months.

Significantly higher proportions of treatment-emergent adverse events (AEs) were reported among patients treated with lenalidomide compared with placebo-treated patients. At least one drug-related AE was reported in 42% of the placebo group, 87% in the lenalidomide 5 mg group and 90% in the lenalidomide 10 mg group. The most frequent drug-related AEs were neutropenia (15% in the placebo group and 74% in each lenalidomide group) and thrombocytopenia (2% in the placebo group and 32 and 36% in the 5 mg and 10 mg groups, respectively).

Health-related quality of life (HRQoL) outcomes were assessed during the MDS-004 trial using the Functional Assessment of Cancer Therapy-Anemia (FACT-An) questionnaire, which was administered at baseline, and weeks 12, 24, 36 and 48. The EQ-5D was administered at baseline only. Compared with placebo, treatment with lenalidomide was associated with improvements in HRQoL (FACT-An scores) during the initial 12 weeks of therapy. Improved HRQoL was maintained in patients who remained on double-blind treatment with lenalidomide. Among patients who switched from placebo to the lenalidomide 5 mg group, improved HRQoL was observed.

Critique of the clinical evidence and interpretation

According to the ERG, some of the literature searches of the manufacturer were unnecessarily restrictive. For AEs, other study designs could have been included and longer-term data could have been sought. It was also not clear how studies were identified for inclusion. Nevertheless, the ERG was unaware of any relevant trials that had been missed, and agreed with the manufacturer that the MDS-004 trial was most likely the best source of clinical evidence for the effectiveness of lenalidomide versus BSC.

Data extraction from the MDS-004 trial was reported for the intention-to-treat (ITT) population (N=205) as well as for the modified ITT (mITT) population (N=139). The primary reason for exclusion from the mITT was an inadequate bone marrow sample, preventing confirmation of the diagnosis of low- or intermediate-1 risk MDS del5q. The manufacturer considered the ITT to match more closely with the daily practice population as defined in the NICE scope; therefore the rates from this population were used in the health economic model. Response rates for lenalidomide were based on the lenalidomide 10 mg group (60.9%) while the response rate in the model was 7.5% for the placebo group. Nevertheless, the ITT population included patients not fulfilling the inclusion criteria, i.e. del5q mutation and bone marrow morphology.

Although serious infections were explicitly mentioned as a relevant outcome in the NICE scope, reporting on this outcome in the MS was minimal. Additional data were obtained from the clinical study report and showed that serious infections occurred in the lenalidomide groups twice as often as in the placebo group.

Due to the crossover design after 16 weeks and dose reductions of the trial, the chances of detecting attributable prolonged survival or acceleration of leukaemia progression were limited.

Cost-effectiveness

The manufacturer submitted a *de novo* economic evaluation on the cost-effectiveness of lenalidomide compared with BSC in low- or intermediate-1 risk MDS del5q patients who are transfusion-dependent. An Excel-based Markov model was developed with 14 health states that reflected transfusion requirements, iron chelation, progression

to AML and complications associated with both transfusion dependency and iron chelation therapy. Patients responding to treatment became transfusion-independent while non-responders remained transfusion-dependent. As a simplifying assumption, all trial patients who responded, regardless of timing, were classed as responders from cycle 1 onwards. Transition probabilities for OS and progression to AML were assumed to be different for transfusion-dependent and -independent patients, and estimated based on the initial response for lenalidomide and BSC of the MDS-004 trial. Response rates for iron chelation therapy and iron overload complication rates were based on the literature[38,39].

Lenalidomide treatment (plus BSC) is compared with BSC, which was also the comparator in the MDS-004 trial. However, BSC in the trial consisted of blood transfusions only (plus chelation therapy when iron overload occurred) whereas BSC in the UK may also include the provision of erythropoietin stimulating agents (ESA) or ESA plus granulocyte-colony stimulating factors (G-CSF). The proportion of patients receiving ESA in the model was calculated from the proportion of UK patients in the MDS-004 trial who received ESA prior to the trial, i.e. 28%. Of the side effects associated with lenalidomide, only neutropenia and thrombocytopenia were included in the model since only these were considered serious enough by the manufacturer to warrant inclusion while also being different between the placebo and lenalidomide arms in the trial. Iron chelation is initiated to avoid complications associated with iron overload for transfusion-dependent patients. In the *de novo* model, patients received either desferrioxamine or deferasirox as iron chelation therapy. Since the number of patients included in the trial was insufficient to obtain transition probabilities for AML mortality, transition probabilities were obtained from the adverse risk group in the article by Wahlin et al.[40]. No half-cycle correction was applied. The model had a National Health Service (NHS) perspective and time horizon of 20 years. Costs and effects were discounted at an annual rate of 3.5%.

During the MDS-004 trial, quality of life was measured using the FACT-An at baseline and in weeks 12, 24, 36 and 48. However, the EQ-5D was measured at baseline only (i.e. when all patients were still transfusion-dependent) and therefore, EQ-5D data for transfusion-independent patients were not available. Therefore, utility values were obtained from the study by Szende et al.[41].

Drug acquisition prices were obtained from the British National Formulary (6 March 2013), while the frequency of monitoring associated with the initiation of lenalidomide treatment was based on the summary of product characteristics. Monitoring visits were assumed to occur with a general practitioner (GP). Costs for lenalidomide were based on the dosing observed in the MDS-004 trial and manufacturer's price quotations. An arbitrary standard error of 10% of the mean was assigned to those cost estimates without a standard error.

Table 2.1 Revised base-case cost-effectiveness analysis, incorporating corrections and amendments identified by the ERG and AC

	Best supportive care		Lenalidomide		Incremental		ICER
	Cost	QALY	Cost	QALY	Cost	QALY	Cost per QALY gained
Manufacturer's base case analysis	**£105,726**	**2.58**	**£156,308**	**3.46**	**£50,582**	**0.89**	**£56,965**
Corrected confirmed programming errors	£104,753	2.59	£156,308	3.46	£51,555	0.87	£59,196
Correcting programming errors dose reduction	£104,753	2.59	£162,628	3.46	£57,875	0.87	£66,453
Additional cycle added	£104,753	2.59	£162,628	3.46	£57,875	0.87	£66,453
Half cycle correction	£104,052	2.57	£160,343	3.43	£56,292	0.87	£64,929
Chelation therapy deferiperone added	£102,270	2.64	£158,890	3.49	£56,620	0.85	£66,346
Cost AML adjusted	£100,655	2.64	£157,227	3.49	£56,572	0.85	£66,289
Response over time (mortality based on max response)	£102,839	2.64	£153,817	3.45	£50,978	0.81	£62,773
Cost AEs adjusted	£102,836	2.64	£153,733	3.45	£50,898	0.81	£62,674
ERG revised base case	**£102,836**	**2.64**	**£153,733**	**3.45**	**£50,898**	**0.81**	**£62,674**
Scenario: *Monitoring would be undertaken by a haematologist and progression to AML similar for lenalidomide and BSC*	*£123,241*	*2.95*	*£172,307*	*3.67*	*£49,065*	*0.72*	*£68,125*

ERG Evidence Review Group, AC Appraisal Committee, ICER incremental cost-effectiveness ratio, QALY quality-adjusted life-year, AML acute myeloid leukemia, AEs adverse events, BSC best supportive care

The base case ICER (cost per quality-adjusted life-year (QALY) gained) was £56,965 per QALY gained (Table 2.1). The probabilistic sensitivity analysis (PSA) showed a 0% probability of the ICER being below £30,000 per QALY gained. Sensitivity analysis revealed that utility values, the proportion of patients experiencing dose interruptions, and the curve fitting for progression to AML and overall mortality were key parameters. Table 2.1 shows the revised base case cost-effectiveness analysis, incorporating corrections and amendments identified by the ERG and AC.

Critique of the cost-effectiveness evidence and interpretation

The economic model described in the MS was considered, by the ERG, to meet the NICE reference case to a reasonable extent and was in line with the decision problem specified in the scope. However, the manufacturer's description of the model did not match their own presented figure. The illogical (e.g. from health-state chelation failure to no chelation cardiac disease) and missing transitions (e.g. from health-state transfusion-dependent chelation to transfusion-independent) were corrected by the ERG.

The ERG challenged some of the assumptions of the manufacturer and therefore made the following adjustments in the ERG base case.

- A half-cycle correction was implemented since the first few cycles showed a very significant redistribution of patients over the various health states.
- Deferiprone, a third option for chelation therapy, was included in the model (Table 2.1). This slightly changed the costs per cycle for chelation therapy (from £1,383 to £1,332) but also influenced the QALYs since adding this third option increases the proportion of patients receiving oral instead of intravenous chelation therapy (from 71 to 94.3%).
- Standard errors without a standard deviation estimate were increased from 10 to 20% of the mean for adverse events and complications. The standard errors of 10% were considered too small by the ERG since more variation for costs is usually observed.
- A programming error for the initial response rate for BSC was corrected.
- The effect of G-CSF, in addition to ESA, for non-responders to BSC was added. The initial response rate was used in the model of the manufacturer. G-CSF is only added to ESA for patients who initially do not respond to ESA. As a consequence, the model of the manufacturer did not include the effect of G-CSF for non-responders to BSC.

The revised base case ICER was £62,674 per QALY gained. The PSA results showed a 0% probability that the ICER was below £30,000 per QALY gained.

Remaining concerns

Utilities were obtained from a study that included broad health state descriptions covering a range of health problems[42]. The manufacturer assumed that these descriptions adequately described the difference between the transfusion-independent and -dependent health states. However, this was challenged by the ERG. In addition, the ERG raised questions on the assumption of similar utility values for transfusion-dependent and AML health states. The latter was accepted by the ERG since the impact of the utility value assigned to AML was minimal. Moreover, a reasonable alternative for health state utility values was unavailable.

While the ERG was not entirely convinced that the definition of BSC fully reflects BSC within the NHS, the model outcomes were not very sensitive to changes in the proportion of ESA use.

Additional scenarios were explored by the ERG and these revealed that utility and cost parameters related to AML, complications and AEs have little to no effect on the ICER. The assumption that monitoring occurred with a GP was challenged by the ERG and therefore adjusted to monitoring by a haematologist in an additional scenario analysis together with revised progression rates to AML (Table 2.1).

Conclusion of the Evidence Review Group (ERG) report (before implementation of the Patient Access Scheme [PAS])

According to the ERG there were two main problems with the clinical effectiveness data obtained from the MDS-004 trial and described by the manufacturer in the MS. First, the possibility of crossover after 16 weeks meant that most long-term effectiveness data were unreliable. Second, data were reported for two populations -the ITT and mITT- and it is not clear how differences between these populations influenced results.

The manufacturer base case ICER was £56,965 per QALY gained, while the ERG base case, correcting for the various issues identified, estimated an ICER of £62,674 per QALY gained.

ERG research recommendations

The ERG concluded that further comparisons of lenalidomide and BSC are required in terms of long-term effectiveness, OS, AML progression and incidence of adverse events. The study on which utilities for the transfusion-related health states was based did not conform to the NICE reference case. In order to increase the robustness of the health economic outcome, a quality-of-life study among MDS patients would be of great value. Ideally, such a study would ask transfusion-dependent patients, as well as patients who have become transfusion-independent, to fill out the EQ-5D, after which outcomes would be valued using the UK tariff, which is based on the general population[43].

KEY METHODOLOGICAL ISSUES

Long-term effectiveness, including survival and progression to AML, was compromised by the crossover design of the trial at 16 weeks. The manufacturer stated that survival of patients with MDS is strongly related to transfusion dependency. In order to perform a life-time cost-effectiveness analysis, the model linked OS and progression to AML to transfusion dependency. Therefore, separate time-to-event curves for people who were transfusion-dependent or -independent at 8 weeks were estimated from the data of the MDS-004 trial.

The relationship between survival and transfusion dependency was supported both by data from the MDS-004 trial (achieving transfusion independence was associated with a significant reduction in the risk of death (hazard ratio 0.53; 95% CI 0.31-0.90; p=0.019]) and the literature[44,45].

Utility values used in the model were not obtained according to NICE guidelines and, consequently, the committee needed to decide whether these were acceptable. The STA described here highlights the difficulties of relying on a single RCT with a cross-over design after 16 weeks. The key issue for a decision maker is whether or not these clinical and economic uncertainties cast sufficient doubt on any patient gain from taking the drug.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE GUIDANCE

Key issues considered by the appraisal committee

Regarding effectiveness, the committee concluded that, on the basis of evidence on transfusion independence and HRQoL, lenalidomide is a clinically effective treatment for people with MDS associated with a del5q cytogenetic abnormality although uncertainty about whether lenalidomide improved survival remained. The committee concluded that the serious adverse events associated with lenalidomide could be partly managed by a reduction in dose.

According to the committee, cost-effectiveness estimates from the model were uncertain due to uncertainty in the survival estimates. In addition, there was uncertainty as to whether lenalidomide changed the rate of progression to AML. The committee accepted the utility values reported by Szende et al.[41] after consulting the patient expert. In addition, the committee agreed with the adjustments of the ERG and considered that if the model applied similar rates of progression to AML for both treatment groups then the ICER would most plausibly exceed £70,000 per QALY gained.

Preliminary guidance (first appraisal consultation document [ACD])

After considering the initial evidence submitted by the manufacturer, the ERG report and the testimony of experts and other stakeholders, the AC concluded that lenalidomide could not be recommended for treating transfusion-dependent anaemia caused by low- or intermediate-1 risk MDS associated with a del5q cytogenetic abnormality when other treatments fail.

Response to preliminary guidance (first ACD) and additional analysis submitted by the manufacturer

The manufacturer performed a systematic literature review (July 2013) to better highlight the association between transfusion independence and survival. The AZA-001 trial[46], which demonstrated improved survival after becoming transfusion-independent, was considered most convincing by the ERG. Overall, the committee concluded that while the strength of the relationship over time was uncertain, it was reasonable for the model to include a benefit in OS for patients treated with lenalidomide compared with BSC. Based on the additional submitted evidence[47], the AC also concluded that progression to AML curves should be similar for both lenalidomide and BSC.

Final guidance October 2013

Despite the additional analysis of the manufacturer, the committee did not change the guidance of the first ACD. In the FAD, they concluded that lenalidomide could not be recommended for treating transfusion-dependent anaemia caused by low- or intermediate-1 risk MDS associated with del5q cytogenetic abnormality when other therapeutic options were insufficient or inadequate. Given the uncertainties, the committee concluded that the most plausible ICER was above £70,000 per QALY.

PATIENT ACCESS SCHEME

The proposed PAS

The October 2013 FAD for lenalidomide was withdrawn after the submission of an approved PAS by the manufacturer. Under the PAS, the manufacturer would provide lenalidomide at no cost to the NHS for patients with transfusion-dependent low- or intermediate-1 risk MDS with isolated del5q abnormality who continued with lenalidomide treatment beyond 26 cycles. This PAS is similar to the existing PAS of lenalidomide for patients with multiple myeloma. The PAS therefore reduces the long-term drug costs for patients who receive more than 26 cycles of lenalidomide. A revised version of the model was submitted and reviewed by the ERG. The deterministic ICER with the PAS was £25,544. Minor adjustments to the sensitivity analysis were made by the ERG as

these were also made earlier in the ERG defined base case. At a threshold of £20,000 per QALY, 26.2% of simulations were cost-effective and at a threshold of £30,000, 66.6% of simulations were cost-effective. The ERG reviewed the proposed PAS and economic model. No additional issues apart from those stated earlier were identified.

Preliminary guidance (second ACD)

The main concerns raised by the AC were uncertainties with regard to the ICER. These uncertainties included patient survival, the proportion of patients eligible for the PAS, and the timing of the PAS rebate. These did not only influence the point estimate but also the cloud of possible outcomes around the ICER. For patient survival, the committee concluded that, despite uncertainty regarding the strength of the relationship, it was reasonable to assume a relationship between transfusion independence and OS. Therefore, it was plausible that lenalidomide indirectly improved survival by reducing transfusion dependence. The committee stated that treatment interruptions were not accounted for in the PAS and that the proportion of people surviving beyond 26 cycles in clinical practice was uncertain. Due to the nature of the PAS, cost reductions were obtained from patients receiving treatment after 26 cycles. If this proportion is uncertain in daily practice, the potential cost savings from the PAS are also subject to uncertainty. As a consequence, the ICER could be much higher. As a response to the concerns related to the PAS, the manufacturer included treatment interruptions, leading to a longer period of time before the PAS comes into effect (26 cycles plus 16 days updated the ICER to £25,300). Additional evidence was provided by the manufacturer based on the MDS-004 trial and real-world data that supported the proportion of patients on active treatment currently used in the model (31.9%). They also conducted an additional analysis on the proportion of patients eligible for the PAS, i.e. the proportion of patients on active treatment after 26 cycles. This showed that when 27% or more patients reach 26 cycles of treatment, lenalidomide remains cost-effective at a threshold of £30,000 per QALY.

Final guidance

After considering the available evidence from the manufacturer, the ERG, expert testimony, and other consultees, the NICE AC decided to recommend lenalidomide for treating low- or intermediate-1 risk MDS associated with an isolated del5q cytogenetic abnormality when other therapeutic options were insufficient or inadequate. The committee agreed that the ICER was uncertain but accepted that a commitment from the manufacturer to publish data on the proportion of patients receiving treatment beyond 26 cycles provided reassurance that lenalidomide was a cost-effective use of NHS resources.

CONCLUSIONS

The STA presented here describes the first treatment alternative for MDS del5q patients. Clinical evidence was obtained from a single randomised phase III trial with a crossover design after 16 weeks. The AC decided to accept lenalidomide as treatment for low- and intermediate-1 risk MDS del5q patients although the crossover design of the trial as well as the PAS increased the uncertainty of the ICER. A commitment from the manufacturer to collect data provided reassurance that the uncertainties surrounding the ICER can be reassessed when the guidance is reviewed. Nevertheless, the AC stated that if lenalidomide was not a cost-effective use of NHS resources, the foregone health benefits to other NHS patients until the review cannot be regained. This appraisal illustrated that the AC can accept a treatment as cost-effective under the acceptance of a commitment of the manufacturer to collect and publish data. This case study saw a PAS accepted by the Ministry of Health after the initial FAD. While lenalidomide for treating MDS patients with del5q cytogenetic abnormalities was initially not recommended, the PAS reduced the ICER substantially from approximately £68,100 to £25,300 per QALY. This changed the recommendation from the AC. The generalisability of the cost-effectiveness results to other countries depends on whether such a PAS is also introduced in these countries as well as the transferability of underlying utility and survival estimates.

ACKNOWLEDGEMENT

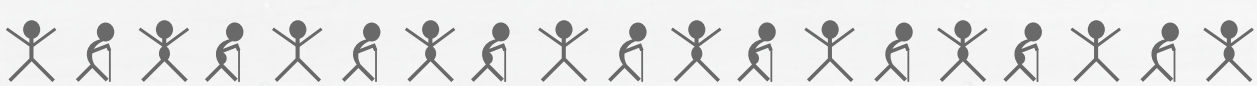
This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Program (project number 11/07/010 (see the HTA program website for further information-http://www.hta.ac.uk). This summary of the ERG report was compiled after AC's review. The views and opinions expressed are the authors' and do not necessarily reflect those of the HTA Programme, NICE, NIHR, NHS, or the Department of Health. The authors, Hedwig Blommestein, Nigel Armstrong, Steve Ryder, Sohan Deshpande, Gill Worthy, Caro Noake, Rob Riemsma, Jos Kleijnen, Johan Severens and Maiwenn Al have no conflicts of interest that are directly relevant to the content of this review.

Chapter 3

Access to expensive cancer drugs in Dutch daily practice: should we be concerned?

Blommestein HM, Franken MG, Verelst SGR, van Agthoven M, Huijgens PC, Uyl-de Groot CA

Netherlands Journal of medicine (2014) 72(4):235-241



ABSTRACT

Background: To investigate whether equal access to bortezomib has been achieved under the Dutch policy regulations that guarantee equal access to expensive inpatient drugs.

Methods: We investigated accessibility to bortezomib treatment at national and regional levels by i) conducting interviews with stakeholders in the Dutch healthcare system to explore prescription barriers and ii) tabulating sales data from 2004-2009 and trial participation rates.

Results: Interviews revealed awareness of the high treatment costs, although prescription barriers were not encountered. National use of bortezomib increased slowly (treating 2% of patients in 2004 to 17% in 2009), indicating a long adjustment period. Furthermore, use remains below the rate estimated by the professional association of haematologists (27%). Regional differences were found for both daily practice use (e.g. ranging from 13-27% in 2009) and clinical trial participation (e.g. ranging from 1-12% in 2006).

Conclusion: Our results were somewhat conflicting: interviews did not reveal any prescription barriers, but quantitative methods showed regional differences, signs of underutilisation, and access inequality. Investigating use and accessibility, based on data triangulation, provides valuable feedback which can enhance evidence-based decision making for both physicians and policymakers. This could improve appropriate and efficient use and ensure equal access to expensive drugs.

INTRODUCTION

Increasing healthcare expenditures may result in limited and unequal access, particularly with regard to new and innovative cancer drugs with high acquisition costs. Policymakers have to make reimbursement decisions considering both rapid and equal accessibility to promising drugs as well as the scarcity of resources. Usually, guaranteeing rapid access means making decisions while available evidence on clinical- and cost-effectiveness is limited[48]. One way of dealing with the need for rapid access and limited evidence is the 'coverage with evidence development' policy; reimbursement under the condition that additional research will be conducted[48,49].

Such policies have been implemented in several countries for surgical procedures, medical devices and pharmaceuticals[49]. Over the last decade, a coverage with evidence development policy was also initiated in the Netherlands, partly triggered by signs of underutilisation and 'zip code prescribing' of trastuzumab[50]. Early access to expensive inpatient drugs is linked with the obligation to gather data on appropriate drug use and cost-effectiveness in daily practice[51]. Drugs meeting the criteria of added therapeutic value and expected budget impact of at least 2.5 million were temporarily included in the policy of 2006-2012. Four years after inclusion, a reassessment will determine whether or not additional financing should continue to exist. At the time we conducted our study, hospitals received 80% of its acquisition costs if a drug was included[52].

Currently more than 30, mostly cancer, drugs are included in this policy. One of these drugs is bortezomib, used for treating multiple myeloma (MM). MM is the second most common haematological cancer. The five-year prevalence in Western Europe is 31,056 while the annual age-standardised incidence rate is 3.2 per 100,000 (IARC GLOBOCAN 2008). Bortezomib obtained European Medicines Agency (EMA) approval in 2004 by demonstrating superior efficacy compared with chemotherapy for the treatment of advanced MM[53-55]; it was included on the Dutch expensive drug list in 2006. Advances in MM treatment in the past decade significantly increased overall survival (44.8 vs 29.9 months[56]), which was largely due to the introduction of autologous stem cell transplantation and new therapeutic agents including thalidomide, lenalidomide, and bortezomib[56,57]. While thalidomide is relatively inexpensive, bortezomib and lenalidomide are expensive drugs. Both are incorporated in professional guidelines[58]. However, the orphan status granted to lenalidomide results in 100% reimbursement for lenalidomide compared with an 80% of reimbursement for bortezomib during our study period. Consequently, accessibility might be an issue, especially for bortezomib. Previous research studied accessibility and use of expensive drugs in the Netherlands[59,60]; however, it remains unclear whether the Dutch policy actually guarantees equal access to expensive inpatient drugs. We investigated whether equal access to bortezomib has been achieved in the Netherlands. We analysed bortezomib use patterns by means

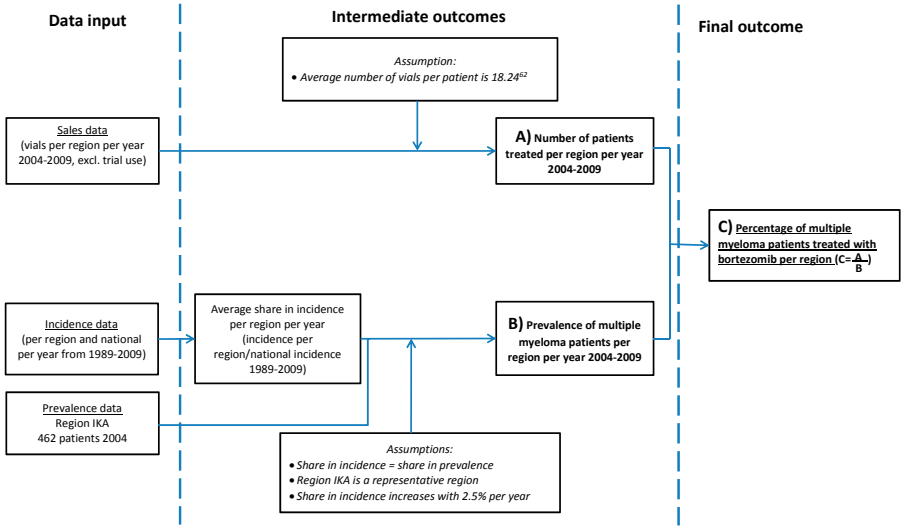
of aggregate sales data and conducted interviews to shed light on perceived or real prescription barriers.

MATERIALS AND METHODS

We took a two-pronged approach. First, seven in-depth interviews were conducted to qualitatively investigate the existence of accessibility issues and prescription barriers. Interviewees were representatives of stakeholders in the Dutch healthcare system: i) a representative of the Dutch Healthcare Authority (NZa), ii) a representative of the Healthcare Inspectorate (IGZ), iii) a hospital director of finance, iv) four haematologists from hospitals varying in size and country location (the North-West, East, South-West, and South). Respondents were selected based on their involvement and knowledge of expensive inpatient drug regulations (NZa and IGZ) or geographical location and type of hospital (haematologists and director of finance). All semi-structured interviews were recorded and analysed according to the steps of Creswell[61], including transcription, coding, interpretation, and description.

Second, we quantitatively investigated the use of bortezomib in daily practice. Because data on bortezomib use at the individual patient level are not available, we combined Dutch sales data (excluding use in clinical trials) from 2004-2009 from the manufacturer, Janssen Pharmaceutical Companies of Johnson & Johnson, with incidence and preva-

Figure 3.1 Flowchart of data input, intermediate and final outcomes



lence data from the Netherlands Cancer Registry[22]. Figure 3.1 provides the flowchart of data used, intermediate and final outcomes and the underlying assumptions. To estimate the number of treated patients ((A) in Figure 3.1), the number of vials sold was divided by the average number of vials used per patient. The average number of vials per patient (18.24) was based on a Dutch observational study of 72 bortezomib patients treated in daily practice from 2004-2008[62].

To investigate bortezomib use across regions, we used the regional division of the nationwide Netherlands Cancer Registry distinguishing eight Comprehensive Cancer Centres[22]. Since these regions differ in size, prescription rates were expressed relative to the number of patients per region. We assumed that equal accessibility to bortezomib would be achieved if the proportion of vials used per region was similar to their proportion of national incidence or prevalence. Regional shares in incidence were calculated over the years 1989-2009. For example, the share in incidence in 2009 for Comprehensive Cancer Centre Amsterdam (IKA) was 18.8%. We calculated this percentage by dividing the incidence of IKA (201) by the national incidence (1069).

Because prevalence numbers were only available for IKA (462 patients in 2004) for one year, we estimated other regional prevalence (B) from their relative shares in incidence. Hereby we assumed i) IKA to be representative for the other regions and ii) the share in incidence per region is equal to the share in prevalence (e.g. if IKA has 19% of the incidence it will also have 19% of the prevalence), and (iii) an annually increasing prevalence of 2.5% (average annual increase over the years 1989-2009[22]) per year because of rises in incidence[57]. Detailed additional information about incidence and prevalence estimates per year is available from the authors upon request.

To obtain a regionally comparable percentage of treated patients (C), we divided the estimated number of treated patients (A) by the estimated prevalence (B). To put regional percentages in perspective, we compared our computed use with the expected percentage of MM patients eligible for bortezomib treatment as estimated by the Dutch professional association of haematologists (the Dutch-Belgian Cooperative Trial Group for Haematology and Oncology (HOVON)). HOVON estimated that about 1600 patients would be eligible for MM treatment per year. Of these patients, one-third would not qualify for treatment with either bortezomib or lenalidomide due to age, the patient's condition or preferences. As result, 1070 patients are eligible for advanced therapy each year.[52] Since patients treated with bortezomib might also be eligible for treatment with lenalidomide and vice versa, HOVON assumed that the number of patients treated with each drug would be similar (50%). To compare the HOVON estimation with the proportion of patients treated with bortezomib per region, we divided the 535 eligible patients (i.e. 1070 divided by 2) by HOVON's estimated prevalence (i.e. 2000 patients), resulting in an estimation of 27% patients. Furthermore, since bortezomib was a novel treatment, clinical trials were conducted during our years of investigation. Because

MM patients are often included in clinical trials, relatively high or low trial participation could distort our computed daily practice use and identified regional differences. Therefore, we selected the two largest clinical studies including bortezomib during our investigated time period and studied trial participation at the regional level. Calculation methods were similar: we divided the number of patients included in trials by regional prevalence to obtain regional trial participation rates for the years 2005-2009. We then combined trial participation with regional daily practice use to compare similarities and differences across regions.

RESULTS

Interview results

Interviewees of the NZa and IGZ did not reveal any accessibility issues for expensive drugs. The IGZ representative, however, admitted that the body had no active role in investigating such issues. Hospitals regulate financial management in various ways. As a result, it may differ per hospital who is responsible for the budget and who is making the financial decisions. According to the interviewed physicians, their financial department divided the total hospital budget by department, whereas physicians organised the division and implementation of the budget within departments. These assumptions were verified and confirmed by the hospital financial manager. Based on these results, we concluded that in the studied hospitals financial management, of both treatment decisions and organisation of care, was the physicians' responsibility.

Generally, all physicians agreed that access to bortezomib is guaranteed in the Netherlands for patients in need. The existence of strict quantitative restrictions was explicitly denied. Physicians adhered to professional guidelines, as far as treatment is concerned, which were frequently mentioned as important. Consultation with colleagues and patient characteristics also seemed to be important factors in the decision (how) to treat. Apart from some variation immediately after the introduction of bortezomib, respondents believed that all eligible patients had equal access.

The Dutch policy of 2006-2012 aimed to facilitate prescription and guarantee access while maintaining incentive for efficiency. According to haematologists, the effects of this policy were two-sided. An additional budget of 80% facilitated prescription but the remaining 20%, financed from the general hospital budget, could hinder prescription. The policy was therefore perceived as ambiguous: while the government relieved the high financial burden, the remainder still had to be financed from the general hospital budget. The situation stimulated local initiatives to manage access to expensive drugs, resulting in a local expensive drug committee to judge appropriate use and structures for consultations with more experienced physicians. Although expensive drugs were

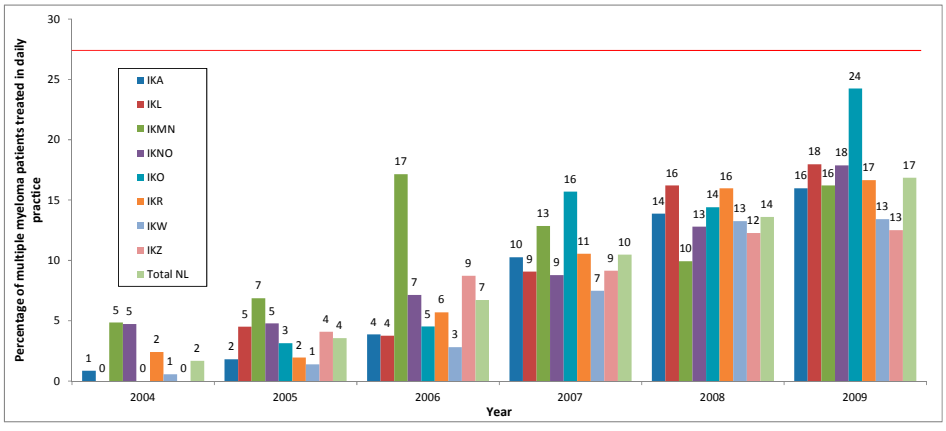
perceived as a high financial burden, according to the respondents, budget played no role in treatment choices.

Data results

Daily practice use.

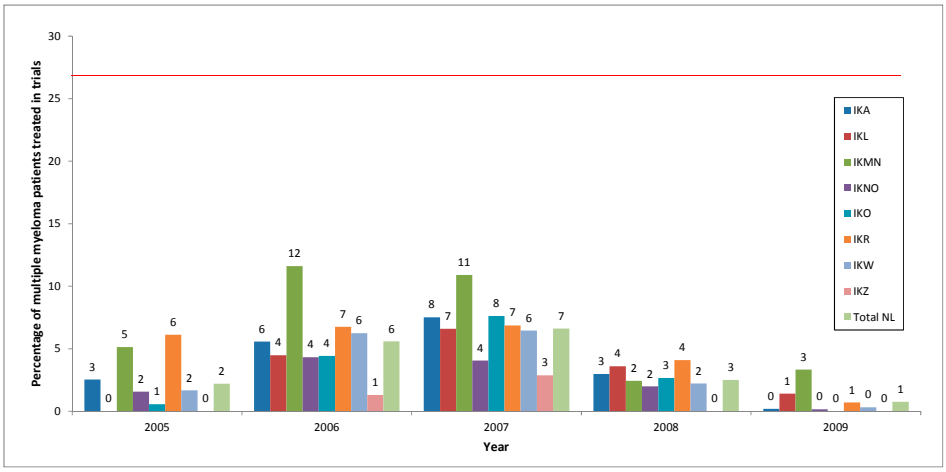
Figure 3.2 presents the percentage of patients treated with bortezomib from 2004-2009 irrespective of treatment line. As mentioned in the method section, HOVON estimated that 27% of MM patients are eligible for bortezomib treatment in daily practice. This is presented as a horizontal line in Figure 3.2. Figure 3.2 reveals relatively low use in 2004 and 2005 for all regions, which was expected since bortezomib was then an innovative treatment and not included on the expensive drug list until 2006. Three regions did not use bortezomib in 2004; all regions used it in 2005. Differences across regions exist in all years with no stable pattern; sometimes regions switched from a high prescription rank in 2005 and 2006 to a low one in 2008. In 2008, two years after inclusion on the expensive drug list, differences between the regions decreased. In 2009, Comprehensive Cancer Centre East (IKO) was the highest prescribing region and Comprehensive Cancer Centre South (IKZ) the lowest, revealing that in one region 24% of patients received

Figure 3.2 Percentage of daily practice multiple myeloma patients treated with bortezomib per region from 2004-2009



IKA: Comprehensive Cancer Centre Amsterdam
IKL: Comprehensive Cancer Centre Limburg
IKMN: Comprehensive Cancer Centre Netherlands Central
IKNO: Comprehensive Cancer Centre North East
IKO: Comprehensive Cancer Centre East
IKR: Comprehensive Cancer Centre Rotterdam
IKW: Comprehensive Cancer Centre West
IKZ: Comprehensive Cancer Centre South

Figure 3.3 Percentage of multiple myeloma patients treated in clinical trials (HOVON 65 and HOVON 86) per region from 2005-2009



IKA: Comprehensive Cancer Centre Amsterdam
IKL: Comprehensive Cancer Centre Limburg
IKMN: Comprehensive Cancer Centre Netherlands Central
IKNO: Comprehensive Cancer Centre North East
IKO: Comprehensive Cancer Centre East
IKR: Comprehensive Cancer Centre Rotterdam
IKW: Comprehensive Cancer Centre West
IKZ: Comprehensive Cancer Centre South

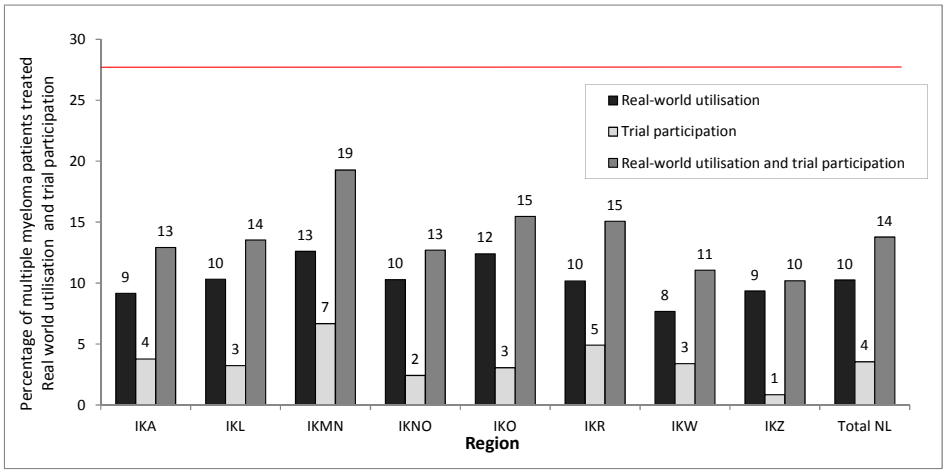
bortezomib while in another only 13% received bortezomib. In all regions the prescription rate was below the 27% of eligible patients as estimated by HOVON.

Use in trials.

Figure 3.3 shows the participation in the HOVON 65[63] and HOVON 86[64] studies per region in the 2005-2009 period. We observed different trial participation rates and, as Figure 3.3 illustrates, trial participation increased from 2005-2007, and decreased in 2008 to almost no participation in 2009. A comparison of Figures 3.2 and 3.3 reveals that the percentage of patients treated in trials is lower than daily practice use of bortezomib.

Finally, Figure 3.4 presents the regional percentages of treated patients aggregated over the years 2005-2009. Comprehensive Cancer Centre Netherlands Central (IKMN) had the highest daily practice use and trial participation (19% were either treated with bortezomib or included in one of the larger trials); IKZ had the lowest (10%). Figure 3.4 also shows that although differences remain, the fluctuation reduced over time. In general, regions with above average daily practice use also had above average trial participation rates.

Figure 3.4 Percentage of multiple myeloma patients treated in daily practice and clinical trials 2005-2009



IKA: Comprehensive Cancer Centre Amsterdam
IKL: Comprehensive Cancer Centre Limburg
IKMN: Comprehensive Cancer Centre Netherlands Central
IKNO: Comprehensive Cancer Centre North East
IKO: Comprehensive Cancer Centre East
IKR: Comprehensive Cancer Centre Rotterdam
IKW: Comprehensive Cancer Centre West
IKZ: Comprehensive Cancer Centre South

DISCUSSION

The aim of our study was to investigate whether bortezomib treatment conformed to policy regulations that were designed to guarantee equal access to expensive inpatient drugs in the Netherlands. Interviews revealed that physicians feel some financial pressure but do not experience prescription barriers and believe that access to expensive cancer drugs is guaranteed. In addition, at that time there were no signs of accessibility issues among IGZ and NZa. Our results, however, also showed that (i) after the introduction of bortezomib, it took one to two years before the drug was prescribed regularly in all regions; (ii) the percentage of patients treated is below the expected 27% of eligible patients; and (iii) there are unexplained regional differences.

In order to investigate accessibility issues and compare regional use levels we had to make several assumptions, especially to calculate the percentage of MM patients treated with bortezomib. While the regions defined by the Netherlands Cancer Registry vary in size, population and available hospital facilities, we expect the baseline patient characteristics to be comparable across regions. Since accurate prevalence numbers were unavailable, we assumed prevalence could be obtained from the distribution of incidence after verifying that the regional distribution of incidence was stable over a long period with a maximum deviation of only 3%. Some uncertainty surrounding total

prevalence, however, remains. Although these assumptions influence the percentage of patients treated, we believe our conclusion of low prescription rates will not be effected. Levels of use would only be closer to HOVON's expected use of 27% if the prevalence of multiple myeloma was much lower (i.e. less than 1700 patients). Considering incidence is 1100 patients per year, prevalence of less than 1700 seems highly unlikely.

Nevertheless, the share in incidence per region was remarkably stable, confirming a stable division between the regions over time. If prescription rates per region were similar, we expected the regions to be accountable for a similar share in bortezomib as their share in incidence. Therefore, regional variation was definitely established, although violations of our assumptions could enlarge or reduce the differences.

Observed regional variation, in both daily practice and trial use, indicates either differences in prescription behaviour or referral of patients to, for example, more experienced hospitals. Because we used sales data aggregated per hospital, we cannot distinguish between patients living in the region and patients referred to the region. Both causes – prescription behaviour and patient referral – limit accessibility. IKZ may have been especially sensitive to regional border crossing because it is the only region without an academic hospital. In this region, use and trial participation is low while relatively high numbers are observed in its neighbouring region (i.e. IKMN). Bortezomib administration, however, does not require specialised skills or hospital facilities, implying that expertise may have been a valid reason for referral immediately after the introduction in 2004, but should be of minor importance in subsequent years.

We studied treatment patterns at an aggregated level, hence neglected other treatment options such as thalidomide and lenalidomide. Because thalidomide is relatively inexpensive in the Netherlands, accessibility should not be an issue. Lenalidomide was accepted for reimbursement at the end of 2007 in Dutch daily practice, creating a competitive alternative treatment option for the years 2008 and 2009 in our analyses. However, lenalidomide does not compensate the low levels of bortezomib prescription. In 2007, 75 patients were treated with lenalidomide and this number increased to 452 and 671 in 2008 and 2009, respectively[52,65].

Regional differences and under-provision have been previously reported in the Netherlands. Large regional differences and under-provision of trastuzumab in the Netherlands were, according to the Dutch Breast Cancer Association[50], mainly due to cost. After the accessibility issues of trastuzumab, the Dutch policy for expensive drugs was revised in 2006. Although bortezomib has been on the market since 2004, it was not until it was admitted to the expensive drug list in 2006 that its use in daily practice doubled compared with the previous year. The increase might indicate that the implemented policy facilitated prescription. Other developments occurred simultaneously, however, including changes in professional guidelines that recommended bortezomib in earlier treatment phases. The relatively low use in the first years might have been

caused by a long adjustment period of physicians who needed to be familiarised with a new drug[65,66]. Bortezomib was, apart from the re-introduction of thalidomide, the first new innovative treatment option for multiple myeloma patients in four decades. It is important that physicians and policymakers are aware of such lags in the regular use of a new innovative and effective drug. Their implementation should receive more attention to accelerate diffusion by, for example, providing feedback about daily practice use. Groot et al.[59] showed that the use of bortezomib in 2005 was almost three times higher in Sweden and France compared with the Netherlands. Furthermore, Dutch use in 2007 was a little less than 35 mg per 100,000 inhabitants while the European average (Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Spain, Sweden, Switzerland and the UK) was above 50 mg per 100,000 inhabitants[60]. Our results also showed that use was below HOVON's expected rate. Despite financial assistance, use and accessibility issues might thus still exist.

It remains subject to further research whether observed regional differences are due to physician prescription behaviour or referral to more experienced or wealthier hospitals. Differences seem to have decreased compared with previous outcomes of the trastuzumab study in 2005, which might be a result of the changes in the policy regulations. However, we should note that the trastuzumab study analysed patients with breast cancer, whose prevalence is much higher than multiple myeloma. Wagelaar et al.[67] studied accessibility of two expensive drugs in the Netherlands, bortezomib and trastuzumab, mainly by investigating whether prescription was in accordance with guidelines at the individual patient level. Medical files were examined and interviews were conducted with physicians, members of hospital boards of directors, and patients. They concluded that guidelines were strictly followed and that recommendations by the professional association and patient characteristics determined treatment decisions. Although the budget of 80% was insufficient according to their respondents, accessibility was not an issue. Interestingly, while their results align with our interview results, they are in contrast with our quantitative findings and our research shows that differences in accessibility might not be revealed by using a qualitative research method only.

In 2012, changes in the regulations increased the additional earmarked budget to full coverage of the 'add-on' diagnoses-related group (i.e. 100% reimbursement of expensive drugs but hospitals and insurers negotiate on the price of the 'add-on'). Although hospital resources remain scarce, this might improve access and reduce remaining regional differences. It will be interesting to closely follow the consequences of this new policy.

We investigated equality in access to bortezomib in the context of Dutch policy regulations for expensive drugs. Use of bortezomib has increased over time although regional differences are still present. We obtained different conclusions using two methods. While interviews did not reveal absolute prescription barriers, regional differences and

possibly underutilisation were observed by comparing sales data with incidence and prevalence data. It seems that appropriate drug use and thus also accessibility depends on various factors, regulatory and organisational characteristics of a healthcare system being two important ones. An evaluation of health policies should therefore be based on mixed methods and data triangulation. Such an evaluation provides insight and valuable feedback that can enhance evidence-based decision making for both healthcare providers and policymakers. This could improve appropriate drug use and ensure equal access to healthcare. In the end, efficient and equitable use of scarce resources increases society's benefits from a healthcare system.

Chapter 4

Real-world costs of autologous and allogeneic stem cell transplantations for haematological diseases: a multicentre study

Blommestein HM, Verelst SGR, Huijgens PC, Blijlevens NMA, Cornelissen JJ, Uyl-de Groot CA

Annals of Hematology (2012) 91:1945-1952

ABSTRACT

Haematopoietic stem cell transplantation (SCT) is an expensive lifesaving procedure, which is increasingly performed in patients with haematological diseases. Developments in the protocol for SCT have resulted in cost estimates that require updating.

We aimed to calculate actual costs for SCT and to identify major cost drivers by means of a daily practice cost study. We randomly selected 191 patients, treated at three university hospitals, who underwent an autologous (auto) SCT or allogeneic (allo) SCT in 2007, 2008 or 2009. Allo-SCT included sibling (sib) donors, matched unrelated donors (MUD) and umbilical cord blood (UCB). Resource use was collected from the hospital registration systems and medical files. The total costs included selection and harvesting of stem cells, transplantation and 1-year follow-up.

The average costs per patient were €45,670 for auto-SCT and €101,919 for sibling allo-SCT. The costs of transplantations from unrelated donors were much higher: €171,478 for allo-SCT-MUD and €254,689 for allo-SCT-UCB. Hospital inpatient days together with laboratory and other activities were the main cost drivers across all types of SCT. Besides, donor search costs were a large cost component in allo-SCT-sib (18 %) and allo-SCT-MUD (12%). Real-world costs were above routine reimbursement and appropriate financing is necessary to guarantee the continuation of SCT.

The costs calculated in this study provide reliable up-to-date input for cost-effectiveness studies and budget revision.

INTRODUCTION

Autologous (auto) and allogeneic (allo) haematopoietic stem cell transplantations (SCT) are effective, often lifesaving, treatments for patients with haematological diseases, which are increasingly applied[68,69]. Auto-SCT includes the reinfusion of earlier-harvested patient-derived stem cells to restore bone marrow function after intensive chemotherapy with or without radiotherapy. The complication rate is low and treatment-related mortality is generally below 5%. In contrast, allo-SCT implies the infusion of stem cells from an allogeneic donor. The stem cells may be obtained from a family donor, including a matched sibling (sib) or haploidentical family donor, a matched unrelated donor (MUD) or from umbilical cord blood (UCB). Donor T cells have a major antitumour effect but may also attack healthy tissues of the patient; graft-versus-host disease (GVHD). GVHD increases the complication rate and treatment-related mortality of allo-SCT, which may increase up to 25–30%, depending on a number of factors[69,70]. Since allo-SCT is increasingly applied and transplants are costly procedures, an increasing demand on health care systems are noted in several countries.

In 2008, more than 800 patients received either an autologous or allogeneic SCT in the Netherlands[71]. As one of the first European countries, the Netherlands introduced managed competition and a diagnosis related group (DRG) financing system. The DRG system detailed hospital care into so-called “care products”, which price is negotiated between health providers and insurers. DRGs include all actions necessary to diagnose and treat the patient in the hospital. Additional reimbursement is only provided for expensive inpatient drugs. Due to the implementation of managed competition, hospitals increasingly face financial responsibility and an urgent need for information about real-world costs is created. Prices of specific care products, such as SCT, are currently considered unsuitable for negotiations and are therefore centrally regulated. Adequate price data were not available, so outdated reimbursement rates were used for SCT. Besides, auto-SCT was only reimbursed for leukaemia patients.

Although the cost of SCT has been studied previously[72-75], up-to-date estimates of average costs are currently unavailable. New indications for SCT have emerged and recent medical developments such as the increasing use of peripheral blood and cord blood grafts affected current transplantation procedures[68]. Furthermore, although advances in the prevention and treatment of graft-versus host disease improve survival, they also increase medical costs. The aim of this multicentre study was to calculate the real-world costs of auto-SCT and allo-SCT for haematological diseases. We aim to identify the major cost drivers and compare real-world costs with reimbursement.

MATERIALS AND METHODS

Study population

The study population consisted of patients with haematological diseases who underwent a SCT in 2007, 2008 or 2009. We randomly selected adult patients from three Dutch university hospitals. As cord blood SCT was performed in only one of the three hospitals, the patients included for this type of SCT were from a single institution. Patient characteristics such as age, sex, and diagnosis were obtained.

Treatment

A complete transplantation consisted of three phases: selection and harvesting of stem cells, transplantation and the ensuing 1-year follow-up. During the first phase, patients were prepared for transplantation. Stem cells were harvested from the patients, or a donor search procedure was initiated. The second phase, in which the transplantation was actually performed, began on the first hospital admission day and continued until discharge. The second phase was followed by a 1-year follow-up for all patients that included monitoring and ongoing support.

Perspective and costs

We took an institutional perspective and incorporated all treatment-related hospital costs. Data were obtained from hospital registration systems that included detailed information about all relevant inpatient treatment activities. By relying on the hospital registration systems, the data gathering corresponded to the principles of the new revised system of activity-based costing. In addition, we collected information from medical files. Data included all relevant treatment activities. These activities were subcategorised into seven cost groups: inpatient days, daycare treatment, outpatient visits, intensive care admission, medication, blood products and laboratory and other activities (e.g. imaging, laboratory tests, radiation and injections). Two additional cost categories were added to the selection and harvesting phase of allo-SCT: donor search costs and human leucocyte antigen (HLA) typing. Donor search costs were based on the total costs of two hospitals divided by the number of allo-SCT-MUD and allo-SCT-UCB. We matched treatment activities to national prices defined by the health authority that included hospital costs and specialist fees[76]. Costs of inpatient days (€712), daycare (€224), outpatient days (€148) and intensive care (€2,100) were obtained from Gaultney et al.[77] who updated the specific prices for haematological departments of Tan et al.[78]. Medication costs for patients at two institutions were available. From these patients, we selected a random sample of 78 patients to calculate medication costs during each of the three phases. Costs were calculated per unit, based on the Dutch Pharmaceutical Advisory Committee (Z-index 2009–2010) and multiplied by the total dosage given.

Analysis

Average and median transplantation costs per patient were calculated for each phase and per cost category. The main cost drivers for each type of SCT were identified by calculating the proportion of total costs of each cost group. In order to reduce the number of categories, costs for blood products and medication were merged. We investigated the correlation between the cost groups and total average SCT costs. The analyses were performed using Microsoft Excel 2010 and STATA SE, version 11.2.

RESULTS

Patients

We included 191 randomly selected patients who received an auto-SCT or allo-SCT in one of three hospitals in 2007–2009. The characteristics of the patients included are presented in Table 4.1. Mean and median age for auto-SCT and allo-SCT-sib were similar. Patients receiving allo-SCTMUD were significantly younger ($p<0.018$) compared to patients receiving auto-SCT or allo-SCT-sib. Table 4.1 shows that multiple myeloma was the most common diagnosis for auto-SCT followed by acute leukaemia and non-Hodgkin lymphoma. Most patients receiving allo-SCT were diagnosed with acute leukaemia. The average number of admission days for auto-SCT, allo-SCT-sib and allo-SCT-MUD patients was similar in the transplantation phase. In contrast, the average number of admission

Table 4.1 Patient characteristics

	Type of SCT			
	Auto[a]	**Allo-sib**[b]	**Allo-MUD**[c]	**Allo-UCB**[d]
	N=68	N=59	N=43	N=21
Patient characteristics				
Average age (years)	51	51	45	47
Median age (years)	53	53	47	52
[range]	[19 - 66]	[19 - 67]	[18 - 66]	[21 - 66]
Sex (male)	57%	36%	67%	48%
Diagnosis				
Aplastic anaemia		3.3%		9.5%
Acute lymphoblastic leukaemia	4.4%	6.7%	19.0%	28.6%
Acute myeloblastic leukaemia	26.5%	50.0%	47.6%	38.1%
Chronic lymphoblastic leukaemia		3.3%	7.1%	14.3%

Table 4.1 Patient characteristics (continued)

	Type of SCT			
	Auto[a]	**Allo-sib**[b]	**Allo-MUD**[c]	**Allo-UCB**[d]
	N=68	N=59	N=43	N=21
Chronic myeloblastic leukaemia		1.7%	9.5%	
Hodgkin lymphoma	4.4%	1.7%		
Multiple myeloma	38.2%	21.7%	4.8%	4.8%
Non-Hodgkin lymphoma	20.6%	5.0%	4.8%	4.8%
Other[e]	4.4%	3.3%	7.1%	
Unknown	1.5%	3.3%		
Admission days for SCT phase				
Average	20.7	19.6	19.1	42.4
Median	20.5	17	15	40
[range]	[6 - 48]	[2 - 62]	[7 - 59]	[20 - 88]

[a] Auto: Autologous stem cell transplantation
[b] Allo-sib Allogeneic sibling stem cell transplantation
[c] Allo-MUD: Allogeneic matched unrelated donor stem cell transplantation
[d] Allo-UCB: Allogeneic umbilical cord blood stem cell transplantation
[e] Other diseases including Mantle cell lymphoma, Ewings sarcoma, Autoimmune diseases, Polycythaemia vera, Prolymphocytic leukaemia, Myelodysplastic syndromes, Primary myelofibrosis

days for allo-SCT-UCB patients was significantly higher ($p<0.01$). From Table 4.1, it can be concluded that the range of admission days was large.

Average and median cost

Table 4.2 presents the average and median cost for autologous and allogeneic SCT. The average cost was €45,668 for auto-SCT. The transplantation phase was the most expensive phase (€21,124). Total median cost (€34,688) was below the average cost. The average total cost for allo-SCT-sib was €101,923. In contrast with auto-SCT, the transplantation phase was the least expensive phase (€24,894). The total average costs of SCT from an unrelated donor were much higher, €171,482 for allo-SCT-MUD compared to allo-SCT-sib and on account of the increased costs during selection and harvesting and follow-up. The total average and median costs for allo-SCT-UCB were €254,690 and €167,289, respectively. Compared to allo-SCT-MUD, costs during the transplantation and follow-up were much higher.

Cost categories

Table 4.3 presents the average costs by category during the three phases for each type of SCT. Inpatient days were a large cost category in all phases and for each type of SCT.

Table 4.2 Average and median costs of autologous and allogeneic stem cell transplantations

Type of transplantation	Phase	Average costs	Median costs	Range [Min - Max]
	Selection/harvesting	€ 11,935	€ 6,732	[€ 2,944 - € 13,539]
Auto-SCT[a]	Transplantation	€ 21,124	€ 19,730	[€ 10,963 - € 37,472]
	Follow-up (1-year)	€ 12,609	€ 6,174	[€ 2,851 - € 80,315]
	Total	**€ 45,668**	**€ 34,688**	
	Selection/harvesting	€ 31,480	€ 29,481	[€ 14,140 - € 61,950]
Allo-SCT-sib[b]	Transplantation	€ 24,894	€ 18,739	[€ 14,384 - € 54,393]
	Follow-up (1-year)	€ 45,549	€ 32,169	[€ 15,435 - € 175,257]
	Total	**€ 101,923**	**€ 88,738**	
	Selection/harvesting	€ 64,876	€ 48,238	[€ 43,726 - € 113,447]
Allo-SCT-MUD[c]	Transplantation	€ 28,581	€ 20,196	[€ 16,727 - € 80,088]
	Follow-up (1-year)	€ 78,025	€ 47,119	[€ 13,646 - € 454,609]
	Total	**€ 171,482**	**€ 129,984**	
	Selection/harvesting	€ 65,398	€ 56,036	[€ 47,131 - € 104,352]
Allo-SCT-UCB[d]	Transplantation	€ 56,277	€ 47,166	[€ 25,008 - € 209,999]
	Follow-up (1-year)	€ 133,015	€ 57,897	[€ 21,464 - € 526,808]
	Total	**€ 254,690**	**€ 167,289**	

[a] Auto-SCT: Autologous stem cell transplantation
[b] Allo-SCT-sib: Allogeneic stem cell transplantation sibling donor
[c] Allo-SCT-MUD: Allogeneic stem cell transplantation matched unrelated donor
[d] Allo-UCB-SCT: Allogeneic stem cell transplantation umbilical cord blood

Table 4.3 Average costs (€) per patient for three phases per transplantation type

Cost category	Phase			Total
	Selection/ harvesting	Transplantation	Follow-up (1-year)	Cost (€)
Auto-SCT[a]				
Inpatient days	6,408	14,869	2,143	
Daycare	128	0	326	
Outpatient days	479	27	2,422	
Intensive care days	0	0	0	
Medication	1,342	2,386	362	
Bloodproducts	720	1,794	1,230	
Laboratory and other activities	2,858	2,048	6,126	
Total auto-SCT				**45,668**
Allo-SCT-sib[b]				
Inpatient days	10,169	14,858	13,887	
Daycare	436	28	809	
Outpatient days	809	41	5,125	
Intensive care days	378	0	2,387	
HLA typing	9,968			
Medication	1,661	4,171	5,469	
Bloodproducts	2,223	1,105	3,733	
Laboratory and other activities	5,836	4,691	14,139	
Total allo-SCT-sib				**101,923**
Allo-SCT-MUD[c]				
Inpatient days	12,610	13,688	33,370	
Daycare	235	0	567	
Outpatient days	882	42	5,553	
Intensive care days	132	657	5,333	
Donorsearch	30,456			
HLA typing	9,968			
Medication	2,778	8,688	8,640	
Bloodproducts	2,661	2,048	5,542	
Laboratory and other activities	5,154	3,458	19,020	
Total allo-SCT-MUD				**171,482**

Table 4.3 Average costs (€) per patient for three phases per transplantation type (continued)

Cost category	Phase			Cost (€)
	Selection/ harvesting	Transplantation	Follow-up (1-year)	Total
Allo-SCT-UCB[d]				
Inpatient days	10,002	30,480	34,598	
Daycare	395	0	2,300	
Outpatient days	869	21	7,580	
Intensive care days	0	2,400	6,838	
Donorsearch	30,456			
HLA typing	9,968			
Medication	3,946	4,596	13,762	
Bloodproducts	2,362	6,089	10,863	
Laboratory and other activities	7,400	12,691	57,074	
Total allo-SCT-UCB				**254,690**

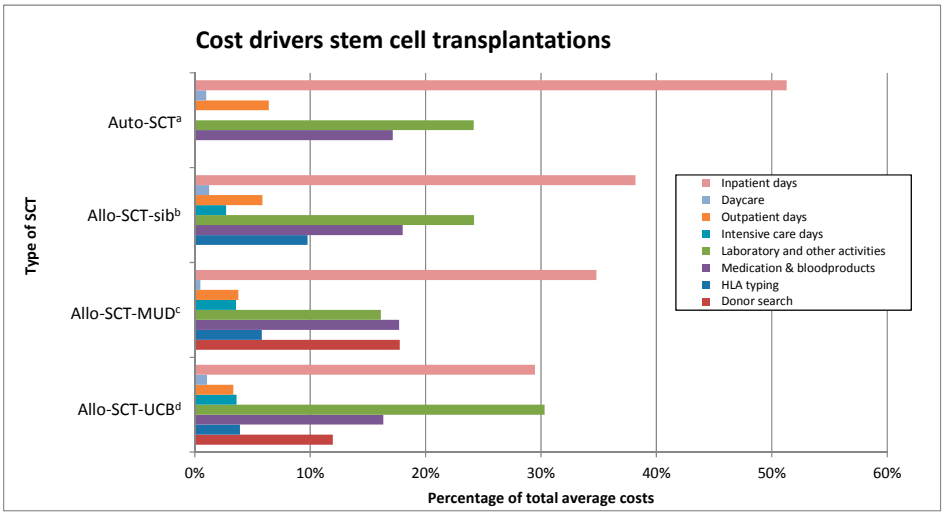
[a] Auto-SCT: Autologous stem cell transplantation
[b] Allo-SCT-sib: Allogeneic stem cell transplantation sibling donor
[c] Allo-SCT-MUD: Allogeneic stem cell transplantation matched unrelated donor
[d] Allo-SCT-UCB: Allogeneic stem cell transplantation umbilical cord blood

For the selection and harvesting phase, there was a large difference between auto-SCT and allo-SCT and between SCT from related or unrelated donors. This difference is to a large extent related to the additional cost categories for allo-SCT, HLA typing (€9,968 and donor search cost (€30,456), which were applied to allo-SCT-MUD and allo-SCT-UCB. During the 1-year follow-up phase, the categories inpatient days and laboratory and other activities were the largest cost categories. We investigated the correlation between cost components, patient characteristics and total cost. Inpatient hospital visits were significantly correlated with total cost ($p<0.001$). None of the patient characteristics had a significant impact on total cost.

Cost drivers

Figure 4.1 shows the percentage of each cost category in the average costs per patient for the three successive phases of auto and allo-SCT. From Figure 4.1, it can be concluded that inpatient days were the main cost driver in auto-SCT, allo-SCT-sib, and allo-SCT-MUD. In allo-SCT-UCB, the cost category laboratory and other activities was a major cost driver and responsible for more than 30% of the costs. However, inpatient days were the second largest cost driver and responsible for nearly 30% of the total

Figure 4.1 Cost drivers for autologous and allogeneic stem cell transplantations



[a] Auto-SCT: Autologous stem cell transplantation
[b] Allo-SCT-sib: Allogeneic stem cell transplantation sibling donor
[c] Allo-SCT-MUD: Allogeneic stem cell transplantation matched unrelated donor
[d] Allo-SCT-UCB: Allogeneic stem cell transplantation umbilical cord blood

costs. While Table 4.3 showed that costs for medication and blood products differed for each transplantation type, Figure 4.1 reveals that their contribution in the total costs was almost identical for each type of SCT, around 18%.

DISCUSSION

This multicentre study calculated the cost per stem cell transplant in daily practice. The average cost of auto-SCT were lower than that of allo-SCT, while costs of allo-SCT increased depending on donor type, especially because of increased donor search costs of €30,456 associated with an unrelated donor as compared to a sibling donor. While the costs of the selection and harvesting phase of allo-SCT-UCB and allo-SCT-MUD were similar, the costs of the second and third phases of allo-SCT-UCB were much higher due to more inpatient days and laboratory and other activities. The admission period during allo-SCT-UCB transplantation compared to allo-SCT-MUD was on average twice as long because of delayed peripheral blood cell recovery. These higher costs of allo-SCT-UCB are predominantly related to the type of transplant, most likely explained by the significant lower number of haematopoietic progenitor cells present in those grafts. Compared to allo-SCT-MUD, regular (cheap) laboratory activities (e.g. haemoglobin, leukocytes,

albumin and protein) as well as expensive laboratory tests (e.g. cytomegalovirus detection with DNA/RNA amplification) were conducted frequently and much more often in allo-SCT-UCB. Allo-SCT-UCB was and is a new innovative treatment performed only in designated hospitals like one of the hospitals included in our study. Treatment strictly followed protocol including extensive laboratory testing. In addition, data on drug use revealed much more anti-viral prescription probably providing another reason why much more laboratory tests were performed during follow-up.

Comparing our results with those of previous studies is difficult due to developments in treatment protocols, different cost analysis methods and country-specific factors which influence the context in which transplantations are performed. However, similarities between previous studies were found. In 2001, a prospective Norwegian cost study[74] calculated the cost of bone marrow allo-SCT to be $106,825 (€95,212 in 2010 after converting and inflation). Their average is comparable to the average costs for allo-SCT from a sibling donor in our study (€101,923). However, their study did not differentiate between sibling and unrelated donor transplantations. According to our results, the difference between these two types of transplantations is substantial. Another multicentre cost study was conducted in Norway among patients with malignant lymphoma and multiple myeloma receiving auto-SCT[75]. Their cost analysis included two transplantation phases that seem similar to the selection and harvesting and transplantation phase of our study. The average cost per patient was $32,160 (€28,664 in 2010 after converting and inflation) and comparable to our results for auto-SCT (€33,059). Majhail et al.[79] presented the median cost of allo-SCT-sib and allo-SCT-UCB for myeloablative and non-myeloablative regimens in USA. While the results of Majhail et al.[79] for allo-SCT-sib are comparable with our results, there is a large difference between the costs for allo-SCT using UCB as a source of stem cells. It should be mentioned that Majhail et al.[79] excluded graft acquisition costs and calculated the median cost for the first 100 days only. The follow-up phase in our study is 365 days and to a large extent responsible for the high average costs associated with UCB as a graft for allo-SCT.

Our results confirm earlier findings of hospital inpatient days as the major cost driver for auto and allo-SCT, and these inpatient days are probably related to post-transplant complications[79-82]. Apart from hospital inpatient days, we also identified laboratory and other activities as a large cost driver. The contribution of total average costs of medication and blood products was around 18% irrespective of the type of SCT. For reimbursement decisions, economic evaluations are frequently conducted as they serve as a means to compare the effects and cost of several treatment strategies. As SCT is a preferred treatment option for many indications within haematology, our real-world cost study provides important input for evaluations of treatment strategies available to haematological disorders. Since we have presented the entire treatment cost of transplantation separately for three phases, our results can be easily adapted to the required input

parameters in future economic evaluations depending on variations in the protocol for the SCT procedure. Furthermore, our estimates can be used to assess whether changes to the protocol can result in cost-savings. For example, strategies to decrease the number of hospital days could reduce overall costs and should be explored. Real-world cost studies are also useful as they can be used to compare the actual treatment costs to the reimbursement rates[83] in use at the moment we conducted our study. For auto-SCT, reimbursement (€44,883) and actual costs (€45,668) were similar. However, hospitals were only reimbursed for auto-SCT if the procedure was conducted among leukaemia patients. The actual cost of allo-sib-SCT was €101,923, which was much higher than the Dutch reimbursement rate of €67,501 Reimbursement rates in use while we conducted our study did not differentiate between MUD and UCB, while according to our results, the difference between both types of SCT was substantial. Both real-world costs of allo-SCT-MUD (€171,482) and allo-SCT-UCB (€254,690) were higher than the reimbursement rate (€145,756). While reimbursement rates should cover all hospital activities, our results revealed a large difference between reimbursement and actual hospital costs. Depending on the type of SCT and the type of patient, the shortage varied between less than €1,000 to more than €100,000 per patient. Similar findings were obtained from Norwegian DRG financing systems[74,75]. Inadequate reimbursement may make transplantations a high financial burden since this type of care is concentrated in a few hospitals. This might impact patient access to care as it could lead to hospitals feeling compelled to refuse treatment or provide treatment at the expense of other patients' treatment. Our study results provided reliable cost calculations for the Dutch government and the reimbursement authorities. After negotiations between haematologists and the Dutch reimbursement agency, reimbursement was adjusted for 2012 based on our results. Future studies are warranted to calculate costs and investigate whether differences between reimbursement and costs are a consistent problem.

Our multicentre study randomly selected patients from three Dutch hospitals. These three hospitals together are responsible for more than 40% of all transplantations in the Netherlands[71]. By including multiple representative hospitals across the Netherlands, we aimed to reduce the hospital variation and establish representative average cost.

Results showed a broad range for costs and admission days. Considering the strong positive relationship between admission days and post-transplant complications, we included patients with and without complications. Besides, the average number of admission days is an important indicator of the total costs. By presenting the average number of admission days per transplantation, we ensure the applicability of our results to other hospitals and settings.

Almost all hospitals in the Netherlands participate in the Dutch-Belgian Cooperative Study Group for Haematological Malignancies, HOVON, a foundation of professionals originally initiated for conducting clinical trials. The group further developed guidelines

for patients outside the context of trials and is the major political spokesman for haematooncology.

Collaboration is often sought with other groups such as the EORTC and national groups in Europe and beyond. Within HOVON, the working group for haematopoietic stem cell transplantations has made consensus statements for stem cell transplantation indications and for all supportive care requirements, all in accordance with international standards of the Joint Accreditation Committee-ISCT Europe (JACIE). Therefore, our results are not only valid for the Netherlands but applicable to other countries which follow the JACIE standards.

We obtained real-world data from the hospital registration systems, medical patient files, and electronic information systems. Our results depend on the completeness of registration, and unregistered hospital activities were not included in our study. This might have caused a slight underestimation of actual cost. On the other hand, while we aimed to select only activities related to the SCT procedure, we cannot guarantee that all activities were fully related to the transplantation process, especially during follow-up. These problems are inherent to real-world cost studies, and since we found no unrelated activities, the effect on our results seemed negligible.

CONCLUSION

In conclusion, auto-SCT and allo-SCT are costly procedures for which hospital inpatient days are the main cost driver across all types of SCT. There was a discrepancy between real-world cost and reimbursement, and based on our data, reimbursement rates changed. Costs calculated in this study provide reliable input for cost-effectiveness studies and can be adapted to the required input parameters.

ACKNOWLEDGMENTS

The authors thank the hospitals that participated and the Dutch Society for Haematology for financing this study. We are grateful to Arjan Bandel, Kafong Cheung, Simone Markus, Koen Meijssen, Geraldine Droog and Ger Boonen for providing the hospital data and Jennifer Gaultney for her proofreading and feedback.

Chapter 5

Cost-effectiveness of rituximab as maintenance treatment for relapsed follicular lymphoma: results of a population-based study

Blommestein HM, Issa DE, Pompen M, Ten Hoor G, Hogendoorn M, Joosten P, Zweegman S, Huijgens PC, Uyl-de Groot CA

European Journal of Haematology (2014):92:398-406

ABSTRACT

Objectives: On the basis of two population-based registries, our study aims to calculate the real-world cost-effectiveness of rituximab maintenance compared with observation in relapsed or refractory follicular lymphoma patients who responded to second-line chemotherapy.

Methods: Data were obtained from the EORTC20981 trial, the Netherlands Cancer Registry and two population-based registries. A Markov model was developed to calculate cost per life year gained (LYG) and quality-adjusted life years (QALYs) for three scenarios.

Results: Our real-world patients were (62 years) 6 to 7 years older and had higher complete response rates to second-line chemotherapy than the trial population. Differences between the real-world rituximab and observation group were observed for second-line chemotherapy and disease progression. Groups were more balanced after using propensity matching. Relying entirely on updated trial results (scenario1) in combination with local cost data resulted in ratios of €11,259 per LYG and €12,655 per QALY. For scenario2, consisting of trial efficacy and matched real-world costs, ratios of €21,202 per LYG and €23,821 per QALY were calculated. Using real-world matched evidence (scenario3) for both effectiveness and costs showed ratios of €10,591 per LYG and €11,245 per QALY.

Conclusion: Although differences in real-world and trial population were found, using real-world data as well as results from long-term trial follow-up showed favourable ICERs for rituximab maintenance. Nevertheless, results showed that caution is required with data synthesis, interpretation and generalisability of results. As different scenarios provide answers to different questions, we recommend healthcare decision-makers to recognise the importance of calculating several cost-effectiveness scenarios.

INTRODUCTION

Follicular lymphoma (FL) is the largest subtype of indolent non-Hodgkin lymphoma. In general, FL is incurable, and for decades, median overall survival (OS) was estimated to be between 8 and 10 years[84]. However, in the past decade, improved median OS has been observed, for example for stage IV, median OS was 12.7 years[85], while median OS was more than 18 years for grade 1 to 2 FL[86]. Rituximab (MabThera®, F.Hoffmann-La Roche Ltd) is a chimeric murine/human anti CD20 monoclonal antibody capable of killing CD20-positive lymphoma cells. In FL, adding rituximab to chemotherapy improves event-free survival[87,88], progression-free survival (PFS) and OS[88-90] in first-line. Moreover, although OS was not significantly different, rituximab maintenance prolongs PFS in first-[87] and second-line[91,92] FL patients. As a consequence, rituximab maintenance is registered in Europe since 2006 and locally reimbursed in the Netherlands for patients with relapsed or refractory FL responding to re-induction therapy. In 2010, this was expanded to patients responding to chemotherapy, irrespective of treatment line.

While the efficacy of rituximab maintenance has been established, policymakers have to decide whether reimbursement of rituximab maintenance is value for money. While randomised controlled trials (RCTs) are the golden standard for establishing efficacy, the data so generated do not necessarily reflect real life[7]. Therefore, healthcare decision-makers are increasingly interested in real-world data. Nevertheless, data are scarce, almost always retrospectively gathered, and experiences of a cohort study demonstrated data synthesis was mostly inevitable to obtain valid cost-effectiveness results[93].

On the basis of two population-based registries, our study aims to calculate the real-world cost-effectiveness of rituximab maintenance compared with observation in relapsed or refractory FL patients who responded to second-line chemotherapy in the Dutch setting. Three scenarios were identified to illustrate the impact of data synthesis and explore how each scenario may fulfil the information need of decision makers. To our knowledge, our study is the first conducting outcomes research for rituximab maintenance on the basis of population-based registries.

MATERIALS AND METHODS

Data sources

Data were obtained from multiple sources: long-term follow-up of the EORTC20981 trial, primary data from the Netherlands Cancer Registry and two population-based registries collecting data on patient characteristics, treatment and healthcare utilisation[20,94]. The Population-based HAematological Registry for Observational Studies (PHAROS)[20,21], located in the north-west and south-west of the Netherlands, aims to cover 40% of the

Dutch population and collects real-world data from medical records. HemoBase[94] is a multidisciplinary Web-based electronic patient record in the north-eastern part of the Netherlands in which data are entered in the database by physicians and laboratory employees.

Patients

We applied the criteria for maintenance treatment and selected indolent FL patients who received at least two lines of chemotherapy and responded to their second-line in 2004 to 2011. In addition, patients receiving rituximab maintenance after a stable disease or unknown response to second-line chemotherapy were included in the description of the real-world patient population. The included patients were divided in two groups: the maintenance and observation group. As real-world patients were not randomly assigned to rituximab or observation, we corrected for observed differences in patient characteristics with the statistical matching technique of propensity scores where each patient received a score that reflects the probability of being in the treatment group given certain pre-treatment characteristics[95]. Propensity scores for rituximab maintenance were estimated using logistic regression with rituximab treatment as the dependent variable and age, years since diagnosis, B-symptoms, FLIPI score, stage and response to re-induction therapy as independent variables. Nearest neighbour matching was used where each person in the treatment group is matched to a patient in the observation group with the closest propensity score to them. The final matched population was used for both effectiveness and cost calculations.

Cost-effectiveness scenarios

Longer follow-up available from the EORTC trial and real-world evidence enabled us to calculate the incremental cost-effectiveness ratio (ICER) from a healthcare perspective for three different scenarios:

1. Effectiveness based on trial efficacy; costs based on treatment protocol EORTC20981[91].
2. Effectiveness based on trial efficacy; costs based on matched real-world patients.
3. Effectiveness based on real-world evidence; costs based on matched real-world patients.

Model structure

A Markov model to calculate cost per life year gained (LYG) and quality-adjusted life years (QALYs) was developed in Microsoft Excel (2010) with a lifetime horizon of 20 years and one month cycle length. This cycle length is the time interval at which patients may switch health states. For scenario1 and scenario2, the model included three health states: progression-free survival (PFS), after progression survival (APS) and death. All patients in the model began in the PFS health state and remained in that state until

relapse/progression (APS state) or death, whichever occurred first. Per cycle (1 month), we calculated the transition between the health states, that is, the proportion of patients per health state. Patients entering the APS state either remain in APS or die. As data on PFS were initially not collected in PHAROS, time-till-next-treatment (TTNT) instead of PFS was calculated in scenario3. The structure of the model was similar except that the PFS state became the maintenance/observation (MOB) state, while the APS state became the next treatment (NT) state.

Model input – effects and utilities

Effects included both LYG and QALYs. Efficacy of rituximab maintenance therapy and transition probabilities for scenario1 and scenario2 were obtained from the EORTC20981 trial[91,92], while effectiveness in scenario3 was obtained from matched real-world data. Extrapolation was performed by fitting parametric distributions to the observed survival times. According to Dutch guidelines, the annual discount rate of future effects was 1.5%[96]. Utility values, 0.88 for PFS, 0.78 for APS and 0 for Death, were obtained from an observational study in the United Kingdom among 215 patients with FL[97].

Model input – cost and prices

We included direct medical costs including hospital inpatient days, day treatment, outpatient days and medication. For scenario1, healthcare use including treatment dosages, follow-up treatments and adverse events was obtained from the EORTC20981 trial. Hospital visits were not registered and were therefore based on the treatment protocol. For scenario2 and scenario3, resource use was obtained from the registries. All costs are in euros (2012) and discounted at 4% annually according to Dutch guidelines[96]. Costs of inpatient days (€548), day care (€177), outpatient days (€116) and intensive care (€2,100) were obtained from Franken et al.[62], while prices for drugs were derived from the national reference lists (medicijnkosten.nl) and multiplied with dosages according to the treatment protocol of the EORTC20981 trial[91]. To derive average costs for adverse events in scenario1, costs for neutropenia (€1,290) and infection (€1,096)[98] were multiplied with the proportion of patients experiencing these adverse events[92]. For scenario2 and scenario3, costs for adverse events were incorporated in the hospital visits. Average treatment costs upon relapse were based on post progression treatments in the third treatment line. It was assumed patients had one day treatment per cycle in scenario1.

Sensitivity analyses

Probabilistic sensitivity analyses were performed to explore the robustness of results. Input parameters varied simultaneously by running 5000 simulations, wherein resource use followed gamma distributions and utilities followed beta distributions. To investigate the effect on the ICER, three scenarios were introduced: A) zero discount rates, B)

using unmatched real-world resource use and C) using unmatched real-world resource use and effectiveness.

RESULTS

Table 5.1 shows the patient characteristics of the real-world population, the propensity matched groups obtained from the real-world data and the characteristics of the patients included in the EORTC20981 trial.

Real-world data

The two registries provided data of 3581 patients. After selecting patients responding to second-line chemotherapy (N = 113), 62 were only observed and 51 received rituximab maintenance. Median age in our real-world patient groups was around 62 years [Range: 31–92] and similar in both groups. In daily practice, a higher proportion of patients treated with rituximab maintenance received rituximab in the re-induction therapy (92% vs. 73%). Besides, disease progression was faster in the observation group; 73% reached the observation phase within 2 years from diagnosis, while 49% reached maintenance within 2 years from diagnosis. Compared with the EORTC20981 trial, real-world patients were 6 to 7 years older, and rituximab was prescribed more often in the re-induction therapy. Interestingly, rituximab maintenance in daily practice was also prescribed to patients with unknown or SD response to re-induction therapy.

Propensity score matching

Using logistic regression with rituximab treatment as the dependent variable and age, years since diagnosis, B-symptoms, FLIPI score, stage and response to re-induction therapy as independent variables, propensity scores, that is, the chance of receiving rituximab treatment, were calculated for patients without missing observations for age, years since diagnosis, B-symptoms, stage and response to re-induction therapy. The average probability of receiving maintenance treatment was 0.56 in the rituximab group, while the probability was 0.34 for patients in the observation group. This difference illustrated that real-world patients in both groups were not identical and patients in the rituximab group had, based on their characteristics, a higher chance (0.56 vs. 0.34) of receiving rituximab treatment. To create comparable groups, patients treated with rituximab maintenance were matched to patients from the observation group based on their chance of receiving rituximab, that is, their propensity score. More balanced groups were created after using propensity score matching, as shown in Table 5.1. For example, the proportion of patients with B-symptoms is similar as well as the re-induction therapy in line 2 and time since diagnosis. In addition, while patients receiv-

Table 5.1 Patient characteristics

	Real-world		Propensity matched real-world		Trial	
	Observation	Rituximab maintenance	Observation	Rituximab maintenance	EORTC Observation	EORTC Rituximab maintenance
Number	62	51	43	43[1]	167	167
Age at diagnosis						
Mean (sd)	57 (12)	59 (12)	57 (10)	59 (12)		
Median [range]	59 [33-80]	59 [29-90]	58 [36-80]	59 [29-90]		
Age at start 2nd line						
Mean (sd)	59 (12)	61 (13)	59 (10)	61 (13)		
Median [range]	61 [34-82]	61 [30-92]	61 [38-82]	62 [30-92]	55 [26-80][2]	
Age after response 2nd line						
Mean (sd)	59 (12)	62 (13)	60 (10)	62 (13)		
Median [range]	61 [34-83]	62 [31-92]	62 [39-83]	63 [31-92]		
Sex						
Male	45%	57%	51%	53%	49%[2]	
Stage at diagnosis						
I	6%	6%	9%	5%		
II	13%	10%	5%	9%		
III	23%	29%	14%	33%		
IV	58%	55%	72%	53%	67%[2]	
B symptoms						
B symptoms	31%	24%	21%	21%	27%[2]	

Table 5.1 Patient characteristics (continued)

	Real-world		Propensity matched real-world		Trial	
	Observation	Rituximab maintenance	Observation	Rituximab maintenance	EORTC Observation	EORTC Rituximab maintenance
missing	5%	2%				
Flipi score %						
Low	8%					
Intermediate	7%	10%	2%	9%		
High	36%	29%	37%	28%	70%	66%
unknown	49%	61%	60%	63%		
Time from initial diagnosis, %						
2 year or less	73%	49%	53%	51%	49%[2]	
More than 2 year	27%	51%	47%	49%	51%[2]	
Rituximab re-induction therapy before the start of maintenance/observation	73%	92%	93%	93%	59%	55%
Response to treatment line 2 %						
CR	37%	45%	74%	53%	29%	29%
CRu	6%	8%	7%	9%		
PR	56%	31%	19%	37%	71%	71%
SD		8%				
PD						
unknown		8%				

[1] Patient number after weighting according to the propensity weights. Unweight population N=23

[2] Patient characteristics reported at the start of the EORTC20981 study only. Therefore, a distinction per group could not be made.

FLIPI: Follicular Lymphoma International Prognostic Index

ing rituximab maintenance after a stable disease or unknown response to second-line chemotherapy were included in the initial real-world patient population, these patients were not included in the propensity matched real-world population. After matching, average propensity scores were similar in both real-world groups (0.56, SD 0.19), meaning the chance of rituximab maintenance was, based on the patient characteristics, similar in both groups. However, this balance reduces the number of patients especially in the observation groups, that is, 23 patients remained with weights between 1 and 5. These weights present the number of times patients in the observation group were matched to a patient in the rituximab maintenance group.

Healthcare use

Table 5.2 shows the healthcare utilisation in daily practice. Real-world average prescription of rituximab (366 mg/m2) and the number of cycles (6) were similar to averages of the EORTC study, 375 mg/m2 and 6, respectively. However, real-world rituximab was prescribed once every 2 instead of 3 months. Hospital visits for the matched observation group were reduced. This reduction was partly caused by four patients with relative high levels of healthcare utilisation, very low propensity scores (mean 0.083) and who died within 5 months after their observation phase started.

Table 5.2 Real-world rituximab prescription and health care use

Health care use	N	Mean (SD)	Median	Range
Rituximab				
Prescription rate (mg/m²)	28	366 (39)	374	[250-459]
Total prescription per patient (mg)	28	3542 (1776)	3345	[700-6800]
Total days maintenance	26	491 (207)	553	[91-763]
Number of cycles	29	6 (3)	7	[1-11]
Days per cycle	26	70 (16)	75	[21-90]
Real-world hospital visits per month				
Observation during MOB				
Outpatient visits	56	0.76 (0.69)	1	[0-2.66]
Day treatment	50	0.13 (0.3)	0	[0-1.31]
Inpatient days	50	0.61 (1.63)	0	[0-7.41]
Maintenance during MOB				
Outpatient visits	44	0.51 (0.47)	0	[0-2.63]
Day treatment	44	0.3 (0.26)	0	[0-1.37]
Inpatient days	34	0.19 (0.49)	0	[0-2.67]

Table 5.2 Real-world rituximab prescription and health care use (continued)

Health care use	N	Mean (SD)	Median	Range
Observation NT				
Outpatient visits	23	0.67 (0.68)	1	[0-2.4]
Day treatment	23	0.18 (0.26)	0	[0-1.12]
Inpatient days	23	1.39 (2.52)	0	[0-9.96]
Maintenance NT				
Outpatient visits	9	1.39 (1.41)	1	[0-4.18]
Day treatment	8	0.63 (0.56)	1	[0-1.72]
Inpatient days	8	1.67 (2.8)	0	[0-6.21]
Propensity matched hospital visits per month				
Observation during MOB				
Outpatient visits	41	0.62 (0.65)	0	[0-2.66]
Day treatment	37	0.18 (0.41)	0	[0-1.31]
Inpatient days	34	0.14 (0.58)	0	[0-3.28]
Maintenance during MOB				
Outpatient visits	39	0.55 (0.48)	0	[0.02-2.63]
Day treatment	40	0.31 (0.27)	0	[0-1.37]
Inpatient days	31	0.21 (0.51)	0	[0-2.67]
Observation during NT				
Outpatient visits	9	0.3 (0.38)	0	[0-0.95]
Day treatment	9	0.17 (0.17)	0	[0-0.4]
Inpatient days	9	0.39 (0.72)	0	[0-1.65]
Maintenance during NT				
Outpatient visits	7	1.79 (1.35)	2	[0.37-4.18]
Day treatment	7	0.72 (0.54)	1	[0.13-1.72]
Inpatient days	7	1.91 (2.93)	0	[0-6.21]

MOB: Maintenance or Observation health state, NT: Next treatment

Cost-effectiveness results

Table 5.3 shows the input parameters and assumptions for the cost-effectiveness of the three scenarios. PFS, TTNT as well as OS were modelled with lognormal parametric distributions.

The results on costs and effects of the three models are presented in Table 5.4. The ICERs for scenario1 were €11,259 per LYG and €12,655 per QALY and almost identical to the results for scenario3; €10,591 and €11,245 per LYG and QALY, respectively. Cost-effectiveness ratios for scenario2 were €21,202 per LYG and €23,821 per QALY. The scenario analyses are shown in Table 5.5. ScenarioA shows the impact of the discount

Table 5.3 Input parameters for cost-effectiveness base case scenarios

Parameter	Scenario1	Scenario2	Scenario3
General Parameters			
Effectiveness	Trial	Trial	Real-world
Discount rate (costs)	4%	4%	4%
Discount rate (efficacy)	1.5%	1.5%	1.5%
Time horizon of analysis (years)	20	20	20
PFS/MOB Parametric distribution	Lognormal	Lognormal	Lognormal
Overall survival parametric distribution	Lognormal	Lognormal	Lognormal
length of cycle (days)	30.4	30.4	30.4
Cost of Drug			
Rituximab 100 mg €	274	274	274
Rituximab 500 mg €	1,369	1,369	1,369
Demographic			
Average age of cohorts	54	60	60
Body weight (kg)	78	76	76
Height (cm)	170	174	174
Rituximab (3-monthly cycles maintenance 2 years at 375mg/m2)			
Number of cycles per month	0.33	0.44	0.44
Treatment duration maximum months	24	24	24
Average dose per administration	697	596	596
Rituximab arm (maintenance phase)			
Drug costs rituximab per month	€ 639	€ 714	€ 715
Cost of monitoring and administering per month	€ 75	€ 223	€ 223
Cost of adverse events per month	€ 238		
Observation arm (maintenance phase)			
Drug costs per month	0	0	0
Cost of monitoring and administering per month	€ 55	€ 172[1]	€ 172[1]
Cost of adverse events per month	€ 96		
After progression treatment costs			
Rituximab arm, costs per month	€ 786	€ 1,992	€ 1,992
Observation arm, costs per month	€ 752	€ 1,332	€ 1,332
Utilities			
Progression-free health state	0.859	0.859	0.859
After progression health state	0.798	0.798	0.798

[1]Costs per month including monitoring and administration as well as costs of adverse events.

Table 5.4 Cost-effectiveness results for three scenarios

	Scenario1 Trial efficacy and costs			Scenario2 Trial efficacy, real-world costs			Scenario3 Real-world effectiveness and costs		
	Rituximab	Observation	Incremental	Rituximab	Observation	Incremental	Rituximab	Observation	Incremental
Results base case analysis									
Effects									
Mean Time in PFS /MOB (yrs)	5.69	3.36	2.33	5.69	3.36	2.33	7.21	3.58	3.63
Mean Time in APS/NT (yrs)	3.70	4.48	-0.78	3.67	4.45	-0.79	2.96	4.35	-1.39
Mean OS	9.39	7.84	1.55	9.36	7.81	1.54	10.17	7.93	2.24
Mean QALYs	7.84	6.46	1.38	7.81	6.44	1.37	8.65	6.54	2.11
Costs									
Costs during PFS/MOB	15,606	4,298	11,308	13,366	6,278	7,088	16,749	6,696	10,054
Cost of Rituximab (€)	11,579	0	11,579	12,943	0	12,943	14,233	0	14,233
APS / NT costs (€)	29,422	34,884	-5,462	74,115	61,478	12,637	57,600	58,150	-551
Mean Total Cost (€)	56,608	39,182	17,425	100,424	67,756	32,668	88,582	64,846	23,736
Costs per LY gained			11,259			21,202			10,591
Costs per QALY gained			12,655			23,821			11,245
Results probabilistic sensitivity analysis									
Effects									
Mean OS	9.40	7.86	1.54	9.35	7.82	1.53	10.18	7.93	2.25
95% CI	(8.2;10.5)	(6.7;9.0)	(0.7;2.4)	(8.2;10.4)	(6.7;9.0)	(0.7;2.4)	(8.2;11.8)	(6.1;9.4)	(1.0;3.9)
Mean QALYs	7.85	6.48	1.37	7.81	6.45	1.36	8.67	6.55	2.12
95% CI	(6.9;8.8)	(5.5;7.4)	(0.6;2.2)	(6.8;8.7)	(5.5;7.4)	(0.6;2.1)	(6.9;10.3)	(5.1;7.8)	(0.9;3.7)
Costs									
Mean Total Cost (€)	56,663	39,153	17,510	100,404	67,832	32,572	88,231	63,861	24,369
95% CI	(41,984; 75,250)	(23,618; 60,116)	(-6,655; 40,348)	(47,886; 185,900)	(30,309; 125,512)	(-43,333; 126,369)	(45,338; 168,030)	(27,051; 121,749)	(-44,362; 105,977)
Costs per LY gained			11,376			21,288			10,829
Costs per QALY gained			12,789			23,919			11,499

PFS: Progression-free survival state MOB: Maintenance or Observation state APS: After Progression survival NT: Next Treatment OS: Overall survival QALY: Quality-adjusted life years LY: Life year

Table 5.5 Results scenario analysis

Base case value	Scenario1	Scenario2	Scenario3
Costs per QALY (€)	**12,655**	**23,821**	**11,245**
Results sensitivity analysis			
A) No discounting			
Costs per QALY (€)	12,697	26,228	12,608
B) Unmatched healthcare costs			
Costs per QALY (€)	-	5,162	-557
C) Unmatched healthcare costs & effects			
Costs per QALY (€)	-		-6,242

rate is marginal. ScenarioB and scenarioC show that using unmatched population-based resource, use and costs lead to different cost-effectiveness outcomes.

DISCUSSION

Using both real-world data and results from long-term trial follow-up showed favourable ICERs for rituximab maintenance compared with observation in patients with FL who responded to second-line chemotherapy. We compared real-world patients from the registries to the population in the EORTC20981 trial and observed differences, for example real-world patients were older while simultaneously, the proportion of complete responses to second-line chemotherapy was higher. Besides, whereas rituximab in the re-induction therapy was randomly assigned in the EORTC20981 trial, most patients treated in daily practice received rituximab in the re-induction therapy. As differences in patient characteristics between the two groups were observed, we performed matching to correct for possible bias. Thereafter, the ICER of three scenarios was calculated.

Relying entirely on updated trial results (scenario1) in combination with local cost data resulted in ratios of €11,259 per LYG and €12,655 per QALY. For scenario2, consisting of trial efficacy and matched real-world costs, ratios of €21,202 per LYG and €23,821 per QALY were calculated. Using real-world matched evidence (scenario3) for both effectiveness and costs showed ratios of €10,591 per LYG and €11,245 per QALY. While results from scenario1 and scenario3 were similar, this should not be interpreted as no need for real-world data. First, as most certainly, a cost-effectiveness analysis based on real-world or trial data will not always generate similar results. Second, scenario1 and scenario3 provide answers to different questions. Due to the randomisation, scenario1 has the highest internal validity of the treatment effect. However, healthcare decision-makers are also interested in external valid ICERs. For example, in the Netherlands, policy

regulations demanding sufficient relevant evidence on real-world cost-effectiveness have been implemented[51]. Expensive drugs receive conditional reimbursement for 4 years after which a re-assessment of the available data occurs. The requirements of the Dutch government seem to be best fulfilled by option B or C. Scenario3C has the highest generalisability because both costs and effects were based on real-world data and all patients were included. However, these results are difficult to interpret for the government because differences between two incomparable groups cannot be interpreted as incremental effects or costs. Scenario2B has the advantages of obtaining the incremental treatment effect from an RCT and including all real-world patients in the cost analysis. While data synthesis is a common approach to overcome problems due to the absence of a randomised design and calculate ICERs of treatment as prescribed in daily practice[93], our results illustrated that caution is required. First, using matched real-world costs but failing to account for the real-world effects (scenario2) results – in our case – to ICERs that were too high. Secondly, interpretation and generalisability problems arise when RCT effects and real-world costs are combined, that is, to which patients the ratios apply.

Even though existing registries reduce start-up costs, the collection of real-world data remains extremely time-consuming and expensive. While registries enabled us to obtain real-world data from 3581 non-Hodgkin patients, patient numbers in the rituximab and observation group were small, resulting in a considerable level of uncertainty (wider confidence intervals). Moreover, although propensity matching is a solid method to correct for imbalances in observational studies[95], it never substitutes randomisation. In addition, it should be noted that the differences in response rate to second-line therapy continued to exists. Another limitation of our study is that the registry data did not allow us to calculate PFS. As TTNT always exceeds PFS, the average time spent in PFS and APS (scenario1 and scenario2) is not directly comparable to the average time spent in MOB and NT (scenario3), respectively. Nevertheless, mean OS is comparable.

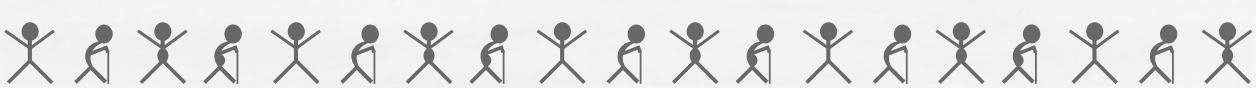
Although limitations exist, our results of scenario1 and scenario3 confirmed previous findings of favourable ICERs for rituximab maintenance of €8,729[99] and €12,600[100]. Incremental QALYs were 1.4 in scenario1 and scenario2 and 2.1 for scenario3, and these values are close to the 1.61 QALYs found by Demousis[101]. Real-world costs from our study in scenario3 (€88,582 and €64,846) corresponds quite well to the total costs of a French observational study, €71,314 and €62,251 for the rituximab and observation group, respectively[99]. The small difference might be explained by the fact that their study was conducted from a French perspective, costs were based on expert opinion, and no costs associated with death were included. Differences in effects between previously published studies were observed as well. Our OS was high in scenario1 and scenario2 compared with results from previous mentioned studies[99,100]. However, it should be noted that these studies were based on the outcomes of the EORTC20981 trial reported

after a median follow-up of 33 months[91], while our study used the long-term outcomes reported after a median follow-up of 6 years[92]. Remarkably, our OS based on real-world data (scenario3) was higher than the OS as observed in the EORTC20981 trial in spite of the older patient population investigated. This might be related to different regimens prescribed as first-line or follow-up treatment and a different time of inclusion. Besides, treatment with rituximab was an exclusion criterion in the trial, while almost 60% of our real-world patients received first-line treatment with rituximab.

The aim of our study was to provide healthcare decision makers with valid cost-effectiveness results that were generalisable to the real-world patient population. While scenarios B and C most closely match the requirements of Dutch policymakers, these options are scientifically less valid. In addition, differences between patients included in the trials and registry were observed. Therefore, we recommend healthcare decision-makers to reconsider their needs and recognise the importance of calculating several scenarios. Our case study illustrated that calculating real-world ratios is possible although caution is required with data synthesis and interpretation and generalisability of these results.

ACKNOWLEDGEMENTS

PHAROS, the Population-based HAematological Registry for Observational Studies, is a cooperation of HOVON (Dutch-Belgian Cooperative Trial Group for Hematology Oncology), the Institute of Medical Technology Assessment at Erasmus University Rotterdam and the Comprehensive Cancer Centers The Netherlands (IKNL) and South (IKZ). We are grateful for the data collection by the various registration teams of the Netherlands Cancer Registry. HemoBase is an initiative of the Working party Hematology Friesland. We are grateful for the data collected by the haematologist, pathologists and clinical chemists in the Frysian hospitals.

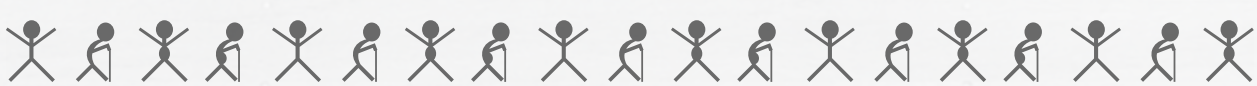


Chapter 6

A practical guide for using registry data to inform decisions about the cost-effectiveness of new cancer drugs: lessons learned from the PHAROS registry

Blommestein HM, Franken MG, Uyl-de Groot CA

PharmacoEconomics (2015): 33(6):551-560



ABSTRACT

Decision makers increasingly request evidence on the real-world cost effectiveness of a new treatment. There is, however, a lack of practical guidance on how to conduct an economic evaluation based on registry data and how this evidence can be used in actual decision making. This paper explains the required steps on how to perform a sound economic evaluation using examples from an economic evaluation conducted with real-world data from the Dutch Population based HAematological Registry for Observational Studies. There are three main issues related to using registry data: confounding by indication, missing data, and insufficient numbers of (comparable) patients. If encountered, it is crucial to accurately deal with these issues to maximize the internal validity and generalisability of the outcomes and their value to decision makers. Multivariate regression modelling, propensity score matching, and data synthesis are well-established methods to deal with confounding. Multiple imputation methods should be used in cases where data are missing at random. Furthermore, it is important to base the incremental cost-effectiveness ratio of a new treatment compared with its alternative on comparable groups of (matched) patients, even if matching results in a small analytical population. Unmatched real-world data provide insights into the costs and effects of a treatment in a real-world setting. Decision makers should realise that real-world evidence provides extremely valuable and relevant policy information, but needs to be assessed differently compared with evidence derived from a randomised clinical trial.

INTRODUCTION

Considerations of costs and cost-effectiveness are increasingly important for decision making on healthcare resource allocation. Economic evaluations enable a comparison of the cost-effectiveness of alternative treatments, and are thus especially important for decision making on reimbursement of new expensive drugs. Until recently, economic evaluations mainly consisted of cost-effectiveness analyses (CEAs) modelled from randomised controlled trial (RCT) data. RCTs aim to demonstrate the efficacy of interventions and ensure internal validity by randomly assigning which patients receive the new intervention. The circumstances in especially phase III trials are, however, not generalisable (i.e. externally valid) to a more heterogeneous group of patients treated in a real-world setting. Therefore, many uncertainties remain regarding the relevance of the results of RCTs in a real-world setting. Cost-effectiveness evidence based on RCT data may, therefore, not be sufficiently informative for decision makers. In such cases, evidence needs to be obtained from other sources, for example patient registries. A patient registry enables the evaluation of specified outcomes for a population defined by a particular disease, condition, or exposure, and when thoroughly designed and performed a patient registry can provide real-world evidence of clinical practice, patient outcomes, safety, and comparative effectiveness[13].

Guidelines on conducting and reporting economic evaluations are readily available[4,102,103], as well as questionnaires to assess the relevance and credibility of observational studies[104]. However, barriers still exist to use evidence from economic evaluations in actual decision making[105,106]. This necessitates the evaluation of the strengths and limitations of different types of evidence[12]. Moreover, practical guidance on using registry data for economic evaluations as well as on how these evaluations can be used in decision making is currently lacking.

This paper presents a practical guide on how to use registry data to inform decisions about the cost effectiveness of new drugs. We discuss the required steps of conducting a sound economic evaluation; the steps are explained by using the Population based Haematological Registry for Observational Studies (PHAROS) as an example. Although using registry data imposes some challenges, we illustrate that it is feasible to conduct an economic evaluation. We also discuss potential issues and limitations of economic evaluations based on registry data. The last section highlights the value of real-world economic evaluations for decision makers.[4]

PHAROS AND ITS CONTEXT

In the Netherlands, outcomes research requirements were implemented in 2006 for new expensive drugs to ensure timely access to promising drugs. If a drug is included in this policy, hospitals receive an additional ear-marked budget; however, with the obligation to gather data on appropriate drug use and real-world cost-effectiveness[12,107]. A reassessment after 4 years determines whether or not additional financing will continue. Real-world data are often collected within a patient registry.

One of the first Dutch patient registries was PHAROS. PHAROS is a population-based disease registry that started in 2010 with three haematologic malignancies (non-Hodgkin lymphoma, multiple myeloma, and chronic lymphatic leukaemia) in three regions; these regions cover 40% of the Netherlands[21]. PHAROS expanded over the years to other haematological malignancies (chronic myeloid leukaemia, myelodysplastic syndromes, and myelofibrosis) and is currently expanding to a nationwide coverage. Like many other registries, PHAROS was created to serve multiple purposes including measuring and improving the quality of care and determining the clinical and cost effectiveness of treatments used in a real-world setting. This paper uses examples of the economic evaluation[108] based on data from PHAROS. This economic evaluation was conducted to inform the reassessment of rituximab maintenance therapy for patients with follicular lymphoma, a subtype of non-Hodgkin lymphoma. A Markov Model was used with a 20-year time horizon to compare rituximab maintenance therapy in patients who responded to second-line chemotherapy with best supportive care (i.e. observation after a response to second-line chemotherapy). For further details we refer to Blommestein et al.[108].

CONDUCTING SOUND ECONOMIC EVALUATIONS WITH REGISTRY DATA

Economic evaluations typically include a number of steps, irrespective of the source of data. These steps, comprising existing guidelines in academic literature[4,102,103] are presented in Table 6.1.

The policy issue

Above all, it is important to define a clear objective for the economic evaluation and ascertain its relevance to healthcare decision making. One of the reasons to initiate PHAROS was to support decision making on the reimbursement of expensive drugs for three haematologic malignancies. Consequently, PHAROS data should facilitate the conduction of economic evaluations with real-world data.

Table 6.1 Steps of an economic evaluation

Step	Description
Policy issue	Define the objective of the economic evaluation and ascertain its relevance for health care decision making
Research question	Determine the main research questions (including what is studied for whom)
Perspective	Define the perspective of the study
Comparator	Identify the relevant alternative treatment(s)
Identify, measure, and value costs	Identify the relevant costs and measure these costs and value the unit costs
Identify, measure, and value outcomes	Identify the relevant outcomes and measure and value these outcomes
Calculate the cost-effectiveness ratio	Obtain the incremental costs and effects and calculate the incremental cost-effectiveness ratio
Sensitivity analyses	Analyse the uncertainty of the outcomes using deterministic, probabilistic and scenario analysis
Presentation and discussion of results	Present the results and discuss all issues of concern

Define the research question

It is crucial to determine the main research questions of the economic evaluation before setting up a registry that should collect the required data. For example, if a registry needs to be able to answer questions about the incremental cost-effectiveness ratio (ICER), relevant costs, and effects of at least two groups of patients are to be collected. Decision makers in the Netherlands require real-world evidence on appropriate use, effectiveness, and incremental cost effectiveness of drugs. Based on these requirements, the following research questions were defined for PHAROS: i) To whom and how is the drug of interest prescribed in daily practice? ii) What is the real-world effectiveness of this drug? iii) What is the real-world incremental cost effectiveness of this drug?

Regarding the first research question, PHAROS needed to include detailed data on baseline patient characteristics (including prognostic information) of patients who were treated as well as of patients who were not treated with the drug of interest. While a registry can be intervention based, PHAROS was set up as a disease-based registry. The advantage of using a disease-based registry is that all patients are included who meet the disease criteria. Therefore, PHAROS included patients eligible for treatment as well as patients ineligible for treatment. This also enabled identifying patients eligible for treatment but not treated with the drug of interest; these patients may serve as a comparator group. In addition, PHAROS needed to provide evidence on how drugs were used in daily practice. PHAROS not only included data on types of treatment, but also data on treatment regimens, dosages, dose modifications, treatment interruptions, and

treatment duration. Furthermore, from a policy perspective, it is important to obtain insight into equitable access to (expensive) drugs. Population-based registries can serve to obtain evidence on uptake by hospital and region; they may thus serve to reveal differences in access to a drug between regions and between university and general hospitals. In cases where data are based on a non-population-based registry, it is crucial that the selection is representative for the entire patient population as well as that a sufficient number of patients is included to ensure generalisability.

Regarding the second research question, PHAROS had to provide evidence on real-world effectiveness of the drug of interest. RCTs are the gold standard to demonstrate efficacy and assure internal validity by random assigning patients to a treatment strategy. In contrast, registries involve observational data and provide details on patients treated in daily practice. Reimbursement decisions may depend on the real-world use, effectiveness, and costs; in cases where a drug is not effective or not cost effective in daily practice, reimbursement of the drug may be reconsidered. If well designed, a registry includes information that enables accounting for heterogeneity in daily practice patients, physician variation, and the healthcare context. Therefore, effectiveness estimates based on registry data assure external validity and are thus generalisable to the real-world patient population. Ideally, the data should cover all treatments from diagnosis until death. However, this also depends on the length of follow-up and the time an analysis is required for policy making.

Regarding the third research question, PHAROS data needed to be able to demonstrate incremental real-world cost effectiveness of the drug of interest. Similarly to the second research question, a well-designed disease registry enables the estimation of incremental real-world effects, costs, and cost effectiveness simultaneously.

Define the perspective of the study

The perspective of the economic evaluation determines what type of costs and outcomes are to be included in the analyses. Most economic evaluations are conducted from a third-party payer or societal perspective. A societal perspective implies the inclusion of all relevant costs (direct and indirect, medical and non-medical costs) and relevant outcomes (quality of life and life-years). In contrast, in a third-party payer perspective non-medical costs are not included (e.g., traveling costs, productivity costs). Other used perspectives are healthcare, hospital, and patient. Requirements regarding the perspective may differ per country. It is, however, best to define the perspective before the start of data collection because it determines what costs and outcomes are needed for the economic evaluation. The objective of PHAROS was to gather evidence for the reassessment of expensive drugs in the Netherlands. Such a reassessment requires a societal perspective in the Netherlands.

Identify the comparator(s)

Economic evaluations involve a ''comparative analysis of alternative courses of action in terms of both their costs and consequences''[4]. The choice of comparator is crucial for the outcomes of the economic evaluation and it may potentially be a source of bias. In economic evaluations based on real-world data, it may not always be clear which alternative treatment is the most appropriate comparator and it may depend on the policy issue at stake. The most relevant alternative for decision makers is usually the current standard of treatment, this may also be best supportive care or a wait-and-see policy[108]. The inclusion of control groups to a registry adds to its complexity, time, and costs[13], but it allows the performance of a sound economic evaluation that compares a new treatment with the current standard of care. Collecting data over a long time period increases the chance that a registry includes an appropriate comparator group and avoids incomparable patient groups because of for example a rapid uptake of a new drug. This was, for example, illustrated by a Dutch observational study among patients with stage III colon cancer. Patients ineligible for the drug of interest had higher levels of unfavourable prognostic factors, i.e. carcinoembryonic antigen levels at baseline[109]. PHAROS included patients diagnosed from 2004 to 2012 and included relatively more patients in the comparator group who were included in the earlier years of the registry, while the intervention group included more patients who were diagnosed at the later years of the registry.

Identify, measure, and value relevant costs

Costs can be identified in the following categories; hospital resources, community care resources, patient and family resource use, and resource use in other sectors[4]. Guidelines regarding economic evaluations and valuation of unit costs can differ per country, as can the available data. We used Dutch data and the methods as set forward by Dutch guidelines[110].

Relevant cost items for inclusion in the registry depend on disease characteristics, the patient population, treatment strategies of interest, and the perspective of the study. It is usually not efficient to collect all potential cost components and a balance needs to be established between the relevance of the cost item relative to the burden of collection[13]. This balance can be based on previous research findings and/or determined in collaboration with treating physicians and based on professional guidelines. In PHAROS, data on hospital resource use were collected for outpatient visits, daycare treatment, inpatient days, and intensive care days. In addition, data on drug dosages, treatment duration, and supportive care were collected. Data on services provided outside the hospital were not collected.

Generally, data on hospital resource use can be collected from electronic hospital records and patient files. However, data can only be retrieved if it has been adequately

reported by physicians. Adequate reporting may be hampered in daily practice because physicians are not dictated by strict criteria as in trials. Patient questionnaires can be used to collect data on additional direct medical costs (e.g., healthcare providers outside the hospital, concomitant medication), direct non-medical costs (e.g., traveling costs), and indirect non-medical costs (e.g., productivity costs). It is important to note, however, that the inclusion of cost items other than direct medical may be hampered in a registry in which data are retrospectively collected. In PHAROS, we encountered several issues. First, information on resource use outside hospitals was expected to be extremely fragmented, especially in cases of severe diseases with centralised treatment. Patients in the PHAROS registry were often discharged from hospital and referred to different rehabilitation centres. Second, although PHAROS was initiated as a prospective registry, clinical and costs data were mainly collected retrospectively at several points in time. In other words, we started in 2010 to collect data from patients diagnosed from 2004 onwards. Patients were identified using the nationwide Netherlands Cancer Registry. This resulted, however, in a delay in the inclusion of patients.

Regarding productivity costs, PHAROS was supplemented with information from the Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship (PROFILES) study. This longitudinal cross-sectional study was conducted to obtain insight on amongst others quality of life and productivity losses of patients with follicular lymphoma[111]. However, the reassessment of the drug of interest was bounded by a 4-year re-evaluation period. At the time, our economic evaluation needed to be conducted for Dutch decision makers, the number of patients included in the longitudinal study was limited and data could not be matched to the disease states in our model. Therefore, the economic evaluation did not include productivity costs. We assumed that this was a conservative approach because the productivity costs for rituximab maintenance are most likely lower compared with the best supportive care group[100].

Furthermore, economic evaluations should only concern costs related to the disease and/or its treatment instead of the costs induced by unrelated diseases occurring simultaneously. It is important to note, however, that establishing such a relation is not always easy or clear-cut when using registry data. For example, admission of older patients to a nursing home may either be related to the disease but may also have occurred for other reasons. Moreover, determining which costs are related to the disease and/or its treatment is even less straightforward for an older population and in cases where comorbidities are present. Therefore, the inclusion of some cost items may be debatable.

The inclusion of cost items in the PHAROS economic evaluation was based on our previous experiences and supported by the literature that reported the same main cost drivers in treating haematologic patients[77,108]. Therefore, it was believed that an appropriate balance was achieved between registration burden and relevance of the cost

items. Such an evaluation of assumptions is crucial and depends on the characteristics of the patient population and the type of drug of interest. More detailed information regarding the included cost items and the unit costs are reported elsewhere[108].

The definition of the policy issue and research questions determines the cost components included in a registry. It is possible that researchers who conduct the economic evaluation are not yet involved at the start of the registry and must therefore rely on available data. In these cases, confirmations from the literature should be obtained to ensure that the most important cost components are included in the economic evaluation.

Identify, measure, and value outcomes of each alternative

The most preferred effectiveness outcomes for policy makers are overall survival (OS)/ life-years gained (LYG), and quality-adjusted life-years (QALYs) but also other clinical objectives linked to improvement in patients' outcomes can be included[110]. The follow-up in registries is generally much longer compared with RCTs and data are collected on subsequent treatments. Therefore, registries usually provide more information on OS. In addition, if data on life-time follow-up are collected, extrapolation of survival data, associated with uncertainty, is no longer necessary. Life-time follow-up is extremely valuable for economic evaluations because a lifetime horizon is usually required to incorporate all potential differences in effects and costs for the remainder of the patient's life[4]. However, because economic evaluations should provide timely results, it may be necessary to conduct evaluations prior to reaching the ideal follow-up time.

Regarding other effectiveness outcome measures, it is important to be aware that they may differ from the endpoints of an RCT. For example, primary endpoints of RCTs in cancer are most often response, time to progression, and progression-free survival; OS rarely is a primary endpoint in an RCT. In observational registries, however, data on response and progression may be biased because this may not be accurately captured in patient files[93]. Moreover, physicians in daily practice often do not report using standardized response criteria[112], whereas RCTs dictate response criteria. This may especially be the case when data are retrospectively collected by other individuals than the treating physician. The moment at which progression is established may also differ from an RCT because there is no strict monitoring scheme; progression could thus be established much later than it occurs. Therefore, we advocate using time-to-next-treatment (TTNT) as a proxy for progression, additional to survival, in economic evaluations based on registry data. Whenever a physician changes to another treatment, there must be a reason for doing so; progression can be one of them. In PHAROS, we used TTNT to model final outcomes (i.e. LYG and QALYs).

Regarding the adverse effects of treatments, these should be accounted for in the economic evaluation. However, identifying and measuring toxicity data may be ham-

pered in a registry. Although adverse events and their severity grading were collected in PHAROS, we encountered substantial issues establishing causal relations between the treatment and the adverse event.

Regarding the outcome quality of life, these data can be collected in a registry using patient-reported outcome measures. As mentioned previously, the number of patients included in the PROFILES study was still limited, and we could not match the data to the disease states in our model. Therefore, we based the utilities on findings in the literature.

Calculate the ICER

This step usually involves modelling methods such as Markov modelling or patient-level simulation modelling[177]. It is important to carefully select the model that best fits the data from the registry[185]. This step can greatly differ from only using data from an RCT. The main issues in calculating the real-world incremental cost effectiveness are associated with confounding by indication, missing data, and insufficient numbers of (comparable) patients. These issues will be further discussed in the next section. The ability to deal with these issues determines whether it is possible to develop a feasible model for the economic evaluation and to obtain valid incremental estimates based on real-world data only[93]. We used the methods as set forward by Dutch guidelines. Detailed information on the cost-effectiveness calculations performed with PHAROS data is reported elsewhere[108].

Assessment of uncertainty

The outcomes of an economic evaluation are surrounded with uncertainties, irrespective of whether the economic evaluation is based on data from an RCT or a registry. Therefore, it is important to extensively conduct analyses of the most important uncer-

Table 6.2 Scenario analysis of the PHAROS economic evaluation[108]

Scenarios	Data for effects (cases and controls)	Data for costs (cases and controls)	ICER per QALY	Total costs		Total QALYs	
				Cases	Controls	Cases	Controls
Scenario 1	RCT	RCT	€ 12,655	€ 56,608	€ 39,182	7.8	6.5
Scenario 2.1	RCT	Matched RW	€ 23,821	€ 100,424	€ 67,756	7.8	6.4
Scenario 2.2	RCT	Unmatched RW	€ 5,162	€ 96,720	€ 89,629	7.8	6.4
Scenario 3.1	Matched RW	Matched RW	€ 11,245	€ 88,582	€ 64,846	8.7	6.5
Scenario 3.2	Matched RW	Unmatched RW	€ -557	€ 85,096	€ 86,271	8.7	6.5
Scenario 3.3	Unmatched RW	Unmatched RW	€ -6,242	€ 81,231	€ 95,830	9.4	7.1

RCT: data from randomised clinical trials, RW: data from real-world practice
Table adapted from Blommestein et al.[108]

tainties. This information may be crucial for deciding on the adoption of a new drug. The uncertainty of input parameters can be analysed by scenario analysis as well as probabilistic and univariate sensitivity analyses[4]. In PHAROS, we observed great patient heterogeneity which resulted, in combination with small numbers of eligible patients treated with the drug of interest, in wide confidence intervals. In addition, as presented in Table 6.2, different scenarios based on different assumptions lead to different cost-effectiveness ratios (e.g., costs per QALY ranged from €11,499 to €12,789 to €23,919 in three scenarios[108]. While information regarding the assumptions for the model and appropriate sensitivity analyses on assumptions apply to all economic evaluations, we believe this is even more important when using registry data. Assumptions to calculate incremental outcomes might be because of the absence of randomisation, which is less straightforward.

Presentation of the results and discussion of all issues of concern to users

Presenting and discussing the results in an understandable matter is of utmost importance for the use of economic evaluations in decision making[105]. This may even be more important when the economic evaluation is based on data from registries because registry data are often less straightforward and more prone to bias. Topics that need to be reported depend on the conducted economic evaluation but should at least include: information on confounders, methods to account for missing values, validity, and generalisability of the results. The latter two are extremely important to determine usefulness of the results for decision makers[4]. It is also important to separately report both the effects and costs per alternative. Extremely high ICERs may, for example, indicate large cost differences between alternatives, but they can also result from small incremental effects.

THE MAIN ISSUES IN ECONOMIC EVALUATIONS BASED ON REGISTRY DATA

There are three main issues with conducting economic evaluations with real-world data from registries: i) confounding by indication; ii) missing data; and iii) insufficient number of patients. If encountered, it is crucial to appropriately deal with these issues to maximize the validity of the results of the economic evaluation and its value to decision makers.

Confounding by Indication

One of the main concerns about observational data raised in academic literature is the lack of a randomised controlled setting, which results in problems with internal validity[113-115]. Instead of treatment being randomly assigned as in an RCT, the choice of treatment is made by the treating physician based on characteristics of the patient. In addition, insurance coverage or national guidelines may also influence outcomes[12]. It is important to be aware that confounding by indication is a major challenge for economic evaluations based on observational data from registries. PHAROS showed that the real-world patient population was highly heterogeneous. When baseline patient characteristics associated with the outcome of interest differ between the treatment groups, the results of a study are biased if not appropriately corrected for these differences. We are aware that no correction method can substitute randomisation, but there are several methods that can be used to increase the validity of the outcomes.

Methods to deal with confounding by indication are for example multivariable regression modelling, propensity score (PS) matching, and data synthesis. Multivariable regression modelling has been the conventional method to reduce bias related to confounding by indication. Potential confounders are included simultaneously in a regression model that estimates final outcomes. Using multivariable regression models for registry data requires information on patient and disease characteristics. In the past decade, there has been an increasing trend of using PS matching techniques[116]. This technique allows the calculation of the chance of receiving the treatment of interest by using observed patient characteristics[95]. Propensity scores are then used to match a treatment group to a comparator group based on patients who have similar chances (i.e. propensity scores) of receiving the treatment of interest. Other applications of the PS include stratification, covariance adjustment, and weighting[95,117]. Although PS matching techniques are increasingly and successfully used[116,118], these techniques are less attractive when multiple treatment strategies are compared simultaneously. A better understanding of the benefits and limitations in practical circumstances of PS matching vs multivariate risk modelling is still needed[116].

Finally, in case correcting for confounding is hampered (e.g., missing values or a lack of a control group), data synthesis can be used to model incremental outcomes. For example, it may be a good option to synthesize efficacy data from an RCT with effectiveness data from daily clinical practice, especially when an appropriate comparator group is lacking[119]. However, it is important to be aware that there was an initial need for data from daily practice because patient baseline characteristics may differ between patients treated and not treated in an RCT.

Missing data

Even when a registry is well designed and executed by an active interdisciplinary collaborative research group, it is to be expected that missing values on certain variables will exist. Therefore, only analysing complete cases is most likely not possible. Although imputing mean values might be less of a problem for RCT data, this method is not to be recommended because the patient population in daily practice is usually far more heterogeneous. We recommend using the multiple imputations method because this method not only imputes missing values but also accounts for the uncertainty associated with the imputed value by creating multiple datasets[120]. Missing values are imputed based on observed variables. To account for the uncertainty of the predicted variables, each missing value is imputed multiple times resulting in several complete datasets. The analyses of the combined datasets produce overall estimates and standard errors that reflect the uncertainty around the imputed variables. However, it is important to note that this method can only be used for missing values that depend on known and observed variables (i.e. variables missing at random[121]).

Insufficient number of comparable patients

Sufficient numbers of patients and follow-up data are required for conducting a sound economic evaluation with registry data. This is, however, sometimes difficult to realise in daily practice. A large difference may exist between the actual patient population (i.e. the population included in the registry) and the analytic patient population (i.e. the population that met the criteria for analysis[13]). RCTs usually base the number of patients included on power calculations and continue including patients until the desired number has been reached. This is, however, not possible in daily practice; for example, if physicians no longer use the alternative treatment, the analytic population will be small. The minimal required number of patients also depends on the extensiveness of the heterogeneity of the real-world patients, which may not be known in advance. The option to actively search for extra patients treated with the drug of interest has to be balanced with a potential diminishing generalisability.

In PHAROS, we faced confounding by indication, missing data, as well as a small analytical patient population. First, confounding by indication was present because the comparator group included relatively more patients with a worse prognosis compared with the treatment group[108]. We used PS matching methods to correct for observed differences in patient and disease characteristics. After matching, both groups were more balanced regarding characteristics of re-induction therapy, B symptoms, and disease progression. Table 6.2 illustrates the variation on outcomes of our scenario analyses in which we used both matched and unmatched data. Second, we encountered a small analytical patient population in PHAROS. The actual population included nearly 700 patients with follicular lymphoma. However, the required analyses were too early

for most patients because the patients did not (yet) receive a second line of chemotherapy. Therefore, only 14% of the actual population was included in the analytic population. To increase the number of patients, data were obtained from Hemobase, a multidisciplinary Web-based electronic patient record in the north-eastern part of the Netherlands that collected similar data. Although this increased the analytic population from 89 to 113 patients, the number of patients remained small. The rather small and highly heterogeneous population led to wide confidence intervals for treatment with rituximab maintenance (e.g., OS of matched real-world effects ranged from 1.0 to 3.9 years and costs ranged from -€44,362 to +€105,977). Third, because missing data were present for relevant outcomes (e.g., response rates), the number of patients included in our analyses reduced even further after applying PS matching (e.g., N = 51 reduced to N = 43 in the rituximab group).

THE VALUE OF REAL-WORLD ECONOMIC EVALUATIONS FOR DECISION MAKERS

Decision makers often make limited use of evidence from economic evaluations[122,123]. There is, however, a higher chance that decision makers use such evidence if the evidence is accessible (i.e. timeliness and understandable) and acceptable (i.e. accuracy and validity of research methods given institutional requirements)[106]. Above all, it is crucial that decision makers realise that registry data differ from RCT data and that the outcomes of their economic evaluations should thus be assessed differently. This should, however, not be seen as a drawback, but rather as an important opportunity. Both data sources complement each other; they allow balancing internal validity and generalisability and answer different questions.

The economic evaluation based on PHAROS data demonstrated these differences by calculating different scenarios. Table 6.2 presents these scenarios as well as their outcomes. We discuss the value of each scenario for healthcare decision makers regarding whether the research methods were accessible and acceptable.

Scenario 1 was only based on RCT results; no real-world data were included in the analyses. Randomisation ensured the internal validity; therefore, the difference between the intervention group (i.e. patients who received rituximab maintenance therapy) and the control group (i.e. patients who were only observed) could be attributed to the treatment. In other words, treating patients with rituximab maintenance therapy costs €12,655 per QALY gained compared with observation only. This scenario used well-known conventional methods (RCT data) and may thus be highly accessible and acceptable to decision makers. Accessibility and acceptability is ensured by understandable results, i.e. economic evaluations based on trial data are intuitive because

conventional methods are used. This is, however, at the cost of generalisability, because no data were used from daily practice. The results do not inform decision makers on the expected costs and effects in the real-world patient population while this was the policy issue at stake. As a consequence, none of the questions raised by decision makers (i.e. to whom and how is the drug prescribed and what is the real-world cost effectiveness) can be answered with scenario 1.

In scenarios 2.1 and 2.2, efficacy data from the RCT were combined with matched and unmatched real-world cost data, respectively. This resulted in substantial differences in the estimated costs per QALY gained (€23,821/ QALY for scenario 2.1 and €5,162/QALY for scenario 2.2). Because both scenarios combined RCT data with real-world data, the interpretation of the outcomes may be more complicated because it is unclear to whom the results apply, i.e. trial, real-world patients, or both. In other words, results are less accessible for decision makers. The effectiveness estimates are internally valid because they are based on RCT data, but they do not inform decision makers on the effectiveness in daily practice. In contrast to scenario 1, both scenarios 2.1 and 2.2 provide information on real-world costs. It should be noted, however, that the accuracy of the incremental costs in scenario 2.2 may be impeded because patients treated and not treated with rituximab maintenance therapy were not comparable and we did not correct for these differences by using a matching method. Moreover, it is questionable for whom the ICER is actually valid (i.e. the efficacy estimates apply to trial patients while the cost estimates apply to the real-world patient population). Therefore, both ICERs should be carefully interpreted.

Scenarios 3.1, 3.2, and 3.3 used real-world data for both cost and effectiveness estimates. Consequently, the results are generalisable to the real-world patient population and applicable to the policy issue at stake. Because decision makers are less familiar with interpreting real-world data, these scenarios may be less accessible for decision makers. It is, therefore, crucial that the methods and results are extensively reported in an understandable language. Unmatched data as used in scenarios 3.2 and 3.3 inform decision makers on the real-world costs and effects, but a major drawback is that differences cannot be assessed between cases and controls because the incremental estimates are not sufficiently valid. Both scenarios 3.2 and 3.3 show higher total costs for the control group while the opposite was expected and shown by the other scenarios. Although matching methods reduced the analytical population, we believe that scenario 3.1 provides the most accurate and valid results because matching methods were used for both costs and effects to reduce bias related to confounding by indication.

Decision makers were interested in real-world outcomes and, in the Dutch case, required evidence from daily clinical practice to reduce the uncertainty of both real-world costs and effects of rituximab maintenance therapy. We believe that the computed ICERs can only be used if the applied methods are accurate and valid. In other words,

incremental outcomes of economic evaluations can only be used when cases and controls are comparable or when appropriate methods are used to correct for differences in baseline characteristics (scenarios 1 and 3.1). In cases where baseline characteristics greatly differ between patient groups and no matching methods have been used, the outcomes of an economic evaluation should not be acceptable for decision makers because the incremental outcomes are not accurate and not valid. We believe that scenario 3.1 is most valuable to decision makers because this scenario achieves an appropriate balance between generalisability and internal validity. The estimated cost-effectiveness ratio (€11,245) also provides reassurance to decision makers that efficacy from the trial can be realised at favourable costs in the real-world patient population. However, because a formal decision has not yet been made, it is currently unknown how decision makers interpreted and evaluated the outcomes.

FURTHER RESEARCH AREAS FOR REGISTRY DATA

Expensive cancer drugs are increasingly developed for patient populations stratified by genetic characteristics and this trend illustrates an increasing role for biochemical, histological, and genetic markers to aid treatment decisions[124]. While the PHAROS registry focused on expensive drugs, registries may also be used to collect information on biochemical, histological, and genetic markers, which can be used for economic evaluations of these markers. This may be an important subject for further research using registry data.

FINAL REMARKS

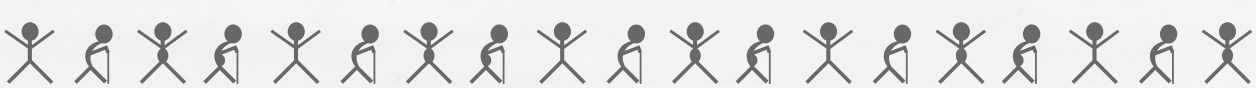
It is important for decision makers that a drug provides sufficient value in relation to its costs in daily practice. Economic evaluations based on real-world data can provide extremely valuable insights into real-world incremental cost effectiveness[108,119,125]. In PHAROS, both matched and unmatched outcomes seem favourable for the decision to adopt rituximab maintenance therapy. In other cases, the variation in outcomes can be much greater and less favourable than in PHAROS, which necessitates a careful evaluation of the causes of the conflicting results between RCT and real-world data. Moreover, it may not always be possible to develop a feasible model with real-world data to calculate incremental estimates[93]. We advocate that incremental estimates (ICERs) should always be based on matched patients in case patient groups are incomparable. However, unmatched real-world data are still valuable for decision makers because they provide evidence on costs and effects of a treatment in a real-world setting, although

not incremental[77,93,123]. Real-world evidence can also be used to obtain a certain level of reassurance regarding the extent to which the evidence from the RCT is applicable to the real-world patient population. It is, however, crucial that decision makers realise that the outcomes of an economic evaluation based on registry data should be assessed differently compared with the outcomes of an economic evaluation based on RCT data. The need for generalisable outcomes has to be balanced with the need for internally valid outcomes. While registries are able to provide insight into the use, effectiveness, and costs of a therapy in routine clinical practice and therefore offer healthcare decision makers with realistic expectations for outcomes in real-world patients, it should be noted that other solutions exist to balance internal and external validity. For example, pragmatic trials can include a broad patient population and can thus also ensure generalisability. Pragmatic trials have the major advantage of randomising treatment but are on the other hand, however, associated with logistical, ethical, and sample size challenges as well as high resource investments[126].

In PHAROS, we demonstrated that it was feasible to conduct a real-world economic evaluation using registry data. We believed that we provided decision makers with acceptable and accessible information and showed that the real-world outcomes confirmed the efficacy of the trial. In our opinion, this provided reassurance to decision makers about a drug's value for money in daily clinical practice.

ACKNOWLEDGMENTS

PHAROS, the Population-based HAematological Registry for Observational Studies, is a cooperation of HOVON (Dutch-Belgian Cooperative Trial Group for Hematology Oncology), the Institute of Medical Technology Assessment at Erasmus University Rotterdam, and the Comprehensive Cancer Center, the Netherlands (IKNL). We are grateful for the data collection by the various registration teams of the Netherlands Cancer Registry. PHAROS is financially supported by ZonMW 2009–2012 (#80-82500-98-01007) and through unrestricted grants from Bayer, Roche, Janssen, Celgene, Sanofi, MSD, Novartis, GSK, Mundipharma, and BMS.



Chapter 7

A cost-effectiveness analysis of real-world treatment for elderly patients with multiple myeloma using a full disease model

Blommestein HM, Verelst SGR, de Groot S, Huijgens PC, Sonneveld P, Uyl-de Groot CA

European Journal of Haematology (2015) Apr 18 Epub ahead of print

ABSTRACT

Objectives: To study the impact of novel treatments for elderly (≥66 years) patients with multiple myeloma (MM) in daily practice by comparing real-world effects (overall survival (OS) and quality-adjusted life years (QALYs)) and costs over time. Also, we calculate cost-effectiveness of treatment sequences commonly prescribed to predict effects and costs if patients had received a different treatment sequence.

Methods: Real-world data including patient and disease characteristics, treatment information and resource use were collected from 1054 elderly patients with MM. Patients received first-line treatment during 2004–2007 (cohort 1) and 2008–2013 (cohort 2). The two cohorts were compared using a patient-level simulation (PLS) model comprising regression models which used patient and disease characteristics to estimate time to next treatment and death. Effects and costs from cohort 2 were compared to 4 commonly prescribed real-world sequences.

Results: Utilisation of novel agents was higher for cohort 2 compared to cohort 1. Modelled average OS for cohort 1 was 38 months (median 25) and total costs €44,200. OS for cohort 2 was 42 months (median 28) and total costs €69,017. The model identified potential OS gains if all patients were to be treated using combinations containing thalidomide, lenalidomide and bortezomib in that particular order. This sequence had, compared to real-world treatment, the most favourable incremental cost-effectiveness ratio, €24,618 per life year gained and €34,875 per QALY.

Conclusions: Our patient-level model enabled to study the effects and costs of entire treatment sequences and to compare real-world treatment patterns over time. Increased utilisation of novel agents improved survival and increased costs for real-world patients with MM in the Netherlands.

INTRODUCTION

Like many cancers, multiple myeloma (MM) is an incurable disease and treatment is characterised by sequential treatment lines aiming to prolong progression-free and overall survival (OS). MM is the second most common haematological malignancy with 20,180[127] and 37,200[128] newly diagnosed patients in the United States (2010) and Europe (2008), respectively. Median age at diagnosis is 70 years and the number of patients with MM is expected to increase due to ageing of the population[129,130]. The clinical course of the disease is very heterogeneous with a wide variation in OS[131,132]. While median OS is 3.7 years for all patients, OS decreases steadily with increasing age, that is OS is 5.2 years in patients ≤50 and 2.6 years in patients ≥80[133]. During the past decade, improved OS in a real-world setting has been reported initially mostly for younger patients[134] but recently also for elderly patients[135]. Improvements in survival were linked to advances in treatment, but besides therapy, many patient and tumour characteristics contribute to the final outcome[136,137]. The majority of the patients is ineligible for high-dose therapy followed by stem cell transplantation due to age (age ≥66 yr.) or coexisting conditions[130]. The introduction of innovative treatments such as thalidomide, bortezomib and lenalidomide during the past decade changed treatment of MM and improved OS[138-141]. However, these improvements imply increasing health-care costs[142] emphasising the relevance of studying costs and effects simultaneously in cost-effectiveness analysis (CEA).

Randomised controlled trials (RCTs) are the golden standard for establishing efficacy and are frequently used for CEAs[143-146]. However, generalising data of RCTs to patients in daily clinical practice are hampered because MM mainly affects elderly and/or heterogeneous patients generally not included in RCTs[147]. So while data from RCTs are available for this patient population and have demonstrated efficacy of, for example, thalidomide[139], it is unclear whether similar outcomes are achieved among patients treated in daily practice. For example, what is the effectiveness of thalidomide treatment in a real-world setting given the patient and disease characteristics of real-world patients and the context in which treatment is provided. In addition, RCTs are designed to compare treatments covering only a limited time period. Consequently, using RCTs, a CEA covering multiple treatment sequences in a row cannot be made.

While several RCTs and cost-effectiveness studies were conducted for patients with MM, these studies provide us only with pieces of information that – even after combining them – do not represent the complete real-world picture. Hence, a CEA of treatment sequences based on real-world data is of utmost importance. The aim of this study was twofold: to provide insight into real-world costs and effects over time and to compare real-world cost and effects of different sequences for real-world patients.

Regarding the first, the novelty of this study is that it provides insight into real-world OS and costs. These data on effectiveness provide additional important information to the efficacy findings from RCTs currently available. Dutch treatment guidelines recommended thalidomide as first-line treatment from 2008 onwards, and we hypothesised that this change influenced both effects and costs of MM treatment. Therefore, results were calculated separately for two cohorts; before and after thalidomide was recommended as first-line treatment.

Regarding the second aim, real-world data enabled us to study treatment sequences, that is multiple treatment lines in a row, instead of only covering one line of treatment. We calculated the cost-effectiveness of different treatment sequences to identify whether effects and costs for patients would have been better if patients had received a different treatment sequence. These sequences included drugs such as thalidomide, bortezomib and lenalidomide. Effectiveness of the drugs in these scenarios was obtained from real-world data controlling for patient and disease characteristics.

METHOD

Clinical analysis

Data sources and patients.

Data were derived from the Netherlands Cancer Registry (NCR) and the Population based Haematological Registry for Observational Studies (PHAROS)[21,148]. In PHAROS, patients from three regions in the western and southern part of the Netherlands diagnosed with MM between 2004 and 2011 were included. From these patients, we selected those patients who actually received treatment for MM and were at least 66 yr. old (N = 1054) as these patients were not eligible for a stem cell transplantation. Based on changed guidelines, the patients were divided into cohorts: cohort 1 included patients receiving first-line treatment between 2004 and 2007, while cohort 2 included patients receiving first-line treatment between 2008 and 2013. Missing values were present in the data set and imputed using multiple imputations by chained equations for each variable[120,149]. Ethical approval for the study was obtained from the Dutch Medical Ethics Committee.

Time and events.

For each patient from the PHAROS data set, the duration of first-line treatment (TTE1) was calculated. This was the time from first-line to second-line treatment for patients receiving more than one line of treatment. For patients who received only first-line treatment, the time to death was calculated. The duration of second-line treatment (TTE2) was obtained similarly. For patients who received a third-line treatment, the

time from third-line treatment to death (TTE3) was calculated. OS was obtained by summarising TTE1, TTE2 and TTE3.

Regression models

Health economic evaluations require data on the costs and benefits of treatments over the lifetime of patients, and extrapolation beyond follow-up was performed by fitting parametric distributions (i.e. exponential, Weibull, gamma, Gompertz, lognormal and log-logistic[150]) to the observed TTE1, TTE2 and TTE3. Parametric functions were assessed for their goodness of fit to the data using Akaike information criteria (AIC), Bayesian information criteria (BIC) and visual inspection. After identifying the distribution that best fitted the data, potential parameters that could be included in the regression models were identified from PHAROS[151], because this registry included prognostic variables for survival or the start of a new treatment. We used a forward selection method to decide which variables should be included in the regression models to predict TTE1, TTE2 and TTE3. Variables were included if the inclusion resulted in better AIC or BIC and significance levels were acceptable ($P < 0.1$). The types of events were predicted with logistic regression models. Variables for the type of event models were selected through forward selection and impact on the R^2 of the models and significance levels ($P < 0.1$).

Simulation study

Model type

Modelling is necessary to calculate life-time costs and effects and the cost-effectiveness of several scenarios. To model the cost-effectiveness of MM treatments properly, a flexible modelling method was required. Patient-level simulation (PLS) models (a subcategory of discrete event simulation models) focus on entities (patients) with attributes (characteristics) and events and can easily incorporate treatment history and patient characteristics[152].

Base case and scenario analyses

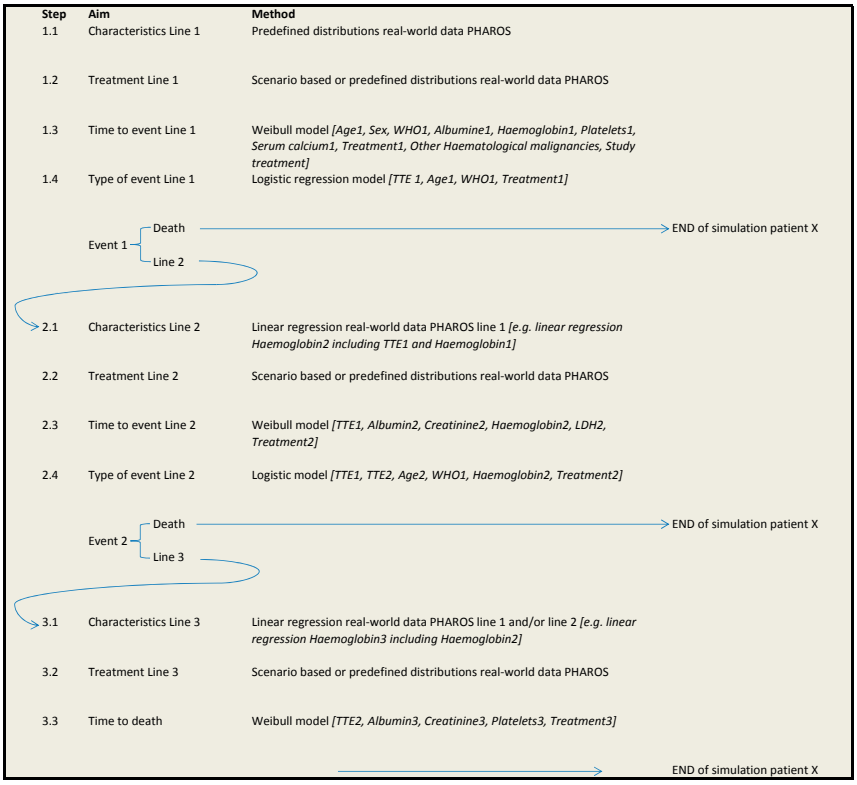
Real-world treatments were categorised in five groups including the following regimens: melphalan–prednisone (MP), thalidomide-based (Thal), bortezomib-based (Bmib), lenalidomide-based (Lena) and other treatments (Other). Base case analysis modelled the effects and costs of real-world treatment and showed these outcomes separately for cohort 1 and cohort 2 to answer the first research question. Real-world treatment in the base case consisted of a mix of therapies prescribed during these time frames, that is real-world treatment patterns derived from PHAROS. To answer the second research question to see whether real-world treatment could be improved, costs and effects of

alternative scenarios were calculated. These scenarios included treatment sequences most commonly prescribed during our observation period: MP–Thal–Bmib, MP–Thal–Lena, Thal–Bmib–Lena and Thal–Lena–Bmib. In the first two scenarios, all patients (100%) in the first-line were treated with MP, while in the third and fourth scenario, all patients received first-line treatment with Thal. Second- and third-line treatments were only assigned to patients who did not die after first- or second-line treatment, respectively. The number of patients who received second- and third-line treatment per scenario is included in the Supporting information. Costs and effects of the sequences were calculated and compared to real-world treatment from the most recent cohort (cohort 2).

Model simulation

Each model run included the simulation of 1054 patients. As it is a PLS model, each patient is simulated individually. The model simulation of one patient consists of a maximum of three treatment lines (Figure 7.1). The first treatment line included four steps. First (1.1), patient and disease characteristics were assigned to the patient by drawing

Figure 7.1 Example of the simulation of one patient



random numbers from predefined distributions based on real-world data. Second, treatment was allocated to the patient (1.2). Depending on the scenario, treatment was either obtained from predefined distributions based on real-world treatment patterns or assigned based on one of the four scenarios. The third step (1.3) calculated the time to the first event (TTE1). TTE1 was obtained from individual survival curves based on the parametric survival model included in the Supporting Information. The fourth step (1.4) determined the type of event (Event 1), next treatment or death. The simulation ends after one line of treatment for patients who died. Second-line treatment was simulated with four similar steps (2.1, 2.2, 2.3 and 2.4) for patients who received a second-line of treatment. Third-line treatment was similar to the previous treatment lines except that time to death (TTE3) was modelled.

Input parameters

Utility values (necessary to calculate quality-adjusted life years (QALYs)) for MM are unavailable for elderly real-world patients in the literature. A population based cross-sectional study was conducted in the Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship (PROFILES) registry[111]. In 2009, 101 MM patients of 66 yr. or older received a questionnaire including – because the official 5-level version was not available yet – a preliminary version of the EuroQol-5 dimensions 5 level. Based on 61 patients, the average utility vale was 0.76 (SD 0.21, range: 0.005 to 1.0), calculated with the Dutch EQ-5D-5L Value Set[153].

Resource use including outpatient and day care visits, inpatient wards and intensive care days per treatment were obtained from PHAROS. Costs of inpatient days (€562), day care (€182), outpatient days (€120) and intensive care (€2,377) were obtained from Gaultney et al.[77] Drug dosages were obtained from guidelines, and prices were derived from the official Dutch price list[154]. For generalisability reasons, the price for thalidomide – which is substantially lower in the Netherlands compared to European countries – was based on the literature[155]. Costs were determined for the year 2014, and according to Dutch guidelines, discount rates were 1.5% for effects and 4% for costs[156]. Additionally, discount rates were 0% to calculate undiscounted results.

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) was performed to explore the possible outcomes and the likeliness of these outcomes. In the PSA, probability distributions for input parameters were used instead of point estimates to reflect the uncertainty of these parameters. Input parameters were varied simultaneously, and the model was run 1000 times for 1054 patients. Input parameters for resource use followed gamma distributions, unit costs normal distributions and utility values beta distributions. Regression

models were varied using Cholesky decomposition to retain the correlations between the parameters seen in the covariance matrix[157].

RESULTS

First, the clinical analysis based on real-world data will be discussed; these results served as input for the simulation study for which the results are discussed in the second part of this paragraph.

Clinical analysis

Patient characteristics and treatment patterns

From PHAROS, we obtained data from 1054 elderly patients with MM. Median follow-up was 50 months. Table 7.1 shows the patient and disease characteristics at diagnosis of the real-world data as well as the characteristics after multiple imputation. Table 7.2 presents real-world treatment patterns for line 1, line 2 and line 3 per cohort. Different treatment patterns were observed for cohort 1 and cohort 2. For example, MP was

Table 7.1 Patient and disease characteristics at diagnosis and first line treatment of the original real-world data, imputed data and the simulated patient population

Patient characteristics at diagnosis		**Real-world patients**	**Imputed data**	**Simulated patients***
		N=1054	N=1054	N=1054
Age	Mean	76	76	76
	Median [Range]	75 [66-93]	75 [66-93]	76 [66-93]
	Missing	0%		
Male sex	N(%)	552(52%)	552(52%)	553(52%)
	Missing	0%		
WHO status	0	298(36%)	378(36%)	377(36%)
	1	388(46%)	482(46%)	480(46%)
	2	111(13%)	141(13%)	141(13%)
	3	30(4%)	41(4%)	45(4%)
	4	8(1%)	11(1%)	11(1%)
	Missing	21%		
ISS[1]	1	191(29%)	292(28%)	287(27%)
	2	226(34%)	344(33%)	354(34%)
	3	247(37%)	417(40%)	413(39%)

Table 7.1 Patient and disease characteristics at diagnosis and first line treatment of the original real-world data, imputed data and the simulated patient population (continued)

Patient characteristics at diagnosis		Real-world patients	Imputed data	Simulated patients*
		N=1054	N=1054	N=1054
	Missing[2]	37%		
Serumß$_2$ Microglobulin	Mean	5.8	6.2	6.2
	Median (Range)	4.4 [0.41-35]	4.5 [0.03-38]	4 [0.03-38]
	Missing	37%		
Albumin level	Mean	34	35	35
	Median (Range)	35.4 [1.9-45]	36 [1.9-45]	36 [1.9-45]
	Missing	15%		
Haemoglobin	Mean	6.7	6.7	6.7
	Median (Range)	6.6 [2.8-10.4]	6.6 [2.8-15]	7 [2.8-15]
	Missing	0.3%		
Creatinine	Mean	137	137	138
	Median (Range)	98 [3-998]	98 [3-998]	98 [3-998]
	Missing	1%		
Platelets	Mean	243	243	244
	Median (Range)	229 [16-1170]	229 [16-1170]	231 [16-1170]
	Missing	2%		
LDH[3]	Mean	271	272	272
	Median (Range)	238 [17-1083]	239 [17-1083]	240 [17-1083]
	Missing	13%		
Serum calcium	Mean	2.4	2.4	2.4
	Median (Range)	2 [1.27-6.3]	2 [1.27-6.3]	2 [1.27-6.3]
	Missing	5%		
Salmon Durie stage	I A/B	160(16%)	165(16%)	165(16%)
	II A	185(18%)	191(18%)	190(18%)
	II B	53(5%)	54(5%)	53(5%)
	III A	487(48%)	506(48%)	513(49%)
	III B	131(13%)	138(13%)	133(13%)
	Missing	4%		
Comorbidity Other malignancy		165(22%)	230(22%)	228(22%)
	Missing	28%		
Study treatment line 1		147(14%)	147(14%)	151(14%)

[1]ISS denotes International Staging System [2] Proportion of missings partly caused by availability of ISS since may 2005 [3]LDH denotes Lactate dehydrogenase

* Average of ten simulation runs

Table 7.2 Real-world treatment patterns per line per cohort

	Cohort 1 First-line treatment 2004-2007	Cohort 2 First-line treatment 2008-2013
Line 1		
N	396	658
Melphalan/Prednisone	187 (47%)	43 (7%)
Thalidomide	158 (40%)	460 (70%)
Bortezomib	3 (1%)	73 (11%)
Lenalidomide	2 (1%)	54 (8%)
Other	46 (12%)	28 (4%)
Line 2		
N	213	288
Melphalan/Prednisone	36 (17%)	33 (11%)
Thalidomide	103 (48%)	40 (14%)
Bortezomib	34 (16%)	124 (43%)
Lenalidomide	16 (8%)	79 (27%)
Other	24 (11%)	12 (4%)
Line 3		
N	112	109
Melphalan/Prednisone	10 (9%)	2 (2%)
Thalidomide	29 (26%)	12 (11%)
Bortezomib	33 (29%)	34 (31%)
Lenalidomide	26 (23%)	51 (47%)
Other	14 (13%)	10 (9%)

the most common (47%) first-line treatment in cohort 1, while Thal was most often prescribed as first-line treatment in cohort 2 (70%). In addition, for cohort 2 compared to cohort 1, higher prescription rates of Bmib and Lena were observed in both second- and third-line treatments.

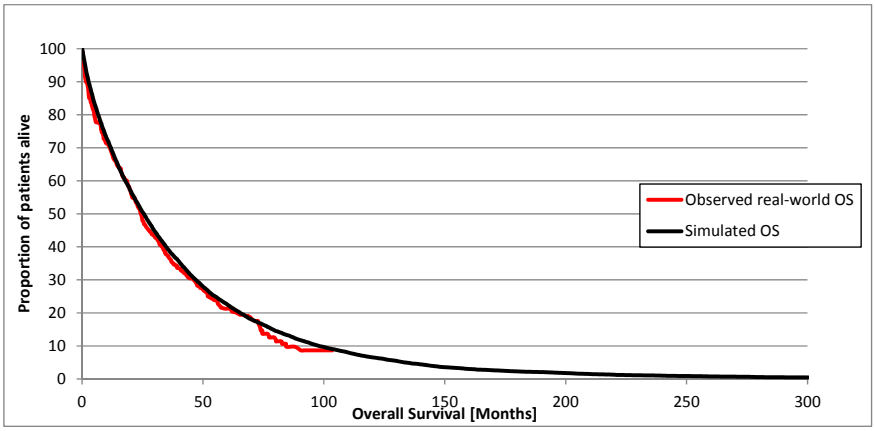
Parametric survival models

After identifying the distribution that best fitted the data (i.e. a Weibull distribution) and potential predictors that could be included, three multivariate regression models were built to estimate time to event for the first, second and third line, respectively. Two logistic models were built for the type of event in the first and second line. The parameters included in the multivariate survival and event models are presented in Figure 7.1. The regression coefficients including standard errors and confidence intervals are included in the Supporting information.

Figure 7.2 Overall survival as observed in real world and according to the model

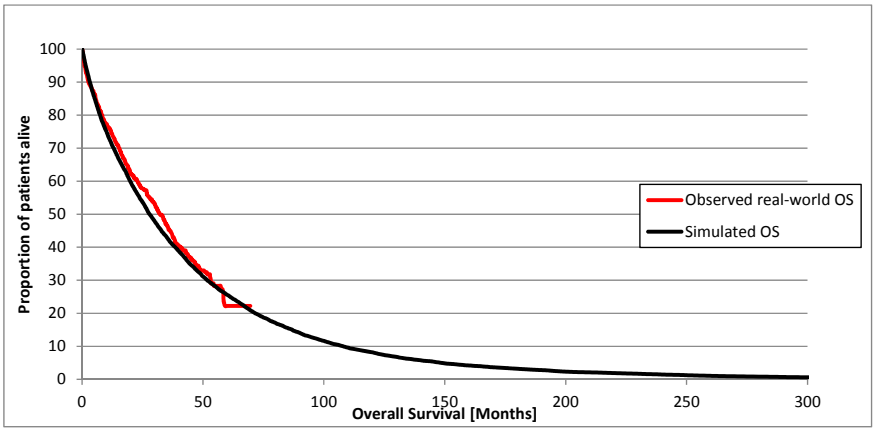
(A) Cohort 1 (first-line treatment between 2004 and 2007)

A



(B) Cohort 2 (first-line treatment between 2008 and 2013)

B



7

Simulation study

Using the results of the clinical analysis (i.e. real-world patient characteristics, treatment patterns, multivariate parametric survival models for the time to event and logistic regression models for the type of event) combined with the PLS model (Figure 7.1), effects and costs of various treatment sequences were calculated. We compared the survival as modelled by our PLS model to the observed data, and Figure 7.2A,B shows similar observed and modelled OS for cohort 1 (i.e. first-line treatment during 2004–2007) and cohort 2 (i.e. first-line treatment between 2008 and 2013) confirming the model's validity. The modelled OS together with the QALYs and costs of real-world treatment

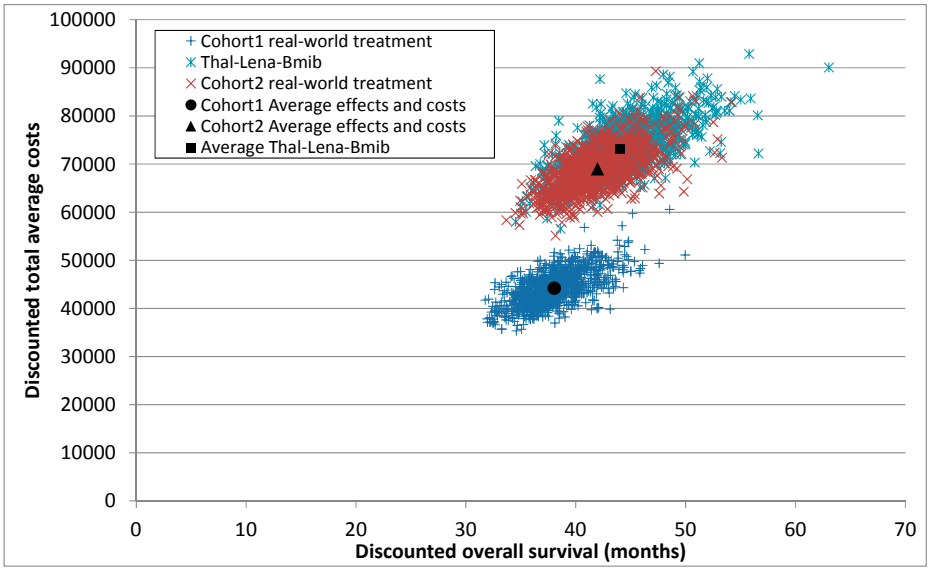
Table 7.3 Median and average OS, Costs and QALYs per treatment sequence discounted and undiscounted

Treatment		Median OS months	Discounted						Undiscounted			
			Average OS months [years]	Average QALYs	Average total costs	Average costs per life year	Incremental costs per life year	Incremental costs per QALY	Average OS months [years]	Average QALYs	Average total costs	Average costs per life year
Real-world	Cohort 1 (2004-2007)	25	38 [3.17]	2.36	€ 44,200	€ 13,934			40 [3.34]	2.48	€ 50,150	€ 15,030
Real-world	Cohort 2 (2008-2013)	28	42 [3.5]	2.6	€ 69,017	€ 19,730			44 [3.7]	2.75	€ 77,931	€ 21,064
Scenario 1	MP-Thal-Bmib	24	36 [2.98]	2.21	€ 38,249	€ 12,837			37 [3.11]	2.31	€ 43,272	€ 13,895
Scenario 2	MP-Thal-Lena	24	38 [3.13]	2.33	€ 47,795	€ 15,263			39 [3.29]	2.44	€ 54,577	€ 16,608
Scenario 3	Thal-Bmib-Lena	29	43 [3.59]	2.66	€ 75,375	€ 21,011			46 [3.79]	2.82	€ 87,162	€ 22,974
Scenario 4	Thal-Lena-Bmib	30	44 [3.67]	2.72	€ 73,202	€ 19,962	€ 24,618	€ 34,875	47 [3.88]	2.88	€ 82,955	€ 21,356

OS denotes Overall survival, QALY denotes Quality-adjusted life-year

(i.e. cohort 1 and cohort 2) and frequently observed treatment sequences (i.e. scenario 1–4) are presented in Table 7.3. Median and average OS for cohort 1 were 25 and 38 months, respectively. Total costs per patient were on average €44,200 and these costs show, from a hospital perspective, the average total costs of the disease per patient from first-line treatment to death. Outcomes for cohort 2 were higher; median OS and average OS were 28 and 42 months, while average total costs were €69,017. Table 7.3 also presents the costs and effects of the four most commonly prescribed treatment sequences in daily practice during our study period. Depending on the treatment sequence, average OS ranged from 36 to 44 months for MP–Thal–Bmib and Thal–Lena–Bmib, respectively. Average total costs from diagnosis till death ranged from €38,249 for MP–Thal–Bmib to €75,375 for Thal–Bmib–Lena. Treatment sequences starting with first-line Thal had highest OS but also the highest average costs. The sequence Thal–Lena–Bmib was the most effective scenario with average OS of 44 months and average total costs of €73,202. Figure 7.3 presents the results of the PSA for cohort 1, cohort 2 and Thal–Lena–Bmib (i.e. the most effective scenario). The model was run 1000 times and effects (OS in months) and costs (discounted total costs in euro's) of each model run are presented as one observation point in Figure 7.3. The spreading of the observations shows the uncertainty around the estimates. Figure 7.3 demonstrates that both effects and costs were higher for cohort 2 compared to cohort 1. Compared to real-world treat-

Figure 7.3 Results of probabilistic sensitivity analysis: total discounted average effects and costs of real-world tretament per cohort and the most-cost-effective treatment sequence



ment patterns in cohort 2, the scenario where all patients would have been treated with Thal–Lena–Bmib yielded better outcomes at higher costs. Average OS increased with 0.17 yr. (0.12 QALYs) and aver- age costs with €4,185. The incremental cost-effectiveness ratio would be €24,618 and €34,875 per life year and QALY gained, respectively.

DISCUSSION

To our knowledge, this is the first study using real-world patient-level data to investigate the cost-effectiveness of complete treatment sequences for elderly real-world patients with MM. A disease track model was designed to calculate cost and effects of daily practice treatment patterns as well as the cost-effectiveness of commonly used treatment sequences, including innovative novel treatments.

Guidelines changed and different treatment patterns were observed between the earlier and the later years of our observation period (i.e. increasing use of novel agents). From 2008 onwards, Thal was recommended as part of first-line treatment followed by Bmib or Lena in second- or third-line treatment regimens. This change was indeed observed in our real-world data, and we illustrated the impact on outcomes by calculating the real-world effects and costs separately for cohort 1 (first-line treatment between 2004 and 2007) and cohort 2 (first-line treatment between 2008 and 2013). While almost 60% of the patients in cohort 1 did not receive first-line treatment with a novel agent, almost 90% did receive a novel agent as first-line treatment in cohort 2. A comparison between these cohorts revealed that both OS (+4 months) and costs (+€24,817) have increased for patients treated according to the real-world treatment patterns as observed during cohort 2. In other words, the shift from MP to Thal induction was clearly followed and proved to be effective. Although we performed an observational study, we are able to relate the improved outcomes to the different treatment patterns observed because we modelled similar patient populations for both cohorts, that is the only change in our model was treatment patterns as observed for cohort 1 and cohort 2. It should be noted that due to limited numbers of patients, we based treatment effectiveness on all patients instead of distinguishing different effectiveness estimates for cohort 1 and cohort 2. We believe this is a conservative approach because effectiveness of treatments might improve if physicians gain more experienced with treatments over time. This might mean that the observed improvement in OS between cohort 1 and cohort 2 might be higher than the 4 months now reported.

In addition to the cohort comparison, the scenario analyses clearly demonstrated improved survival with increasing use of novel treatments. For example, OS was on average 5–8 months better (average 43–44 months versus 36–38 months) if all patients received first-line treatment with thalidomide, a novel agent, instead of only MP. Costs

for patients treated with MP–Thal–Bmib or MP–Thal–Lena were €38,249 and €47,795, respectively. Total costs were €73,202 and €75,375 for Thal–Lena–Bmib and Thal–Bmib–Lena, respectively.

Outcomes of the most effective scenario (Thal-Lena-Bmib) were compared to the real-world treatment patterns as observed for cohort 2. Although novel agents were already prescribed, the comparison revealed higher OS for the scenarios in which Thal-Lena-Bmib would have been the standard treatment approach. In this scenario Lena and Bmib were only assigned to patients who did not die after either first- or second-line treatment. Cost-effectiveness ratios were acceptable, €24,618 per life year gained and €34,875 per QALY.

Although Bmib is recommended as first-line treatment in the Netherlands since 2013[158], we did not study treatment sequences that used Bmib as first-line treatment. PHAROS included patients diagnosed since 2004–2011. Of these patients, only 7% was treated with Bmib in the first-line.

The internal validity of our model was confirmed by the good correspondence with the real-world data with almost overlapping survival curves. Our results confirmed previous findings[135,139,159] that, in general, elderly patients benefit from novel agents and show that OS of real-world patients improves by increasing the proportion of patients treated with novel agents. Survival for treatment sequences is unavailable. However, the modelled OS can be compared to the OS as reported in the literature. Results from a meta-analysis of 1685 patients included in RCTs showed median OS for first-line MP treatment was 32.7 months and for first-line thalidomide treatment 39.3 months[139]. This is above the median OS from our model, 24 and 30 months, respectively. Lower OS was expected given the age and disease characteristics of our real-world patient group. Nevertheless, the improvement in median survival in our study (5 months) is almost as high as the effect of first-line thalidomide as established by the meta-analysis of Fayers et al.[139] (6 months). Median OS of cohort 1 and cohort 2 in our study was below the median OS of 33.6 months for non-high-dose patients as reported by a Swedish population based study[159]. This difference might be related to different prescription of treatments or patient characteristics. For example, our study included a higher proportion of patients with International Staging System III, 38% compared to 25%.

Real-world costs of patients with relapsed/refractory MM were according to Gaultney et al.[77] €72,968 and this is above the average total costs in our study. However, we included newly diagnosed patients instead of only relapsed refractory patients. In addition, the study of Gaultney et al. was more detailed on for example concomitant medication costs. Evidence on the cost-effectiveness of novel agents for MM is scarce[142]. Only one other CEA, conducted from the US perspective, is available. Garrison et al.[143] showed – depending on subsequent treatments – higher or comparable lifetime costs, €46,458 ($63,294, exchange rate $1 = €0.7340) for first-line MP, while costs for first-line

Thal were higher €104,560 ($142,452). The total cost for treatment sequences containing Thal might be higher due to the branded price for thalidomide used by Garrison et al.[143,160]. Information on health-related quality of life in patients with MM treated with novel agents is limited and a comparison between different treatments cannot be made[161]. Utility values, necessary to calculate QALYs, are even scarcer and only available for transplant eligible patients[73]. Therefore, we relied on a small cross-sectional subsample of a real-world patient population and were unable to assign distinctive utility values to the treatment strategies. We assume that all treatments have an equal impact on the quality of life, that is neglecting the differences in toxicities between the treatments. Nevertheless, this is subject for further research.

Real-world cost-effectiveness can only be calculated with real-world data, as these data provide insight into real-world effects. In addition, due to longer follow-up compared to most RCTs, real-world data also provide additional insights into OS. However, the absence of a randomised design is an important limitation of real-world data. This might be problematic for a disease such as MM, in which many factors inherent to the patient and the tumour itself contribute to the final outcome. To correct for imbalances, we included a large number of patient and disease characteristics such as age, laboratory values and disease stage that are considered to influence OS in MM.

Although the PHAROS database provided us with information of 1054 elderly patients with MM, uncertainty around the model estimates was presented by the cost-effectiveness clouds in the sensitivity analysis. Nevertheless, robust estimates were obtained from the sensitivity analysis after 1000 model simulations. Due to the number of patients, we were not able to make a distinction between patients receiving maintenance therapy and patients who did not receive maintenance therapy, and only the hospital visits for maintenance therapy were accounted for. In addition, the number of patients only allowed us to develop a model with three treatment lines and we neglected the drug costs of fourth and subsequent treatment lines. Furthermore, real-world costs were based on the most important cost drivers, that is hospital visits and drug costs. The effectiveness of maintenance therapy and subsequent treatment lines as well as detailed real-world cost analysis including laboratory tests and other procedures remains subject for further research.

This study shows that with the advent of newer treatments that have proven efficacy, real-world elderly patients with MM live longer but at an increased cost. A shift from MP to Thal has clearly been effective. OS improved with 4 months for patients receiving first-line treatment between 2008 and 2013 compared to patients receiving first-line treatment between 2004 and 2007, while average costs per patient increased with €24,817. Although OS was a little lower, efficacy of novel agents as obtained in RCTs was confirmed in Dutch daily clinical practice. By comparing real-world treatment to hypothetical scenarios, we identified that real-world treatment in the Netherlands

could have been improved further by increasing the utilisation of novel agents. The incremental costs were €24,618 per life year gained and €34,875 per QALY if the sequence Thal–Lena–Bmib was standard treatment for all patients. Recommended treatment for MM rapidly changes, and therefore, further research is necessary to identify whether outcomes could be improved in the future as well by increasing use of novel agents in daily clinical practice.

ACKNOWLEDGEMENTS

The authors would like to thank Annemieke Leunis for sharing her experience and expertise in designing a full disease model. The authors are grateful to Simone Oerlemans and Lonneke van de Poll for conducting the cross-sectional quality of life study. Furthermore, the authors are grateful to the various registration teams of the Netherlands Cancer Registry for the data collection. PHAROS, the Population-based HAematological Registry for Observational Studies, is a cooperation of HOVON (Dutch-Belgian Cooperative Trial Group for Hematology Oncology), the institute of Medical Technology Assessment at Erasmus University Rotterdam and the Comprehensive Cancer Organisation, the Netherlands (IKNL). PHAROS is financially supported by ZonMW 2009–2012 (#80-82500-98-01007) and through unrestricted grants from Bayer, Roche, Janssen, Celgene, Sanofi, MSD, Novartis, GSK, Mundipharma and BMS.

Supporting information

Regression coefficients to model time-to-event for line 1

Weibull model

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Constant	3.433	0.544	0.000	2.366	4.500
Age1	-0.018	0.005	0.001	-0.028	-0.007
Sex (Reference category male)	0.195	0.068	0.004	0.062	0.327
WHO status 0	Reference category				
WHO status 1	-0.093	0.079	0.243	-0.249	0.063
WHO status 2	-0.338	0.109	0.002	-0.551	-0.124
WHO status 3	-0.689	0.215	0.002	-1.120	-0.258
WHO status 4	-1.372	0.638	0.049	-2.736	-0.009
Albumine1	0.014	0.005	0.010	0.003	0.024
Haemoglobin1	0.095	0.029	0.001	0.037	0.152
Platelets1	0.001	0.000	0.022	0.000	0.001
Serum calcium1	-0.190	0.093	0.042	-0.374	-0.007
Treatment1 MP	Reference category				
Treatment1 Thal	0.342	0.080	0.000	0.186	0.498
Treatment1 Bmib	0.227	0.150	0.130	-0.067	0.520
Treatment1 Lena	0.441	0.203	0.030	0.043	0.838
Treatment1 Other	-0.401	0.136	0.003	-0.668	-0.133
Comorbidity (Reference category no comorbidity)	-0.215	0.092	0.022	-0.398	-0.032
Included in study (Reference category not included	0.259	0.118	0.028	0.028	0.490
Shape	1.071	0.030		1.014	1.131

Regression coefficients to model type of event line 1

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Constant	-8.024	1.068	0.000	-10.118	-5.930
TTE1 (ln)	-0.694	0.078	0.000	-0.848	-0.541
Age1	0.115	0.014	0.000	0.089	0.142
WHO status 0	Reference category				
WHO status 1	0.392	0.204	0.057	-0.011	0.795
WHO status 2	0.616	0.273	0.025	0.079	1.154
WHO status 3	0.298	0.438	0.496	-0.562	1.159
WHO status 4	0.295	0.820	0.720	-1.318	1.907

Regression coefficients to model type of event line 1 (continued)

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Treatment1 MP	Reference category				
Treatment1 Thal	0.240	0.192	0.210	-0.135	0.616
Treatment1 Bmib	0.814	0.363	0.025	0.101	1.526
Treatment1 Lena	0.407	0.433	0.347	-0.441	1.255
Treatment1 Other	0.570	0.337	0.090	-0.089	1.230

Regression coefficients to model time-to-event for line 2

Weibull model

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Constant	1.009	0.428	0.019	0.166	1.853
TTE1	0.008	0.004	0.045	0.000	0.016
Albumine2	0.036	0.010	0.000	0.017	0.055
Creatinine2	-0.002	0.001	0.001	-0.003	-0.001
Haemoglobin2	0.095	0.043	0.028	0.010	0.180
LDH2	-0.001	0.000	0.000	-0.002	-0.001
Treatment2 MP	Reference category				
Treatment2 Thal	0.267	0.152	0.080	-0.032	0.566
Treatment2 Bmib	0.025	0.154	0.873	-0.278	0.327
Treatment2 Lena	0.310	0.179	0.083	-0.040	0.660
Treatment2 Other	0.028	0.211	0.896	-0.387	0.442
Shape	1.069	0.042		0.989	1.156

Regression coefficients to model type of event line 2

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Constant	-4.429	1.789	0.013	-7.935	-0.923
TTE2 (ln)	-0.620	0.117	0.000	-0.850	-0.390
TTE1	-0.024	0.011	0.024	-0.046	-0.003
Age2	0.075	0.021	0.000	0.034	0.116
WHO status 0	Reference category				
WHO status 1	0.551	0.267	0.040	0.026	1.075
WHO status 2	0.534	0.415	0.198	-0.281	1.348
WHO status 3-4	0.207	0.647	0.749	-1.069	1.483
Haemoglobin2	-0.201	0.105	0.056	-0.408	0.005
Treatment2 MP	Reference category				
Treatment2 Thal	1.128	0.401	0.005	0.342	1.913

Regression coefficients to model type of event line 2 (continued)

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Treatment2 Bmib	1.291	0.406	0.001	0.496	2.086
Treatment2 Lena	1.758	0.456	0.000	0.863	2.652
Treatment2 Other	1.051	0.538	0.051	-0.003	2.105

Regression coefficients to model time-to-event (death) for line 3

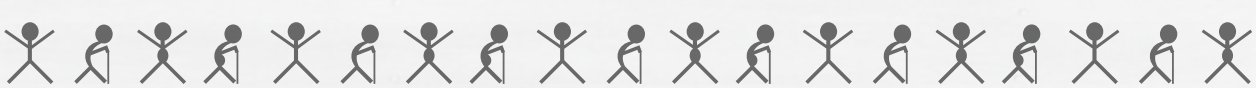
Weibull model

Variable	Coefficient	SE	p-value	[95% Conf.	Interval]
Constant	0.595	0.728	0.414	-0.836	2.026
TTE2 (ln)	0.286	0.104	0.006	0.083	0.490
Albumine3	0.042	0.017	0.013	0.009	0.076
Creatinine3	-0.004	0.001	0.000	-0.006	-0.002
Platelets3	0.003	0.001	0.004	0.001	0.004
Treatment3 MP		Reference category			
Treatment3 Thal	0.517	0.358	0.148	-0.184	1.218
Treatment3 Bmib	0.178	0.343	0.603	-0.493	0.850
Treatment3 Lena	0.391	0.349	0.263	-0.293	1.075
Treatment3 Other	0.007	0.417	0.987	-0.810	0.824
Shape	1.033	0.067		0.909	1.174

Number of patients receiving first, second and third-line treatment per scenario*

	Line 1	Line 2	Line 3
Scenario1	Treatment MP	Treatment Thal	Treatment Bmib
MP-Thal-Bmib	N=1054	N=658	N=409
Scenario2	Treatment MP	Treatment Thal	Treatment Lena
MP-Thal-Lena	N=1054	N=658	N=410
Scenario3	Treatment Thal	Treatment Bmib	Treatment Lena
Thal-Bmib-Lena	N=1054	N=659	N=394
Scenario4	Treatment Thal	Treatment Lena	Treatment Bmib
Thal-Lena-Bmib	N=1054	N=659	N=358

* Based on 1000 simulations

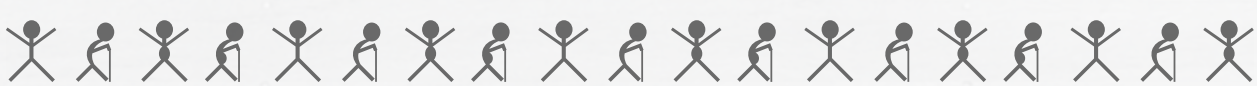


Chapter 8

Real-world cost-effectiveness of sequential treatments: a practical guide on the construction of a full disease model

Blommestein HM*, de Groot S*, Redekop WK, Uyl-de Groot CA
* Both authors contributed equally to the manuscript

Submitted



ABSTRACT

Objective: Although models comprising multiple treatment lines are increasing in importance, experience is scarce and while guidelines for good modelling practices are available, they are sometimes too brief to help researchers develop models that are valid and credible. This study provides practical recommendations in constructing a discrete event simulation (DES) model to support real-world cost-effectiveness analyses of treatment strategies spanning multiple treatment lines.

Methods: Based on experiences with two DES models used in cost-effectiveness analyses of treatment strategies in cancer, we discuss how best practices, mainly derived from the ISPOR-SMDM Task Force, can best be implemented. Additional recommendations were provided wherever best practices were unavailable or not applicable.

Results: The following steps in constructing DES models were considered particularly important: assigning patient characteristics when simulating a patient population, estimating time-to-event (i.e. next treatment line or death) and event type for each patient, assigning costs and outcomes to all treatment lines, discounting, and conducting sensitivity analyses. Modelling multiple treatment lines using a DES model and real-world data imposes several challenges. First, it is necessary to correct effectiveness and costs for patient characteristics (including previous treatments). This could be addressed by including patient characteristics and effectiveness of previous treatments in the survival models. Second, when modelling a heterogeneous population, valid extrapolation of survival outcomes beyond observation is required. This could be achieved by using multiple survival models per treatment line. Third, the timing of competing events needs to be estimated appropriately. As recommended by the Task Force, one single survival model should be used together with a regression technique to determine event type.

Conclusions: Developing good-quality models comprising multiple treatment lines requires guidance beyond the existing guidelines and practical recommendations are currently lacking. The guidance based on hands-on experience with two DES models can improve validity and credibility of future disease models and cost-effectiveness analyses.

INTRODUCTION

As more treatments become available (within and beyond treatment lines), traditional economic evaluations may not provide sufficient information, since these do not assess costs and effects of treatment strategies spanning multiple treatment lines and are not able to determine the optimal order (i.e., sequence) in which treatments should be provided.

As a consequence, full disease models comprising multiple treatment lines are expected to increase in importance, but experience is scarce. Tosh and colleagues[162] called for a methodological framework for economic evaluations of sequential therapy for chronic conditions, since they found that methods have not been consistently applied, which has led to varied estimates of cost-effectiveness and uncertainty in respect of the most appropriate analytic methods. Although guidelines for good modelling practices are available including the series commissioned by the ISPOR-SMDM Task Force[163], they are sometimes too brief to help researchers develop models that are valid and credible.

This study provides practical recommendations in constructing a discrete event simulation (DES) model to support cost-effectiveness analyses of treatment strategies spanning multiple treatment lines. Best practices derived from the ISPOR-SMDM Task Force and additional sources are cited, followed by a description of how these were implemented in our DES models to estimate the real-world cost-effectiveness of new treatments in metastatic renal cell carcinoma (mRCC) and multiple myeloma (MM).

CASE STUDIES

Metastatic Renal Cell Carcinoma (mRCC)

115,200 patients were diagnosed with kidney cancer in Europe in 2012[164]. Renal cell carcinoma represents 80% of all kidney cancers. Median overall survival (OS) of patients with advanced disease is 43, 27 and 8.8 months for patients with a favourable, intermediate or poor prognosis, respectively[165]. Health outcomes are influenced by prognostic factors[166].

A number of first- and second-line targeted therapies (e.g. sunitinib, sorafenib and everolimus) for mRCC have been introduced since 2006[165]. These therapies improve health outcomes, such as progression-free survival (PFS) and OS[167-172]. However, a Dutch population-based registry showed that almost half of the patients presenting with mRCC did not receive any targeted therapy[173]. A DES model was developed to study the real-world cost-effectiveness of several treatment strategies applied in patients with mRCC comprising one or more sequentially administered drugs. Potential health out-

comes and costs of hypothetical treatment scenarios were calculated by assuming that all treatment-eligible patients were treated according to a particular treatment strategy.

Multiple myeloma (MM)

In 2012, 38,900 patients were diagnosed with MM in Europe[164]. MM is a heterogeneous disease with a wide variation in OS[131,132]. Depending on the stage of the disease, median OS ranges from 29-62 months[174].

Like for many cancers, treatment of MM is characterised by sequential treatment lines aiming to prolong PFS and OS. In the past decade, several treatment options have become available including the thalidomide-, bortezomib- and lenalidomide-based regimens. While most of these treatments were first recommended as treatment for third or subsequent lines, they are now recommended as induction therapy. Health outcomes are also influenced by prognostic factors, mainly patient and disease characteristics[136,137]. A DES model was developed to study the real-world cost-effectiveness of sequential use of novel agents for elderly MM patients[148]. Furthermore, by studying treatment sequences, we aimed to identify the optimal treatment strategy.

Comprehensive data on patient and disease characteristics of patients with mRCC and MM, as well as data on treatments and outcomes were collected in two population-based registries, the mRCC registry (PERCEPTION) and the MM registry (PHAROS)[21,151].

MODEL STRUCTURE AND DESIGN

Best practice:

> *"If, (…), a valid representation of any aspect of the decision problem would lead to an unmanageable number of states, then an individual-level state-transition model is recommended."*[176]

> *"DES is an attractive option in nonconstrained models (…) when individual pathways through the model are influenced by multiple characteristics of the entity; and when recording individual entity experience is desirable."*[175]

The first stage in developing a decision model involves choosing an appropriate model structure. According to the best practice commissioned by the ISPOR-SMDM Task Force, DES is the preferred modelling method if it is difficult to model the disease course of the average patient, and when the course of the disease, including its treatment, would require too many health states. As stated in the previous paragraph, patients with mRCC

and MM in daily practice represent a heterogeneous population, and characteristics of these patients have a large impact on the costs and effects of treatment. In order to incorporate individual patients and allow for variability between patients, a DES model was developed to calculate the cost-effectiveness of various treatment scenarios as recommended by the ISPOR-SMDM Task Force[175,176]. DES models allow individual patients to have their own characteristics, such as age and health state, which may also change over time[177].

Furthermore, treatment of both mRCC and MM is characterised by sequentially administered drugs. Instead of modelling single treatment options, a comparison of complete treatment strategies was needed. DES models can easily include the effect of previous therapies, in contrast to Markov models, which cannot incorporate history of patients without constructing a large amount of health states. Therefore, a DES model seemed a better choice for modelling treatment strategies spanning multiple treatment lines for mRCC and MM[176]. Caro et al.[163] also argued that a DES provides an alternative, more natural, way to simulate clinical reality, whereas a Markov model requires all aspects of a disease including patient and disease characteristics and treatment history to be captured in a health state. Although various methods exist which can include memory in Markov models (e.g. tracker variables), the required number of tracking variables would have been quite large in a model of sequential therapies.

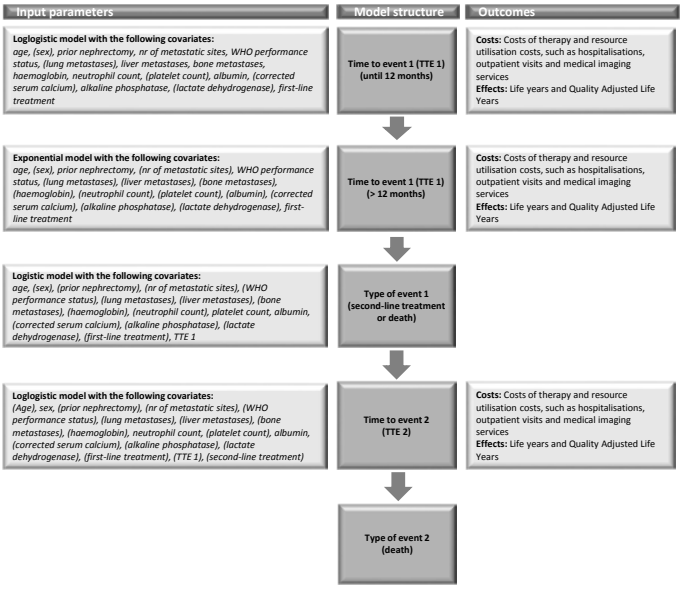
In addition, data from the mRCC registry and MM registry revealed that some patients died very soon after treatment was initiated while some patients survived much longer. In a micro simulation Markov model, patients can only experience one transition per cycle and this would require many cycles with a small cycle length, which favoured a DES model allowing to include time continuously.

The DES models for mRCC and MM comprised entities (i.e. patients), attributes assigned to the entities, and events. Attributes were obtained from patient-level data from either the mRCC or MM registry by selecting clinical factors, biochemical and haematological factors known to impact mRCC or MM outcomes, respectively. Events were either second-line treatment, third-line treatment (in the MM model only) or death. The time horizon of the models spanned the patients' lifetime. The structures of the mRCC and MM model are presented in Figure 8.1. Characteristics and sources for input parameters of both models are presented in Table 8.1.

Figure 8.1 Model structure of the full disease models spanning multiple treatment lines

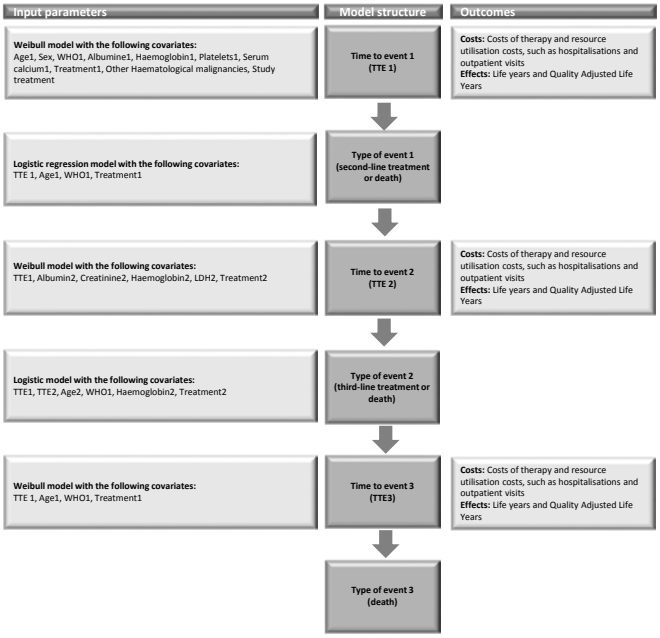
a) mRCC model

A



b) MM model

B



NOTE. Covariates between brackets were considered for inclusion in the survival models and logistic regression model, but excluded through backward and/or forward selection.

Table 8.1. Model characteristics and sources for input parameters of the DES models

	Metastatic renal cell carcinoma	Multiple myeloma
Model characteristics		
Aim	Model real-world cost-effectiveness for patients with metastatic renal cell carcinoma	Model real-world cost-effectiveness for elderly patients with multiple myeloma
Perspective	Health care	Health care
Patients	Patients with metastatic renal cell carcinoma	Elderly patients with multiple myeloma
Outcomes	Effects (OS and QALYs) and costs (€)	Effects (OS and QALYs) and costs (€)
Model type	Discrete event simulation	Discrete event simulation
Time horizon	Lifetime	Lifetime
Parametric distribution	Loglogistic and exponential distribution (line one) and loglogistic distribution (line two)	Weibull distribution for all lines
Disease pathways (Base case)	Real-world treatment including two subsequent lines	Real-world treatment including three subsequent lines
Disease pathways (scenarios)	Hypothetical pathways including two lines of treatment	Hypothetical pathways including three lines of treatment
	No targeted therapy	MP-thalidomide-bortezomib
	Sunitinib - Sorafenib	MP-thalidomide-lenalidomide
	Sunitinib - Everolimus	Thalidomide-bortezomib-lenalidomide
	Sunitinib - Other	Thalidomide-lenalidomide-bortezomib
Sensitivity analysis	Univariate and probabilistic sensitivity analyses (1000 simulations)	Univariate and probabilistic sensitivity analyses (1000 simulations)
Sources for input parameters		
Data (Patient and disease characteristics, treatment effects and patterns, health care utilisation)	Real-world data from the mRCC registry (PERCEPTION)	Real-world data from the MM registry (PHAROS)
Unit prices	Dutch reference price lists and literature	Dutch reference price lists and literature
Discount rates	Dutch guidelines (effects 1.5%, costs 4%)	Dutch guidelines (effects 1.5%, costs 4%)
Utilities	Literature	Cross-sectional study

OS: overall survival QALYs: Quality-adjusted life years MP: melphalan prednison

TIME-TO-EVENT

Best practice:

> *"It is (...) very important to justify the particular extrapolation approach chosen, to demonstrate that extrapolation has been undertaken appropriately and so that decision makers can be confident in the results of the associated economic analysis."*[178]

As survival data is often not fully observed, extrapolation beyond the observation period is needed. The method to extrapolate this data should be chosen in a systematic way in order to ensure valid and clinical plausible extrapolation. Time-to-event data derived from either the mRCC or MM registry were extrapolated using a range of parametric models (Exponential, Weibull, Log-logistic, Log-normal, Gamma and Gompertz). These models were assessed for their goodness of fit to the data using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Furthermore, each parametric function was assessed graphically as recommended by Latimer[178].

For the mRCC model, a loglogistic distribution best fitted the time to the first event (TTE1) and the time to the second event (TTE2). Nevertheless, visual inspection showed that TTE1 was underestimated after 12 months. Additionally, as a consequence of the functional form of this distribution, mean TTE1 was highly influenced by a small proportion of the population with very long TTE1 estimates. Therefore, an alternative model (i.e. exponential) was chosen for TTE1 after 12 months, based on the AIC and BIC. This approach was also conducted by Leunis et al., who specified different survival models for different time periods[179]. For the MM model, a Weibull distribution best fitted the time to the first, second and third event. A Weibull distribution had the best goodness of fit based on the AIC/BIC and was also considered appropriate based on visual inspection.

COMPETING EVENTS

Best practice:

> *"Where feasible, when estimating times to competing events, methods of analysis that estimate the timing of competing events jointly are preferred to approaches that estimate separate time to event curves for each event."* [178]

In survival analysis, competing events are present when an individual is at risk of several different types of events but can have only one event at a time[150,175]. For the mRCC and MM models, time to next treatment was calculated from patient level data. Since some patients died before a new treatment was initiated, next treatment and death were competing events. Generally, there are two approaches to analyse competing events[175]. The first approach is to perform separate survival analysis for each event where the other event is treated as censored. Then, for each event a time is sampled, with the patient moving to the event with the shortest time. The second approach is to perform one single survival analysis but make no distinction between the competing events; a separate sampling process in the model determines which event a patient will experience.

The second approach was adopted in the mRCC and MM models, as recommended by the ISPOR-SMDM Modelling Good Research Practices Task Force[175]. Whereas survival analysis assumes that censoring is non-informative, we hypothesised that death was mostly treatment-related and, as a consequence, that censoring the patients who died might have altered the probability of experiencing a next treatment. Furthermore, the graphic presentation and interpretation of the single survival analysis are straightforward whereas the interpretation of the two separate survival curves is less intuitive. While the second approach was adopted in the mRCC and MM models, we applied the first approach to validate our results and hypothesis. Interestingly, both methods yielded very similar results in the models. Since there is no difference between the two methods in terms of ease or speed (i.e. both methods require the estimation of two statistical models), we decided to align with current guidelines.

ASSIGNING PATIENT AND DISEASE CHARACTERISTICS

Best practice:

> *"The expected costs and benefits across the sampled group [..] provide an unbiased estimate provided that a sufficiently large sample is simulated and any covariance between the different patient characteristics is correctly taken into account."*[180]

While the mRCC and MM registries provided patient level data, simulation of the population including patient and disease characteristics was needed to study what would have happened to a patient if that patient had been treated differently. No recommendations were made by the ISPOR-SMDM Task Force about simulating a population. However,

Davis et al. emphasises the need to account for the covariance between patient and disease characteristics[180].

In the mRCC and MM models, patient and disease characteristics were simulated similarly; random numbers were drawn from predefined distributions. These distributions were derived from patient-level data from the mRCC registry or the MM registry, respectively.

To account for the covariance between characteristics, different distributions were used for patients with a different prognosis. For example in the mRCC model, first WHO performance status before treatment was simulated by drawing random numbers from a predefined distribution as obtained from the PERCEPTION registry. In simulating additional patient and disease characteristics; for each characteristic, a different distribution was used for patients with a WHO performance status of 0-1, and for patients with a WHO performance status of 2-4. This method was adopted to increase the likelihood that the combination of patient and disease characteristics per individual matched the original data.

Besides patient and disease characteristics, treatment needed to be assigned to each patient. Two multinomial logistic regression models (i.e. one to assign first-line treatment and one to assign second-line treatment) were used, including patient and disease characteristics as well as treatment history as covariates, to assign real-world treatment patterns to the patients in the mRCC model. This process guaranteed that the patients who received the treatments in the model were similar to the patients who received these treatments in daily clinical practice. This method was not feasible in the MM model, since some novel agents (i.e. bortezomib and lenalidomide) were prescribed to very few patients during the follow-up period. These numbers were too small to run a multinomial logistic regression model. Treatment in the MM model was therefore simulated in the same way as patient and disease characteristics; the probability of receiving a certain treatment was based on the distribution of treatments as observed in daily clinical practice using different distributions for patients with a different WHO performance status. Having simulated patient and disease characteristics, and treatment for all patients, the patient's time to an event (either TTE1, TTE2 and TTE3) was estimated taking these characteristics into account.

ACCOUNTING FOR PREVIOUS THERAPIES

The mRCC and MM model aimed to calculate the cost-effectiveness of several treatment strategies comprising one or more sequentially administered treatments. Therefore, it was important to correct for the effectiveness of previous treatments when estimating the effectiveness of second- and third-line treatments. In addition, the effectiveness of

subsequent therapies should be taken into account when estimating overall survival of first- and second-line treatments. Ideally, the effectiveness of, for example, bortezomib after thalidomide is solely based on patients treated with thalidomide followed by treatment with bortezomib. Although the registries included a substantial number of patients, these were not adequate to provide a stable estimate of the effectiveness of all treatment sequences. Recommendations to account for previous therapies in full disease models do not exist.

Therefore, we chose to correct for the effectiveness of previous therapies by including the TTE of the previous line in estimating the TTE of the subsequent therapy. For example, TTE1 was included in the parametric survival model estimating TTE2. This allowed us to obtain the effectiveness of second-line treatment accounting for the effectiveness of first-line treatment given the patient's characteristics. In the MM model TTE1 had a significant association with TTE2 as well as with the type of event. The coefficient corresponding to TTE1 was not treatment specific, i.e. a TTE1 of 2 months obtained by treatment with thalidomide is similar to a TTE1 of 2 months obtained by treatment with bortezomib. Since adding type of treatment to the model did not improve it's explanatory value, we believe this method can be used to correct for the effectiveness of previous therapies.

COSTS AND OUTCOMES

Best practice:

> *"Costs and quality of life weights are attached to events and time spent with different health conditions to estimate long term costs and health outcomes."*[175]

According to this guideline, health outcomes and costs depend on events and time and this practice was adopted in both models. Total life years (i.e., OS) were calculated by summing TTE1, TTE2 (and TTE3). Besides total life years, total quality adjusted life years (QALYs) were calculated by weighting LYs for the quality of life during these years using utility weights. In the mRCC model, various utility weights were used for patients with a favourable or intermediate prognosis, and patients with a poor prognosis before either first-line therapy or second-line therapy since their quality of life was expected to differ. Treatment-specific (including the effect of adverse events) or utility weights for different risk groups (or disease stages) were unavailable for elderly patients with MM. Therefore, an average utility weight was used in the MM model, obtained from a Dutch population-based cross-sectional study in MM.

Based on real-world data, average treatment-specific resource use per month was obtained in order to calculate total costs per month. For example the number of outpatient

visits per month for mRCC patients treated with sunitinib or the number of hospital days per month for MM patients treated with thalidomide. Average total costs per patient were calculated by multiplying treatment specific total costs per month with TTE.

DISCOUNTING

Best practice

> *"A (common) real discount rate should be applied to future costs and, when used in a cost-effectiveness analysis, to future outcomes."*[184]

> *"Discounting methods should accord with general guidelines for economic evaluation."*[185]

Future effects and costs should be converted to their present value in order to account for factors such as time preferences and uncertainty. As recommended, future costs and effects in the mRCC and MM model were discounted to their present value using discount rates based on the Dutch guideline for pharmacoeconomic research[181].

While a Markov model with a fixed cycle length provides a convenient structure to discount future costs and effects, discounting future costs and effects in a DES model including treatment strategies comprising one or more sequentially administered treatments is more challenging. First, a DES model produces individual TTE estimates, and as a consequence LYs and QALYs need to be discounted for each patient separately. Furthermore, different utility values were assigned to the treatment lines in the mRCC model and therefore, total QALYs needed to be discounted for each treatment line separately. Second, unit costs per month differed between treatment lines. For example, a patient with MM treated with melphalan-prednisone, followed by a bortezomib-based regimen, and then followed by a lenalidomide-based regimen, incurs different hospital and drug costs per month during first-, second- and third-line treatment. As a consequence, total costs need to be discounted for each treatment line separately.

Since total costs per treatment line were obtained by multiplying unit costs per month by the corresponding TTE, it was decided to discount TTE and multiply unit costs per month by the discounted TTE. The same approach was adopted to discount future QALYs in the mRCC model. While discounting time was a convenient approach in our models, this is not possible for DES models where costs are obtained from multivariable regression models. Since these models include undiscounted time as an explanatory variable, it is not possible to calculate and discount the total costs per line. Instead,

costs should be calculated and discounted for different time frames, e.g. per year, which adds both complexity and computational burden to the model.

PROBABILISTIC SENSITIVITY ANALYSIS

Best practice

> *"The inner loop evaluates the outcomes across the simulated population for the given parameter values, and the outer loop samples those parameter values to reflect uncertainty in the model inputs. In a cohort-level model, only the outer loop is required, thus PSA computation time for a cohort-level model is likely to be lower than for an equivalent patient-level model."*[180]

In other words, the inner loop aims to calculate costs and effects for one simulated population (with constant patient and disease characteristics), whereas the outer loop changes all input parameters according to their probability distributions to examine the impact of the joint uncertainty across all input parameters. In the model for mRCC and MM, the values of input parameters varied across simulations. Due to the probabilistic structure of the models, the values of input parameters could also vary within one simulation. For example, an inpatient day could cost €402 while calculating costs of treatment scenario A and €646 while calculating costs of treatment scenario B. Furthermore, patient and disease characteristics could vary within one simulation. For instance, 26% of the population could be assigned a WHO performance status of 2-4 while calculating costs and effects of treatment scenario A and 35% could be assigned a WHO performance status of 2-4 while calculating costs and effects of treatment scenario B. However, in each single simulation, parameters that are not related to a certain treatment scenario should have the same values in all treatment scenarios. This approach reduces the 'noise' or random variation that is introduced by setting unit costs and patient and disease characteristics twice in each simulation, once for Scenario A and once again for scenario B. It also increases the model's efficiency, since fewer simulations are needed to get a stable estimate of the ICER.

Apart from probabilistic sensitivity analysis, univariate sensitivity analyses can be performed to examine the impact of alternative input parameters on the incremental cost-effectiveness ratios (ICERs) as illustrated in the mRCC model.

DISCUSSION

Economic evaluations mostly require a lifetime time horizon in order to capture all health and economic consequences[4]. Such a time horizon makes a full disease model including treatment strategies spanning multiple treatment lines inevitable. However, the development of the full disease models for mRCC and MM revealed several challenges, including the optimal ways to correct effectiveness and costs for patient characteristics (including the effectiveness of previous treatments), extrapolate survival outcomes beyond observation for a heterogeneous patient population, and estimate the timing of competing events. Best practices, including solutions to these challenges, were not always found in the literature. Also, existing disease models did not often provide suitable solutions to these challenges since these models differed in aim, characteristics of the disease or treatment varied and comprehensive data was unavailable. Therefore, guidance beyond the existing guidelines and practical recommendations are necessary to improve validity and credibility of future disease models.

Based on hands on experiences with two DES models the following recommendations can be made in constructing a DES model to support real-world cost-effectiveness analyses of treatment strategies spanning multiple treatment lines.

First, the inclusion of patient characteristics (including the effectiveness of previous treatments) as covariates in survival models, makes it possible to derive more valid estimates of costs and effectiveness. Second, using multiple survival models per treatment line ensures valid extrapolation of survival outcomes beyond observation for a heterogeneous population. Third, as recommended by the ISPOR-SMDM Task Force, when competing events exist, one single survival model should be used together with a regression technique to determine which event type will occur.

Although the mRCC and MM models enabled the estimation of the cost-effectiveness of several treatment strategies comprising one or more sequentially administered drugs,[148,173] these models could have been improved further. Based on our experiences, the following recommendations can be made to improve future models of treatment strategies spanning multiple treatment lines. First, in our models, patient and disease characteristics were simulated by drawing random numbers from predefined distributions. Covariance between characteristics was accounted for by using different distributions for patients with a different prognosis. This method does, however, not guarantee valid relationships between all patient and disease characteristics. Multivariable regression models could overcome this problem as illustrated by Goossens et al. in a study on propensity score matching[182]. This method generated patient and disease characteristics based on a set of regression models; one for each characteristic. While the first characteristic was defined using a predefined distribution, all other characteristics were simulated using regression models that included as covariates all of the characteristics

already assigned to the patient; this preserved the covariance between the different characteristics that was observed in the original data.

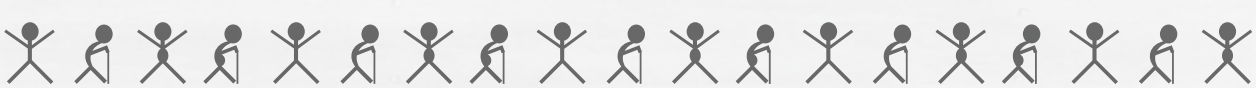
Second, in the mRCC and MM models, total costs per patient were derived by multiplying mean monthly costs (per treatment) by the individual patient's time to an event. However, the distribution of cost data is skewed which means that a limited number of patients is responsible for a high proportion of the costs[4]. As a consequence, by multiplying mean monthly costs by the individual patient's time to an event, total costs per patient might be overestimated. Again multivariable regression models could solve this problem. Besides type of treatment and time to event, these models could include patient and disease characteristics as covariates to estimate total costs per patient. In this way, overestimation of costs will be prevented.

This practical guide is a first attempt to document how best practices in modelling, derived from the ISPOR-SMDM Task Force and additional sources, can be interpreted and implemented. Experiences in implementing best practices were based on two studies only; these studies had rather similar aims, they both focussed on treatment strategies in cancer, and the available data was comparable. Although, treatment strategies spanning multiple treatment lines are common in other disease areas (e.g., rheumatoid arthritis), we recommend further research to be done to ascertain whether this practical guide helps others with different goals working in other disease areas with different data sources. We therefore recommend them to share their findings from constructing and using a DES model including treatment strategies spanning multiple treatment lines.

It should be clear that a DES model is not necessarily the best choice when constructing a full disease model. While a DES model was a feasible option to study the cost-effectiveness of several treatment strategies comprising one or more sequentially administered drugs for patients with mRCC and MM, disadvantages of this model structure may include the type and amount of required data as well as the time needed for model building and simulation[183]. Data for the mRCC and MM model were derived from population based registries. In these registries comprehensive data were collected on patient and disease characteristics. In addition, compared to randomised trials, both registries had a long follow-up duration. This enabled us to study the impact of multiple treatment lines on overall survival. If comprehensive data is not available, a different model structure might be more appropriate. Additionally, the time needed for model building and simulation should be balanced against the benefits of modelling patients individually in a DES model.

CONCLUSION

In order to secure the validity and credibility of models, guidelines were developed summarising best practices in modelling[175,176,178,180,184,185]. Unfortunately, these guidelines are sometimes too brief to be used in constructing full disease models comprising multiple treatment line. Extra instruction is therefore needed. This study aimed to help filling this gap by providing practical guidance on constructing a DES model. Specifically, it explores how to apply the guidelines by describing how they were actually implemented in two DES models, and it provides additional recommendations which may help to further improve the validity of full disease models.



Chapter 9

Discussion

BACKGROUND

Health care expenditures are increasing[1] and trigger intense debate regarding the extent to which new developments should be publicly reimbursed with scarce (public) resources[4]. Health technology assessment (HTA) examines the consequences of adopting or using a new technology in a transparent manner by examining clinical, organisational, economic and patient-related aspects[6]. From the beginning, HTA was based on evidence from randomised controlled trials[186] (RCTs) which are conducted to establish the efficacy and safety of treatments. However, results from RCTs are often not generalisable to patients treated in daily clinical practice for many reasons, e.g. due to differences in patient characteristics and differences in the context of health care delivery. Therefore, there is globally an increasing interest in real-world data[7,12,187]. This thesis evaluated the added value of real-world data for health care decision makers regarding organisational, clinical and economic aspects of HTA using haematological malignancies as a case study.

In the introduction of this thesis, three research questions (RQs) have been formulated;

1. What are the shortcomings of data from RCTs for health care decision makers?
2. What is the added value of real-world data for health care decision makers regarding
   A) organisational aspects of health technology assessment?
   B) clinical aspects of health technology assessment?
   C) economic aspects of health technology assessment?
3. What are the methodological challenges for using real-world data to inform health care decision makers?

This chapter discusses the main findings, evaluates, the impact of real-world data on health outcomes and describes recommendations for further research.

RQ1 WHAT ARE THE SHORTCOMINGS OF DATA FROM RCTS FOR HEALTH CARE DECISION MAKERS?

RCTs aim to establish the efficacy and safety of treatments compared to placebo or another active treatment. To estimate efficacy most accurately, bias should be reduced as much as possible. Treatment is randomly assigned to ensure similarity between the groups for observed and unobserved patient and disease characteristics. Strict in- and exclusion criteria are used and the circumstances of care delivery are controlled by stringent treatment protocols to reduce the chances that contextual factors influence efficacy and safety estimates. The strict criteria and controlled circumstances increase

the likeliness of finding an unbiased treatment effect. For that reason, RCTs are considered the golden standard and the rigidity of RCT protocols ensure high internal validity. However, this also leads to severe limitations of the external validity of the information regarding organisational, clinical and economic aspects of HTA for health care decision makers.

Organisational aspects

Organisational aspects of HTA include the uptake, accessibility and utilisation of new treatments (see Figure 1.1, Chapter 1). RCT data are unable to provide generalisable insights in organisational aspects of new treatments. The uptake includes whether a technology is adopted by physicians in daily practice and accessibility refers to availability for patients who may benefit from the drug. Utilisation includes how many patients received the drug and by which treatment regimens and in what time schedule. The uptake, accessibility and utilisation of a technology in daily clinical practice is, however, influenced by characteristics of the patient, technology and physicians, and by the context of care delivery. For example, recommendations from the professional group and financing conditions set by the government can have a major impact on organisational aspects. In reverse, the same characteristics that enable a description of organisational aspects of HTA (e.g. the context of care delivery) are potential sources of bias for finding a treatment effect and are therefore eliminated in RCTs. As RCTs lack data regarding organisational aspects of HTA, information cannot be derived on the uptake, accessibility and utilisation in daily practice.

Clinical aspects

Clinical aspects of HTA include amongst others efficacy, effectiveness and safety. RCTs compare a new treatment to, usually, one other treatment and aim to ensure internal validity. Therefore, RCTs are seen as the golden standard for demonstrating the efficacy of a treatment. However, the generalisability (external validity) of RCT results is hampered. Due to the strict inclusion criteria, patients included in the RCT are not likely representative for patients treated in daily clinical practice. When real-world patients are different, RCT data do not reflect the true effectiveness of a technology in daily practice. Furthermore, the efficacy of a treatment is most often not measured by overall survival but by using intermediate outcomes such as laboratory or biomedical end-points (e.g. progression-free survival). In the study described in Chapter 2 efficacy was measured by the proportion of patients that became transfusion independent after treatment. Impact on final outcomes (i.e. OS) could not be obtained due to cross-over, the small number of patients included and the short follow-up. Due to the aim of RCTs, follow-up is often limited especially for patients who discontinue treatment. As a consequence, impact on overall survival is often difficult to establish from RCT data. Moreover, as

follow-up is shorter, (rare) long-term adverse events cannot be observed[187]. Limited follow-up is especially a problem for diseases which have a large likelihood to progress (e.g. haematological malignancies). Furthermore, RCTs are designed to compare treatments during a limited period of time and are not designed to estimate the effectiveness of treatment sequences including multiple lines of treatment.

RCTs compare a new treatment to usually one and occasionally two other treatments. Chapter 2 and Chapter 5 illustrated that placebo and best supportive care can be the most relevant comparators in case there is no (other) active treatment option available. However, this is not necessarily the case and treatment in daily practice is often more diverse. Since the number of comparator treatments that can be included in an RCT is limited, the comparator in the RCT may, therefore, not reflect treatment in daily practice, especially when treatment developments are rapidly evolving.

Economic aspects

Performing economic analyses, including cost-effectiveness analyses (see Figure 1.1, Chapter 1), is not the main goal of most RCTs. Although piggyback analyses alongside RCTs are conducted, most RCTs do not collect health resource use data. Because findings from RCT data are not sufficiently generalisable to real-world patient populations, RCT data do not provide the information for studying the economic aspects of HTA. If the controlled circumstances from RCTs differ from treatment in daily practice, cost calculations based on resource use in RCTs are not representative. For example, due to the strict inclusion criteria, patients in RCTs may be fit and therefore have on average lower resource use (i.e. less hospital admissions) than less or unfit patients. As a consequence, RCT data may underestimate costs. In contrast, it is also possible that RCT data overestimate costs for instance due to extensive monitoring or diagnostic testing.

Our study showed that governmental agencies are often uncertain about the costs in daily practice in case only RCT data are available. Chapter 2 illustrates that the ICER was significantly influenced by the proportion of patients receiving long-term treatment. It was questionable whether or not the proportion of patients that receive long-term treatment was generalisable to the daily practice patient population. As a consequence, the cost-effectiveness of the technology was highly uncertain.

In conclusion, insights in organisational, clinical and economic aspects of HTA that represent daily practice cannot be derived from RCT data. RCT data lack generalisability; usually focus on intermediate outcomes, and have a relatively short follow-up. The comparator treatment may not reflect treatment in daily practice and RCTs are not designed to study treatment sequences.

RQ2A WHAT IS THE ADDED VALUE OF REAL-WORLD DATA REGARDING ORGANISATIONAL ASPECTS OF HEALTH TECHNOLOGY ASSESSMENT?

Real-world data allow studying the uptake, accessibility and utilisation of an innovative treatment in daily clinical practice. Time may pass before a new drug is regularly prescribed and real-world data provide insight in the uptake process. For example, Chapter 3 revealed that it took two years before the innovative treatment was regularly prescribed. Furthermore, our data showed that utilisation substantially increased after additional financing was provided for using this treatment. This illustrates that contextual factors that are eliminated in RCTs have an impact on uptake and accessibility in daily practice.

Uptake and accessibility at the national and regional level can be described with real-world data. This information can be used to study whether equal access per region is guaranteed. In the Netherlands, different utilisation patterns between regions were identified which could not be explained by differences in incidence rates, guidelines or reimbursement[188]. These findings illustrate that although professional guidelines were available, treatment prescription in daily practice may be diverse, possibly influenced by organisational (most likely financial) aspects. However, it should be noted that aggregated real-world data, such as the sales data used in Chapter 3, also have limitations. Aggregated real-world data are not able to show whether patients were appropriately treated because they do not describe which patients were treated. Moreover, it is not possible to define in which treatment line the drug was prescribed, in which regimen, for how many cycles and what the clinical benefit was.

Whether the technology is used appropriately (i.e. the right patient, dosage, frequency and combination regimen) is especially relevant if the drug is used for several indications and administration is not limited to one single patient population. Providing insight in appropriate use requires more detailed real-world data which can be obtained from a population based registry such as the PHAROS-registry[20,21]. This registry included information on patient and disease characteristics as well as detailed data on treatments and resource use. Such detailed real-world data allow comparing utilisation in daily clinical practice to prescription in the RCT or clinical guidelines. First of all, it is possible to describe which patients are treated in daily practice. A treatment may receive approval for a subpopulation, however, patients who do not meet the criteria of the subpopulation may also receive the treatment in daily practice. For example, maintenance therapy was licensed for patients with a response to previous therapy, but in daily practice maintenance was also prescribed to patients with stable disease[108]. Second, detailed real-world data provide insight in the utilisation including dosages,

frequencies and combination regimens. While lower dosages might be expected in case patients are less fit, experience with the treatment over years may also increase dosing or prescription for different indications. Chapter 5 illustrated that average dosages and total number of cycles can be similar to RCT protocol, however, that patients can receive the drug with shorter intervals, i.e. more frequently, in daily practice. Deviations from the RCT protocol in daily practice may have an impact on real-world effectiveness.

While most new treatments are initially recommended to relapsing or refractory patients, over time, treatments may be recommended to newly diagnosed patients. For example thalidomide, lenalidomide and bortezomib, were initially prescribed as third-line treatment. However, these treatments were later also recommended in clinical guidelines as second- or first-line treatments. Uptake of new guidelines can only be studied with real-world data spanning multiple lines of treatment. Furthermore, in case several treatments are available, a full disease model including multiple treatments per line and covering multiple lines of treatment becomes a necessity to determine the optimal treatment sequence. We investigated the uptake of innovative treatments for multiple myeloma in the first, second and third line[148]. Our results in Chapter 7, showed that treatment guidelines, recommending thalidomide as first-line treatment, were followed and the uptake of the new guideline proved to be rather rapid. This might be related to the fact that the treatment (thalidomide) was not a new novel treatment as it was previously used in subsequent treatment lines.

In conclusion, aggregated real-world data provide insight in the uptake, utilisation and accessibility of health technologies in daily practice. More detailed real-world data are necessary to establish whether the technology is used appropriately (i.e. the right patients, dosage, frequency and combination regimen).

RQ2B WHAT IS THE ADDED VALUE OF REAL-WORLD DATA REGARDING CLINICAL ASPECTS OF HEALTH TECHNOLOGY ASSESSMENT?

Real-world data enable to study clinical aspects of HTA including effectiveness in daily practice, final outcome measures, all relevant comparators and the full disease course.

Effectiveness

This thesis showed that effectiveness of treatments can be estimated in daily practice for the real-world patient population with real-world data. Real-world analyses can confirm that technologies are not only effective for patients included in RCTs but also for the patient population of interest in daily practice. That may lead to new findings. In

Chapter 5, real-world patients with non-Hodgkin lymphoma were generally older and had better response rates to the earlier chemotherapy regimens compared to the RCT population. Effectiveness in daily practice is not only influenced by different patient characteristics, the real-world utilisation, including dosing and frequencies, also influence effectiveness.

Final outcome measures

Since the follow-up of real-world data is usually much longer, real-world data provide insight in both intermediate and final outcome measures such as OS. Furthermore, longer follow-up increases the chance of revealing rare adverse events of treatments. However, establishing the causality of adverse events is much more complicated in daily practice. Especially for haematological malignancies where treatment regimens consist of several drugs and adverse events (e.g. anaemia) can be both an adverse event and a symptom of the disease[189].

Comparators

All relevant comparator treatments are included in case real-world data are population and disease based. Incremental effectiveness of a technology can then be compared to the most relevant comparator. The most relevant comparator is not necessarily one treatment, such as in RCTs, but can be a diverse spectrum of treatments or even be different over time.

Treatment sequences

In case real-world data is collected from diagnosis till death, information is included on all lines of therapy which provides insight in treatment sequences and their effectiveness of them. For many diseases, such as in haematological malignancies, multiple treatment options are available. This not only facilitates establishing the effectiveness of a treatment, but also determining the optimal treatment order. To study the effectiveness of treatment sequences, it is necessary to account for previous and subsequent lines of therapy. As shown in Chapter 7 and 8, a micro level simulation model including multiple lines of treatment can be developed with detailed patient level real-world data for diseases with multiple treatment options. Such a simulation model can calculate intermediate (i.e. time-till-next treatment) and final outcomes (overall survival) of different real-world treatments over time. Furthermore, full disease models can identify and estimate potential health gains. Our developed micro level simulation model for multiple myeloma included five treatment options per line and three lines of treatment[148]. Treatment sequence models help to identify the optimal sequence and to facilitate an improvement of outcomes. Chapter 7 showed that an increase in the use of novel agents, would improve health outcomes.

In conclusion, real-world data can establish effectiveness of treatments in daily practice for real-world patient populations. They provide insight in long-term and final outcomes and all relevant comparators can be included in the comparison. Furthermore, real-world data enable to develop full disease models spanning multiple lines of treatment. These models provide insight in the optimal treatment sequence.

RQ2C WHAT IS THE ADDED VALUE OF REAL-WORLD DATA REGARDING ECONOMIC ASPECTS OF HEALTH TECHNOLOGY ASSESSMENT?

Real-world data provide insight in the economic aspects of technologies in daily practice. Chapter 4 illustrates that real-world data, from hospital registration systems and medical files, can be used to study the costs of treatments in daily practice. Cost estimates based on real-world data provide generalisable insight in the costs and cost drivers in daily practice since they represent costs of patients who were treated under real-world circumstances. Real-world data enable reliable estimates of budget impact while identifying cost drivers may indicate potential strategies to reduce the total costs of specific treatment.

Detailed real-world data including patient and disease characteristics, treatment prescription and resource use, as used in Chapter 5 and Chapter 7, allow simultaneously studying effects and costs and enable conducting real-world cost-effectiveness analyses. The cost-effectiveness estimates will be generalisable to patients treated in daily practice. Real-world cost-effectiveness analyses can be focused on one treatment such as in Chapter 5 where the cost-effectiveness was evaluated of one specific treatment (rituximab). However, in case real-world data are collected from diagnosis till death, analyses can also calculate the cost-effectiveness of treatment sequences. This was illustrated in Chapter 7, cost-effectiveness was calculated for real-world treatment sequences in multiple myeloma. Furthermore, a disease model facilitates a comparison of the cost-effectiveness of commonly used treatment sequences. Moreover, a micro level simulation model enables an assessment regarding the extent to which effects and costs for patients would have been better in case patients had received a different treatment sequence. Potential health gains can be identified and the incremental cost-effectiveness ratio can be calculated by comparing treatment strategies.

In conclusion, real-world data not only provide insight in the real-world costs and cost-effectiveness of all treatments independently but also provide insight into costs

and cost-effectiveness of treatment sequences and insight into the optimal order of treatments.

RQ3 WHAT ARE THE METHODOLOGICAL CHALLENGES FOR USING REAL-WORLD DATA TO INFORM HEALTH CARE DECISION MAKERS?

Although real-world data have much added value and high potential, several challenges remain to exist for comprehensively studying the organisational, clinical and economic aspects of HTA. This thesis addressed three main methodological challenges: confounding, missing data and sufficient number of patients.

Correcting for confounders

Correcting for confounders is the first main challenge. The effect of confounders is not eliminated in daily practice since circumstances are not strictly controlled and treatment is not randomly assigned as in RCTs. The population in daily practice is often far more heterogeneous and as a consequence, differences between the treatment and comparator group may exist that hamper the internal validity of results obtained from real-world data[113-115]. For example in Chapter 5, differences observed between the characteristics of the patients treated with the technology of interest compared to patients treated with the comparative treatment regarding prior treatments and disease progression. This would result in bias in case effectiveness would have been directly derived from the difference between the two groups. Propensity score matching (PSM) and multivariate regression are two often used methods to correct for observed confounders[187,190].

Propensity score matching

In Chapter 5, we illustrated that PSM, can be used to control for observed differences in patient and disease characteristics. A propensity score is calculated based on observed characteristics and represents the probability that a patient would have been in the treatment group of interest[95]. This propensity score can then be used to match patients that receive the treatment of interest to control patients. After matching, effectiveness estimates of the treatment can be retrieved from the difference between the matched groups. Although PSM enables a more reliable comparison between patients treated and the control group, it should be noted that this method also has several limitations.

First, while randomisation ensures a balance between both observed and unobserved characteristics between the groups, PSM can only match patients based on observed characteristics. Differences can remain to exist in unobserved characteristics after matching (i.e. residual confounding).

Second, there are certain requirements for using PSM. Ideally, the sample size should be large, the data should include many covariates and the circumstances for the two groups should be similar[191]. Similar circumstances ensure that observed differences between the two groups can be attributed to the treatment. However, such conditions are often not fulfilled for real-world data. As illustrated in Chapter 5, the number of patients in two groups can be limited even when real-world data were collected of many patients. Furthermore, including many covariates in the dataset does not necessarily ensure the inclusion of many covariates in the matching. For example, when certain parameters cannot be retrieved from hospital registrations, they are missing and cannot be included in the matching. Moreover, comparing real-world effectiveness and utilisation implies that organisational circumstances influence outcomes and thus that circumstances can be different. Nevertheless, population based disease registries ensure that data are obtained in a similar manner and from similar sources. While PSM is a feasible option to correct for imbalances for analyses that compare two treatments, Chapter 7 and 8 showed PSM was not feasible for the calculation of the effectiveness of treatment sequences. PSM can compare two treatment options, however, the number of treatment options is much larger when several different treatment options are available in the first-, second- and third-line. In Chapter 7 for example, first-line treatment consists of five different regimens. Furthermore, PSM assumes that it is sufficient to account for the baseline characteristics. However, to study treatment sequences it is necessary to account for characteristics at the start of second and third-line treatment as well, especially since characteristics may change over time and can influence or bias treatment effects.

Multivariate regression

Multivariate regression is another method to correct for confounders[187]. All covariates that potentially influence the outcome of interest can be included in the regression model. Chapter 8 shows that the main challenges in calculating the cost-effectiveness of treatment sequences are challenges including to correct effectiveness and costs for patient characteristics (including previous treatments), to extrapolate survival outcomes beyond observation for a heterogeneous patient population, and to estimate the timing of competing events. It is possible to account for (changing) characteristics over time with multivariate regression models per treatment line. In Chapter 7, time to next treatment in the first line was estimated with a Weibull survival model including age, sex, WHO performance status, laboratory parameters, comorbidities, inclusion in RCT, and treatment. In the second-line, the Weibull survival model also included laboratory parameters, but these were based on values at the start of second-line treatment. Nevertheless, as opposed to strictly controlled randomisation, regression models can still not account for unobserved variables and do not create balance between groups. This

is especially important in case certain characteristics are only present among patients who receive similar treatment since this may bias the regression coefficients.

Missing data

Missing data are the second main methodological challenge for using real-world data. The more variables included in the analysis, the higher the chance that the number of missing values is substantial. In contrast to Chapter 5, it was not possible to analyse complete cases only in Chapter 7. The regression models for time to next treatment included many variables and the proportion of patients who had missing data on at least one of these parameters was simply too large. Multiple imputations by chained equations can be used to impute missing values accounting for the uncertainty associated with the imputed value[120,121]. Uncertainty is accounted for by imputing missing values multiple times where the imputed value may deviate. As a result, multiple datasets are available and analyses of the combined datasets produce estimates and standard errors that account for the uncertainty associated with the imputed values. Chapter 7 uses multiple imputation by chained equations and illustrates that this method can be used to analyse the complete dataset. However, multiple imputations assume that missing values are at random and can be estimated from observed variables. Since missing values are unavailable, this can never be tested. Furthermore, the complexity of analyses significantly increases since all analyses have to be performed on the combined dataset.

Adequate patient numbers

The third challenge of real-world data is a limited number of patients in the control group and/or treatment group. In daily practice, treatment is not randomly assigned and the uptake of newly introduced treatments may be rapid. As a consequence, the number of patients in a defined control group can be limited. Data collection during a longer time frame allows using a historical control group and this increases the chance of identifying appropriate and sufficient patients who can serve as comparator. However, a historical control group should be treated under the same circumstances. As illustrated in Chapter 5, even when data of many real-world patients are collected and when uptake occurs as expected, the number of patients in both groups can be limited due to the moment real-world data collection started. Many of the patients in the real-world datasets had not yet received two lines of therapy at the moment of the analyses. Therefore, most patients were ineligible for inclusion in this study. Continuous data collection of all potential necessary variables increases the chance of having sufficient eligible patients in a dataset for real-world data analysis. However, this may not always be feasible. The studies included in this thesis were performed for haematological malignancies where incidence and prevalence of the diseases is relatively low. Hence, continuous data collection is necessary to have an adequate number of patients but

also feasible. However, if the incidence and prevalence of diseases is much larger such as in colon cancer or lung cancer, data collection would be extremely time consuming and most likely cost-inefficiently.

Timeliness

Another challenge is timely providing relevant results for decision makers. In case the uptake is slow or when the follow-up is not sufficiently long, analyses can be postponed. However, postponing analyses may reduce the relevance of studying real-world data for health care decisions makers. This is especially relevant for a disease area such as haematological malignancies since new treatments are rapidly introduced and treatments shift from later lines to first-line treatment. As a consequence, the initial research question may no longer be relevant. The challenge for real-world data to provide results in a timely matter was also highlighted in Chapter 7. Effectiveness was estimated on real-world treatment data; however, the moment the results were presented, guidelines for first-line treatment had already been revised. Consequently, the relevance of a comparison between real-world treatment patterns and guidelines was diminished. Health care decision makers are more interested in adherence to current guidelines than adherence to past guidelines. Furthermore, since current treatment patterns were most likely already changed, the identified potential health gains were less relevant. Providing insight in the latest treatment patterns was not possible since the number of patients receiving these treatments were still limited. Providing relevant results in a timely manner is crucial. However, a balance needs to be established between timeliness and relevance of the results and sufficient follow-up in order to estimate final outcomes and adequate numbers of treated patients.

In conclusion, using real-world data imposes several methodological challenges including confounding, missing data and sufficient number of patients. PSM, multivariate regression analysis multiple imputation are methods to overcome some of the challenges. Nevertheless, a sufficient number of comparable patients and timely providing results, remain important challenges.

IMPACT AND POTENTIAL USE OF REAL-WORLD DATA FOR HEALTH CARE DECISION MAKERS

Real-world data provide generalisable results for organisational, economic and clinical aspects of HTA for health care decision makers. Although patient related aspects, such as quality of life were not studied in this thesis, previous studies showed that real-world data are also able to describe patient-related aspects[192-197]. Nevertheless, the impact of

real-world data on health care decision making and ultimately, health care outcomes, is less straightforward.

According to Stryer et al.[198], there is a hierarchical order where four levels of impact can be distinguished; i) impact on further research including research on methods, ii) impact on policies, iii) impact on clinical practice and, iv) impact on health care outcomes. According to the hierarchical order, impact on health care outcomes can only be realised if studies have an impact on policies and clinical practice.

The real-world studies included in this thesis did have an impact on further research and methods. Insights were provided in areas in which scientific data were previously lacking. For example, while RCT data established the efficacy of treatments, the effectiveness in real-world practice was unknown for rituximab, thalidomide, lenalidomide and bortezomib. Furthermore, methodological guidance was provided for conducting cost-effectiveness analysis using real-world data and constructing full disease models. These practical guides facilitate future research with real-world data. Finally, our results showed that accessibility in daily practice can best be studied using both quantitative and qualitative methods.

Our studies also illustrate that real-world data influence the second level of impact (impact on policies). Cost estimates based on real-world data provided decision makers with representative results and, as a result, reimbursement rates (of treatment with stem cell transplantations) were revised based on our generated evidence. Our real-world cost-effectiveness studies provided policy makers and clinicians with generalisable insights in the cost-effectiveness of treatments in daily practice. However, the extent to which they had an impact on reimbursement decision remains to be seen. For most expensive drugs, including our real-world cost-effectiveness analysis of rituximab maintenance, a formal decision regarding continuous reimbursement has not been made[123]. Nevertheless, real-world cost-effectiveness data served as input for price negotiations and financial arrangements.

While our generated real-world evidence had an impact on the first two levels (i.e. impact on further research and policies), so far, impact on clinical practice or health care outcomes is negligible. Potentially, the identified major cost drivers from real-world studies could be used to improve the efficiency of health care delivery. However, if costs and effects are not studied simultaneously, it is not possible to illustrate whether reducing costs influences health outcomes. If negative influences on health outcomes cannot be ruled out, the impact of cost data on clinical practice remains limited. It should be noted that even when costs and effects are studied simultaneously, time passes before results from real-world data are available. During this time period, new uncertainties may have arisen, for example new treatment options may have become available. Relevant results should be provided in a timely manner to influence clinical practice and ultimately improve health outcomes.

Regarding the fourth level of impact, our real-world evidence did not yet have an impact on health outcomes. It should be noted that impact on health outcomes depends on many factors. For example, results of real-world data should be accessible (i.e. available and understandable) and acceptable (i.e. scientific, institutional appropriate and ethical)[105]. Timely access to real-world data is a major challenge especially in diseases where patients progress fast to new lines of therapy. In those circumstances, real-world analysis might soon be outdated if they are not directly available.

An online system with direct access may help to improve accessibility and timely access of real-world data. For example, the Dutch Institute for Clinical Auditing fortnightly provides physicians with benchmarked feedback, which allows physicians to compare their performance with national performances[199]. However, extreme caution is required regarding interpretation of results considering potential biases associated with real-world data. Another potential use of real-world data that may improve accessibility and may have an impact on clinical decision making is the use of real-world data for validating and creating nomograms. Nomograms provide, based on patient and disease characteristics, individual estimates for recurrence, cancer-specific and overall survival and benefit of therapies[200]. When nomograms are easily accessible, e.g. the nomogram for gallbladder cancer[201] or breastcancer[202], they can provide a convenient structure to make real-world data more accessible for physicians and hence, may better facilitate improving health outcomes. Nevertheless further research regarding the potential of real-world data for nomograms is necessary.

Future research could also investigate whether it is possible to improve acceptability of real-world results for policy makers and clinicians. The studies included in this thesis illustrated that there are several challenges associated with real-world analyses and these challenges may influence the acceptability of real-world results. Methods, including multiple imputation, propensity score matching and modelling are available to overcome some of the challenges. However, an insufficient number of comparable patients is a serious problem that cannot be resolved easily. Furthermore, real-world data will never reach the same level of internal validity as RCT data. Therefore, if RCT and real-world data show contrasting results, it is important to carefully evaluate the validity of the results of both sources.

This thesis used haematological malignancies as case study and therefore further research may be necessary to confirm conclusions. However, it should be noted that haematological malignancies have characteristics that are applicable to other disease areas. For example, treatment consist of subsequent lines with multiple treatment options available per line creating a need for a full disease model. Chapter 8 illustrated

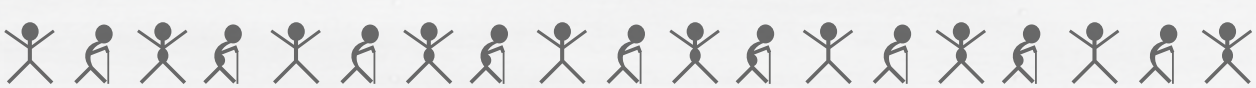
that although other diseases may have different characteristics, modelling approaches can be more or less similar.

CONCLUDING REMARKS

This thesis discusses the added value of data obtained from population based, real-world studies encompassing patients with haematological malignancies, in comparison with data obtained from RCTs. Although we only used real-world data on haematological malignancies, we believe that our findings are important and relevant for other disease areas in which multiple treatments are available in subsequent lines of treatment.

Our study shows that insights in organisational, clinical and economic aspects of HTA that represent daily practice cannot be derived from RCT data. RCT data lack generalisability, usually focus on intermediate outcomes, have difficulties detecting rare adverse events, the comparator treatment hardly ever reflects treatment in daily practice and studying treatment sequences is not possible. Real-world data provide generalisable results for organisational, economic and clinical aspects of HTA for health care decision makers. Nevertheless, using real-world data impose several challenges including correcting for confounders, missing data and limited number of patients. Although there are methods to overcome these challenges, the impact of real-world studies on health care policy and outcome has been limited so far. Impact on health outcomes can only be realised by improving the accessibility and acceptability of real-world evidence, which is the main challenge for future research.

The most valid conclusion from this thesis is that both types of data are necessary for HTA of new technologies for patients at large. The most important step forward is that (cost-)effectiveness can be analysed of treatment sequences spanning multiple lines of treatment.



Summary

INTRODUCTION

Health technology assessment (HTA) is a policy approach that examines the short- and long-term social consequences of the application or use of a health technology. The main dimensions of HTA include organisational, clinical, economic, and patient-related aspects. For a long time, HTA was mainly based on evidence from randomised controlled trials (RCTs). However, more recently, there is growing interest in HTA based on real-world data. RCTs are considered the golden standard for establishing efficacy since they demonstrate whether the treatment works and is safe under optimal and highly controlled circumstances. Although RCTs ensure internal validity, the results are not generalisable to the context of care in daily practice (i.e. external validity). Technologies may not be adopted, adopted under different circumstances or applied to patients who do not fulfil the inclusion criteria of studies. As a consequence, results from RCTs might not be sufficient for making decisions in daily clinical practice. Therefore, real-world evidence is increasingly requested. Real-world data are not collected through conventional RCTs, but from real-world clinical practice (i.e. not under controlled circumstances). While findings from real-world data are of more practical value to health care decision makers, using real-world data also imposes methodological challenges due to the uncontrolled circumstances and absence of random treatment assignment. The evidence base regarding the actual added value of real-world data compared to RCTs is currently inconclusive. This thesis evaluates the added value of real-world evidence for health care decision makers using haematological diseases as case study.

PART ONE: LIMITATIONS OF EVIDENCE FROM RANDOMISED CONTROLLED TRIALS

Chapter 2 reports on the single technology appraisal of lenalidomide, a treatment for patients with myelodysplastic syndromes. Evidence regarding outcomes and costs was primarily obtained from a randomised trial. However, with only RCT data available, several uncertainties remained regarding the effectiveness and costs in daily practice as well as long-term outcomes. The appraisal committee accepted a commitment from the manufacturer to publish, once available, real-world data. This was believed to provide reassurance regarding the value for money.

PART TWO: ORGANISATIONAL, CLINICAL AND ECONOMIC ASPECTS OF HEALTH TECHNOLOGY ASSESSMENT

To obtain insight in the organisational aspects of HTA, **Chapter 3** reports on the utilisation and accessibility of an expensive drug (i.e. bortezomib) for patients with multiple myeloma**.** The accessibility to bortezomib treatment was investigated at national and regional levels using real-world data (i.e. interviews and sales data from 2004-2009). Interviews with stakeholders revealed awareness of the high treatment costs, although prescription barriers were not encountered. Sales data revealed that utilisation increased slowly, indicating a long adjustment period. Furthermore, utilisation remains below the rate estimated by the professional association of haematologists and regional differences were observed. Based on our results, we concluded that utilisation and accessibility can best be studied with both qualitative and quantitative real-world data. Providing these real-world insights can enhance evidence-based decision making and improve appropriate and efficient utilisation of expensive drugs.

In **Chapter 4**, we describe how economic aspects of HTA can be studied with real-world data. The real-world costs for stem cell transplantations (SCTs) were calculated and major cost drivers were identified. Real-world data were obtained from patients, treated at three university hospitals, who underwent an autologous (auto) SCT or allogeneic (allo) SCT in 2007, 2008 or 2009. Allo-SCT included sibling donors, matched unrelated donors (MUD) and umbilical cord blood (UCB). Resource use was collected from the hospital registration systems and medical files. The average total costs per patient were €45,670 for auto-SCT and €101,919 for sibling allo-SCT. The costs of transplantations from unrelated donors were much higher: €171,478 for allo-SCT-MUD and €254,689 for allo-SCT-UCB. Hospital inpatient days together with laboratory and other activities were the main cost drivers across all types of SCT. Besides, donor search costs were a large cost component in allo-SCT-sib (18%) and allo-SCT-MUD (12%). We concluded that real-world costs were above routine reimbursement and appropriate financing is necessary to guarantee the continuation of SCT.

Economic aspects of HTA were further addressed in **Chapter 5**. The real-world cost-effectiveness of rituximab maintenance compared with observation in relapsed or refractory follicular lymphoma patients was calculated. Data were obtained from a trial and from population based registries. A Markov model was developed to calculate cost per life year gained (LYG) and quality-adjusted life years (QALYs) for different scenarios. Using real-world data as well as results from long-term trial follow-up showed favourable incremental cost-effectiveness ratios for rituximab maintenance. Nevertheless, results

showed that caution is required with data synthesis, interpretation and generalisability of results.

PART THREE: PRACTICAL GUIDANCE AND METHODOLOGICAL CHALLENGES

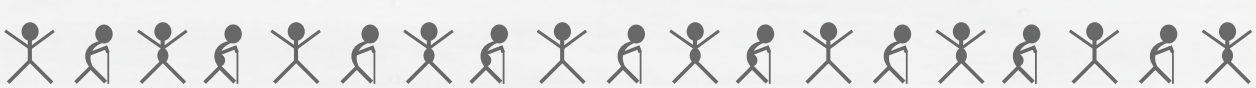
Practical guidance on using real-world data was provided in **Chapter 6.** This chapter describes the required steps necessary to perform a sound economic evaluation using an economic evaluation conducted with real-world data as example. Methodological challenges for using real-world data were identified as well and included three main issues: confounding by indication, missing data, and insufficient numbers of (comparable) patients. If encountered, it is crucial to accurately deal with these issues to maximise the internal validity and generalisability of the outcomes and their value to decision makers. Multivariate regression modelling, propensity score matching, and data synthesis are well-established methods to deal with confounding. Multiple imputation methods should be used in cases where data are missing at random. We concluded that decision makers should realise that real-world evidence provides extremely valuable and relevant policy information, but needs to be assessed differently compared with evidence derived from an RCT.

Calculating real-world cost-effectiveness of strategies spanning multiple lines is one of the main potentials of real-world data. This was illustrated in **Chapter 7** where the cost-effectiveness of real-world treatment sequences for elderly patients with multiple myeloma were calculated. Using real-world data, a patient-level simulation model was designed comprising three treatment lines. The model enabled to study the impact of novel treatments in daily practice and revealed that utilisation of novel agents increased over the years. This improved survival and increased costs. Real-world treatment patterns were also compared to the optimal treatment strategy. This comparison showed that outcomes could be improved at favourable incremental cost-effectiveness ratios.

Although models comprising multiple treatment lines are increasing in importance, experience is scarce. In **Chapter 8**, guidance is provided in constructing a model to support cost-effectiveness analyses of treatment strategies spanning multiple treatment lines. Based on experiences with two models used in cost-effectiveness analyses of treatment strategies in cancer, we discuss how best practices can best be implemented. Additional recommendations were provided wherever best practices were unavailable or not applicable. Guidance based on hands-on experience with two models can improve the validity of future disease models and cost-effectiveness analyses.

DISCUSSION

This thesis discusses the added value of evidence obtained from real-world studies encompassing patients with haematological malignancies, in comparison with evidence obtained from RCTs. Our study shows that insights in organisational, clinical and economic aspects of HTA that represent daily practice cannot be derived from RCT data. Real-world data provide generalisable results for organisational, economic and clinical aspects of HTA for health care decision makers. Furthermore, real-world cost-effectiveness of treatment sequences can be calculated. Nevertheless, using real-world data impose several challenges including correcting for confounders, missing data and limited number of patients. Although there are methods to overcome these challenges, the impact of real-world studies on health care policy and outcome has been limited so far. Impact on health outcomes can only be realised by improving the accessibility and acceptability of real-world evidence, which is the main challenge for future research. The most valid conclusion from this thesis is that both types of data are necessary for HTA of new technologies. The most important step forward is that real-world (cost-) effectiveness can be analysed of treatment sequences spanning multiple lines of treatment.



Samenvatting

INTRODUCTIE

Health Technology Assessment (HTA) is een beleidsaanpak die de korte en lange termijn effecten onderzoekt van het toepassen van een gezondheidstechnologie. De belangrijkste onderdelen van HTA omvatten organisatorische, klinische, economische en patiënt gerelateerde aspecten. Oorspronkelijk was HTA voornamelijk gebaseerd op gegevens uit gerandomiseerde gecontroleerde studies (RCTs). RCTs laten zien of een behandeling effectief en veilig is onder optimale en gecontroleerde omstandigheden. Hierdoor worden RCTs gezien als de gouden standaard voor het vaststellen van de effectiviteit van een technologie. De interne validiteit van de resultaten is groot maar dit gaat ten koste van de generaliseerbaarheid (externe validiteit) voor de dagelijkse praktijk. In de dagelijkse praktijk worden technologieën mogelijk niet geïmplementeerd, geïmplementeerd onder andere omstandigheden of toegepast bij patiënten met andere karakteristieken dan de patiënten in de RCTs. Hierdoor zijn de resultaten van RCTs mogelijk niet voldoende om beslissingen te nemen en neemt de vraag naar gegevens uit de dagelijkse praktijk toe. Deze gegevens kunnen niet verkregen worden via de traditionele RCTs. Hoewel resultaten uit de dagelijkse praktijk van praktische waarde zijn voor het nemen van beslissingen in de dagelijkse praktijk, brengt het analyseren van deze gegevens ook methodologische problemen mee vanwege de niet gecontroleerde omstandigheden en de afwezigheid van randomisatie. De daadwerkelijke toegevoegde waarde van resultaten uit de dagelijkse praktijk ten opzichte van gegevens uit RCTs is op dit moment niet eenduidig. Dit proefschrift evalueert de toegevoegde waarde van resultaten uit de dagelijkse praktijk voor het nemen van beslissingen in de gezondheidszorg. Hematologische ziekten worden hierbij gebruikt als casus.

DEEL ÉÉN: BEPERKINGEN VAN RESULTATEN UIT GERANDOMISEERDE STUDIES

In **Hoofdstuk 2** wordt de behoordeling beschreven van lenalidomide, een behandeling voor patiënten met myelodysplastische syndromen. De effecten en kosten waren voornamelijk gebaseerd op een gerandomiseerde studie. Hierdoor was er onzekerheid over de effectiviteit en kosten in de dagelijkse praktijk, alsmede de lange termijn uitkomsten. Er werd besloten de behandeling te vergoeden, maar onder de voorwaarde dat de fabrikant resultaten uit de dagelijkse praktijk zou verzamelen waarmee een latere herevaluatie mogelijk zou zijn.

DEEL TWEE: ORGANISATORISCHE, KLINISCHE EN ECONOMISCHE ASPECTEN VAN HEALTH TECHNOLOGY ASSESSMENT

In **Hoofdstuk 3** wordt inzicht verkregen in de waarde van gegevens uit de dagelijkse praktijk voor de organisatorische aspecten van HTA. Het gebruik en de toegankelijkheid voor een duur geneesmiddel (i.e. bortezomib) voor patiënten met multiple myeloom is onderzocht op nationaal en regionaal niveau met behulp van interviews en verkoopcijfers over 2004-2009. Uit interviews bleken stakeholders op de hoogte van de hoge behandelkosten, maar barrières om het dure geneesmiddel voor te schrijven werden niet ervaren. De verkoopcijfers lieten zien dat het gebruik langzaam toenam en dit duidde op een lange aanpassingsperiode voordat het geneesmiddel regulier werd voorgeschreven. Het daadwerkelijke gebruik bleef lager dan het verwachte gebruik door de beroepsgroep en daarnaast was het gebruik tussen de regio's niet evenredig aan het aantal myeloom patiënten per regio. Op basis van de gevonden resultaten concluderen we dat het gebruik en de toegankelijkheid van een behandeling het beste bestudeerd kan worden met kwantitatieve en kwalitatieve gegevens uit de dagelijkse praktijk. Het bestuderen van deze gegevens kan gepast en efficiënt gebruik van dure geneesmiddelen verbeteren.

In **Hoofdstuk 4**, beschrijven we hoe de economische aspecten van HTA bestudeerd kunnen worden met behulp van gegevens uit de klinische praktijk. De kosten van stamceltransplantaties (SCTs) zijn berekend en de belangrijkste kostenposten zijn geïdentificeerd. Gegevens uit de dagelijkse praktijk zijn verzameld uit ziekenhuis informatiesystemen en medische dossiers. Patiënten uit drie universitaire ziekenhuizen werden geïncludeerd indien ze in 2007, 2008 of 2009 een transplantatie ondergingen van een autologe (auto) of allogene (allo) donor. Allo-SCTs konden SCTs van verwante donoren, niet verwante donoren (MUD) of navelstrengbloed (UCB) zijn. De totale kosten per patiënt voor een auto-SCT waren €45,670. De kosten voor een allo-SCT van een verwante donor waren €101,919. De kosten van een SCT via een niet verwante donor of navelstrengbloed waren een stuk hoger, respectievelijk €171,478 en €254,689. Opnamedagen in het ziekenhuis en ziekenhuisactiviteiten, zoals laboratorium bepalingen en beeldvormende diagnostiek, waren de grootste kosten posten bij alle type transplantaties. Voor de allo-SCTs van niet verwante donoren of navelstrengbloed vormden de kosten voor het zoeken van een donor ook een grote kostenpost, 18% bij niet verwante en 12% bij navelstrengbloed SCTs. Gegevens uit de dagelijkse praktijk lieten zien dat de daadwerkelijke kosten hoger waren dan de huidige vergoedingstarieven.

Economische aspecten van HTA werden ook bestudeerd in **Hoofdstuk 5**. In dit hoofdstuk is de kosten-effectiviteit in de dagelijkse praktijk berekend van rituximab als onderhoudsbehandeling in vergelijking met geen onderhoudsbehandeling voor patiënten met gerecidiveerd of refractair folliculair lymfoom. Gegevens werden verkregen via de registratie trial en registers. Een Markov model werd ontwikkeld om de kosten te berekenen per gewonnen levensjaar en per voor kwaliteit gecorrigeerd levensjaar voor verschillende scenario's. De incrementele kosten-effectiviteit van rituximab was gunstig, zowel op basis van resultaten uit de dagelijkse praktijk als op basis van langdurige opvolging van de patiënten in de registratie studie. Echter, de resultaten lieten ook zien dat we voorzichtig moeten zijn met het samenvoegen van verschillende gegevensbronnen, de interpretatie en de generaliseerbaarheid van verschillende scenario's.

DEEL DRIE: PRAKTISCHE AANBEVELINGEN EN METHODOLOGISCHE UITDAGINGEN

In **Hoofdstuk 6** worden er praktische aanbevelingen gegeven voor het gebruik van klinische praktijk gegevens. Dit hoofdstuk bespreekt de noodzakelijke stappen voor het uitvoeren van een goede economische evaluatie en gebruikt hierbij een praktisch voorbeeld van een economische evaluatie op basis van gegevens uit de dagelijkse praktijk. Er worden methodologische uitdagingen geïdentificeerd waarvan de afwezigheid van randomisatie, het ontbreken van gegevens, en onvoldoende vergelijkbare patiënten de belangrijkste drie zijn.

Het is belangrijk om deze uitdagingen op de juiste manier te adresseren om de interne validiteit en generaliseerbaarheid van resultaten zo goed mogelijk te waarborgen. Multivariate regressie modellen, propensity score matching en het combineren van meerdere databronnen zijn erkende methoden om de problemen aan te pakken die voortvloeien uit het niet randomiseren van behandelingen. Meervoudige imputatie kan gebruikt worden wanneer de missende gegevens willekeurig zijn. Resultaten gebaseerd op gegevens uit de dagelijkse praktijk zijn zeer waardevol bij het nemen van beslissingen in de gezondheidszorg en bij het maken van beleid. Echter, de resultaten moeten wel anders beoordeeld worden dan resultaten die afkomstig zijn uit RCTs.

Eén van de belangrijkste mogelijkheden van klinische praktijk gegevens is het berekenen van de kosten-effectiviteit van complete behandelstrategieën die bestaan uit meerdere behandellijnen. Dit wordt geïllustreerd in **Hoofdstuk 7** waarin de kosten-effectiviteit wordt omschreven van behandelstrategieën in de dagelijkse praktijk voor oudere patiënten met multiple myeloom. Met behulp van gegevens uit de dagelijkse praktijk is een microsimulatie model gebouwd bestaande uit drie behandellijnen. Met behulp van

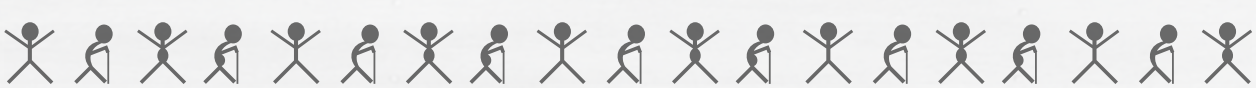
het model kan het effect van nieuwe behandel mogelijkheden in de dagelijkse praktijk bestudeerd worden. Daarnaast lieten klinische praktijk gegevens zien dat het gebruik van nieuwe behandelingen steeg over de jaren. Deze stijging verbeterde de overleving, maar zorgde ook voor hogere kosten. Behandelstrategieën in de dagelijkse praktijk werden ook vergeleken met de optimale behandelstrategie. Deze vergelijking liet zien dat uitkomsten in de dagelijkse praktijk verder verbeterd konden worden tegenover gunstige kosten-effectiviteitratio's.

Ondanks dat modellen voor complete behandelstrategieën steeds belangrijker worden, is de ervaring op dit moment beperkt. In **Hoofdstuk 8** worden er aanbevelingen gegeven voor het construeren van een model om de kosten-effectiviteit te berekenen van behandelstrategieën. Op basis van twee ziektemodellen wordt er beschreven hoe de huidige richtlijnen geïmplementeerd kunnen worden bij het ontwikkelen van een ziektemodel voor behandelstrategieën. Additionele aanbevelingen voor het ontwikkelen van een ziektemodel worden gegeven wanneer de huidige richtlijnen onvoldoende van toepassing zijn of houvast bieden. Deze aanbevelingen kunnen de validiteit van toekomstige ziektemodellen en kosten-effectiviteitsanalyses van behandelstrategieën verbeteren.

DISCUSSIE

In dit proefschrift wordt de toegevoegde waarde van gegevens uit de dagelijkse praktijk omschreven op basis van studies naar hematologische ziektes. Deze studies laten zien dat resultaten op basis van RCTs beperkt generaliseerbaar zijn. Met gegevens uit de dagelijkse praktijk is het mogelijk de organisatorische, klinische en economische aspecten van HTA te bestuderen. Dit geeft resultaten die generaliseerbaar zijn en niet verkregen kunnen worden uit de reguliere RCTs.

Echter, het gebruik van klinische praktijk gegevens brengt ook uitdagingen met zich mee zoals het corrigeren van verstorende variabelen, missende gegevens en een beperkt aantal patiënten. Ondanks dat er methoden beschikbaar zijn om met deze uitdagingen om te gaan, is de invloed van resultaten uit de dagelijkse praktijk op het nemen van beslissingen in de gezondheidszorg tot nu toe beperkt. Invloed kan alleen gerealiseerd worden door de toegankelijkheid en acceptatie van resultaten uit de dagelijkse praktijk te verbeteren en dit is de belangrijkste uitdaging voor toekomstig onderzoek. De belangrijkste conclusie op dit moment over de waarde van gegevens uit de dagelijkse praktijk is dat voor HTA zowel gegevens uit RCTs als uit de dagelijkse praktijk noodzakelijk zijn. De belangrijkste stap voorwaarts is dat de kosten-effectiviteit van volledige behandelstrategieën in de dagelijkse praktijk berekend kan worden.



PhD Portfolio

PhD Portfolio

Name PhD candidate:	Hedwig Blommestein
Erasmus University Rotterdam, department:	instituut Beleid & Management Gezondheidszorg
PhD period:	2010-2015
Promotor:	Prof.dr. Carin A. Uyl-de Groot
	Prof.dr. Peter C. Huijgens

PHD TRAINING

- Advanced modelling methods for health economic evaluations *York March 2010*
- Academic writing in English for iBMG staff *Rotterdam April-September 2010*
- Klaar in vier jaar. Birgitte Hertz *Rotterdam Oktober-November 2010*
- Pharmacoeconomic Modeling – Applications ISPOR short courses *Prague November 2010*
- Patient Registries November ISPOR short courses *Prague 2010*
- Survival Analysis for Clinicians NIHES winter programme *Rotterdam January-February 2011*
- English Biomedical Writing and Communication *Rotterdam March-July 2011*
- Risbo Module Geven van Onderwijs I – kleine groepen *Rotterdam July 2012*
- Risbo Module Toetsing I – beoordeling en feedback *Rotterdam June 2012*
- Propensity Scores and Observational Studies of Treatment Effect ISPOR short courses *Berlin November 2012*
- Thesis Supervision Erasmus University *Rotterdam April-June 2015*

TEACHING

- Methode en Technieken 1: Workgroups (2010-2015) and lectures (2013-2015)
- Statistiek aanschuifvariant: Workgroups (2011)
- Methode en Technieken 2: Workgroups (2014)
- Health Technology Assessment: Workgroups (2013, 2015) and lectures 2015-2016
- NIHES Health Economics summer course: Workgroups (2013, 2014, 2015)
- Supervising and co-reading bachelor and master theses (2011-2015)

PRESENTATIONS AT CONFERENCES

Workshop

18th Annual European Congress of International Society for Pharmacoeconomics and Outcomes Research, Milan Italy 2015

- Challenges and solutions to successfully determine real-world cost-effectiveness

Oral presentation

5th Dutch Hematology Congress, Arnhem The Netherlands 2011

- Costs of autologous and allogeneic stem cell transplantations for haematological diseases in The Netherlands. A cooperative study by HOVON, IMTA, and the Dutch Society for Hematology.

Poster presentations

American Society of Hematology

San Diego December 2011

- Cost of Autologous and Allogeneic Stem Cell Transplantations for Hematological Disease: A Dutch Multicenter Daily Practice Study

New Orleans December 2013

- One line does not make a picture: real-world cost-effectiveness of multiple myeloma treatments using a full disease model

European Hematology Association

Amsterdam June 2012

- Pharmacoeconomic evaluation of rituximab as maintenance treatment for follicular lymphoma: results of a real world population based study
- Valuation of health related quality of life of long term survivors of lymphoma: a population based study

International Society For Pharmacoeconomics and Outcomes Research.

Prague November 2010

- Accessibility and utilisation of bortezomib in Dutch daily practice.

Madrid November 2011

- Cost of Autologous and Allogeneic Stem Cell Transplantations for Hematological Disease: A Dutch Multicenter Daily Practice Study.
- Diagnostics and treatment of patients with non-small-cell lung cancer in daily practice.
- Health related quality of life of long term survivors of lymphoma: a population based study

Berlin November 2012

- Pharmacoeconomic evaluation of rituximab as maintenance treatment for follicular lymphoma: results of a real world population based study

- Can a population-based patient registry improve the feasibility of outcomes research in multiple myeloma?
- Are population-based registries a suitable tool for outcomes research in cancer?
- Experiences from four registries (Nomination)

Dublin 2013

- One line does not make a picture: real-world cost-effectiveness of multiple myeloma treatments using a full disease model (Nomination)

PRESENTATIONS AT OTHER MEETINGS

1e najaarssymposium Werkgroep Hematologie Friesland Leeuwarden October 2012

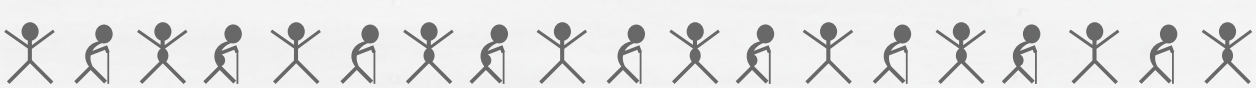
- Farmaco-economische evaluatie van rituximab maintenance bij recidief folliculair lymfoom: een voorbeeldstudie in de farmaco-economie

Werkconferentie PHAROS Vumc Amsterdam March 2013

- "Outcomes research in non-Hodgkin lymphoma" Uitkomstenonderzoek in non-Hodgkin lymfoom

Uitkomstenonderzoek en Registries Seminar Kennisinstituut Geneesmiddelen en Medische Technologie Den Dolder February 2014

- De mogelijkheden en onmogelijkheden van uitkomstenonderzoek voor doelmatigheid



List of publications

LIST OF SCIENTIFIC PUBLICATIONS

Included in this thesis

Blommestein HM, Armstrong N, Ryder S, Deshpande S, Worthy G, Noake C, Riemsma R, Kleijnen J, Severens JL, Al MJ. Lenalidomide for the Treatment of Low- or Intermediate-1-Risk Myelodysplastic Syndromes Associated with Deletion 5q Cytogenetic Abnormality: An Evidence Review of the NICE Submission from Celgene. PharmacoEconomics. 2015 Aug 28. [Epub ahead of print]

Blommestein HM, Verelst SG, de Groot S, Huijgens PC, Sonneveld P, Uyl-de Groot CA. A cost-effectiveness analysis of real-world treatment for elderly patients with multiple myeloma using a full disease model. Eur J Haematol. 2015 Apr 18. doi: 10.1111/ejh.12571. [Epub ahead of print]

Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: lessons learned from the PHAROS registry. PharmacoEconomics. 2015 Jun;33(6):551-60.

Blommestein HM, Franken MG, Verelst SG, van Agthoven M, Huijgens PC, Uyl-de Groot CA. Access to expensive cancer drugs in Dutch daily practice: should we be concerned? Neth J Med. 2014 May; 72(4):235-41.

Blommestein HM, Issa DE, Pompen M, Ten Hoor G, Hogendoorn M, Joosten P, Zweegman S, Huijgens PC, Uyl-de Groot CA. Cost-effectiveness of rituximab as maintenance treatment for relapsed follicular lymphoma: results of a population-based study. Eur J Haematol. 2014;92(5):398-406.

Blommestein HM, Verelst SG, Huijgens PC, Blijlevens NM, Cornelissen JJ, Uyl-de Groot CA. Real-world costs of autologous and allogeneic stem cell transplantations for haematological diseases: a multicentre study. Ann Hematol. 2012 Dec;91(12):1945-52.

Not included in this thesis

Franken MG, Gaultney JG, Blommestein HM, Huijgens PC, Sonneveld P, Redekop WK, Uyl-de Groot CA. Policymaker, please consider your needs carefully: does outcomes research in relapsed or refractory multiple myeloma reduce policymaker uncertainty regarding value for money of bortezomib? Value Health. 2014 Mar;17(2):245-53.

Leunis A, Blommestein HM, Huijgens PC, Blijlevens NM, Jongen-Lavrencic M, Uyl-de Groot CA. The costs of initial treatment for patients with acute myeloid leukemia in the Netherlands. Leuk Res. 2013 Mar;37(3):245-50.

OTHER PUBLICATIONS

Rob Riemsma R, Büyükkaramikli N, Corro Ramos I, Armstrong N, Blommestein H, Westwood M, Portegijs P, Kanters T, Worthy G, Ross J, Al M, Severens JL, Kleijnen. Asfotase alfa for treating paediatric-onset hypophosphatasia: a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd, 2015. Report to the National Institute for Health and Care Excellence

Riemsma R, Al M, Büyükkaramikli N, Armstrong N, Blommestein H, Clay F, Ross J, Worthy G, Severens H, Kleijnen J. Ramucirumab for treating advanced gastric cancer or gastro-oesophageal junction adenocarcinoma previously treated with chemotherapy a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd, 2015. Report to the National Institute for Health and Care Excellence

Riemsma R, Al M, Blommestein H, Deshpande S, Ryder S, Worthy G, Noake C, Armstrong N, Severens JL, Kleijnen J. Lenalidomide for the treatment of myelodysplastic syndromes associated with deletion 5q cytogenetic abnormality: a Single Technology Appraisal. York: Kleijnen Systematic Reviews Ltd, 2013. Report to the National Institute for Health and Care Excellence

Franken M, Gaultney J, Blommestein H, Tan S, van Gils C, Huijgens P, Sonneveld P, Redekop W, Uyl-de Groot C. Pilot outcomes research: Effects and costs of bortezomib in relapsed or refractory multiple myeloma. Report to the Dutch Health Care Insurance Board (CVZ). 2010

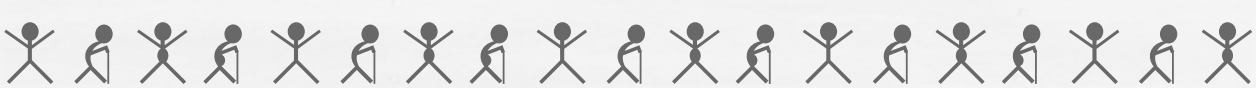
Blommestein HM, Verelst SGR, Cornelissen JJ, Huijgens PC, Uyl-de Groot CA. Kostprijzen van autologe en allogene hematopoietische stamceltransplantaties voor hematologische ziekten. Onderzoeksrapport iBMG 2010.03

Blommestein HM, Blijlevens NMA, Leebeek FWG, Schutgens REG, Uyl-de Groot CA. Kostenrapportage Hemofilie. Onderzoeksrapport iBMG 2011.03

Blommestein HM, Issa DE, Huijgens PC, Uyl-de Groot CA. ZonMW Report Review Rituximab – Diffuse Large B-cell Lymphoma 2011

Blommestein HM, Issa DE, Huijgens PC, Uyl-de Groot CA. ZonMW Report Review Rituximab – Follicular Lymphoma 2011

Blommestein H, Uyl-de Groot C, Visser O, Oerlemans S, Verelst S, Van den Broek E, Issa D, Aarts M, Louwman M, Sonneveld P, Postuma W, Coebergh J, Van de Poll L, Huijgens P. Impact of new systemic treatments of patients with hematological malignancies in the Netherlands: population-based cohort studies of process and outcome as a basis for assessments of cost-effectiveness. Report of PHAROS NHL/CLL/MM 2004-2012. ZonMW report 2014.



Dankwoord

Mijn proefschrift is klaar! Ik heb dit zeker niet alleen gedaan en wil graag iedereen bedanken voor zijn of haar bijdrage.

Een aantal mensen wil ik in het bijzonder bedanken. Allereerst mijn twee promoteren. Carin, vanaf de eerste dag heb je me de kans en het vertrouwen gegeven om als een zelfstandig onderzoeker aan de slag te gaan. Ik bewonder en waardeer jouw enorme enthousiasme voor het onderzoek, grote hoeveelheid kennis en optimisme. In combinatie met je belangstelling voor mij en de mensen om me heen, had ik me geen fijnere promotor kunnen wensen.

Peter, zoals jij, is er maar één. Je bent enorm gedreven en je energie en kennis lijken oneindig. Wat is het fijn om daar gebruik van te mogen maken. Ik wil je bedanken voor alle kansen die je mij de afgelopen jaren geboden hebt en ook nu nog biedt.

Naast mijn promotoren wil ik de commissieleden bedanken voor het lezen van mijn proefschrift en het opponeren tijdens de verdediging.

De artikelen uit dit proefschrift zouden niet tot stand gekomen zijn zonder coauteurs. Allemaal bedankt voor jullie input! Maiwenn, jou wil ik in het bijzonder bedanken voor de ontzettend leerzame en hele fijne samenwerking. Margreet, zonder jou had ik misschien niet eens bij BMG gewerkt! Vanaf dag één ben je zo behulpzaam geweest en heb je altijd het beste met me voorgehad. Bedankt voor je enorme bijdrage aan de artikelen en de introductie en discussie van dit proefschrift. Cobi, je bent weliswaar geen coauteur maar wat zouden we zonder jou moeten!

Het merendeel van de hoofdstukken is gebaseerd op data uit PHAROS. Graag wil ik alle datamanagers, de deelnemende patiënten, de ziekenhuizen, de projectleden en andere betrokkenen bedanken voor hun onmisbare bijdrage. Zonder jullie zou het niet mogelijk zijn geweest om de waarde van gegevens uit de dagelijkse praktijk te onderzoeken. Marieke Koelink, jouw bijdrage wil ik nadrukkelijk benoemen. Bedankt voor je inzet en persoonlijke aandacht op het juiste moment. Het Integraal Kankercentrum Nederland wil ik verder graag bedanken voor de mogelijkheid om het onderzoek met real-world data voort te zetten.

Heel veel uren heb ik de afgelopen jaren doorgebracht bij het iBMG. Uren die echt omgevlogen zijn dankzij alle fijne collega's. Bedankt! Eén van de hoogtepunten die ik in het bijzonder wil noemen is onze deelname aan de Roparun van Parijs naar Rotterdam. Team 179! Een unieke ervaring waar ik nog heel vaak aan terugdenk.

Apo, Matthijs, Pieter en Tim, samen met de meiden vormen jullie voor mij een uniek collectief. Inhoudelijke discussies hebben we de afgelopen jaren afgewisseld met kort even buurten en vele feestjes op of buiten het werk, #onmisbaar!

Lieve Laura, we hebben een hoop uren samen op de kamer doorgebracht. Ik wil je allereerst bedanken voor al het meedenken de afgelopen jaren. Daarnaast hebben je aanstekelijke vrolijkheid (de voorpret voor kerst begon altijd ruim van tevoren!) en razendsnelle humor ervoor gezorgd dat we de afgelopen jaren ontzettend veel gelachen hebben.

Lieve Igna, wat was het fijn om samen met jou aan dit promotietraject te beginnen. Ik heb zoveel aan je gehad en niet alleen als mede trouwe bezoeker bij de lunch of een andere BMG-bijeenkomst. Wanneer er iets niet lekker liep had je het altijd direct door en er samen over praten, vaak aan het einde van de dag, zorgde altijd voor nieuwe energie. Wat fijn dat je mijn paranimf wilt zijn.

Lieve Saskia, wat is het fijn om zo'n lieve vriendin iedere dag om je heen te hebben en een echte luxe om daar ook op zo'n prettige manier samen mee te werken. Ik ben enorm trots op wat we samen ontdekt en geleerd hebben. Zonder jou was ik echt niet tot hier gekomen en daarom ben ik erg blij dat je vandaag mijn paranimf wilt zijn.

Lieve familie en vrienden, er is door jullie allemaal altijd met zoveel belangstelling geïnformeerd naar dit proefschrift. Ik wil jullie allemaal bedanken voor de interesse en aanmoedigingen. Ik hoop dat jullie het mooi vinden.

Lieve Renske, je bent altijd belangstellend en geïnteresseerd, ook in mijn werk. Onze trip naar Rusland was superleuk en zo welkom na het inleveren van mijn proefschrift bij de kleine commissie. Lieve Maja, Anneloes en Janine, als BMG vriendinnetjes hebben we de basis gelegd voor dit proefschrift maar ook voor heel veel leuke tripjes en etentjes. Ik vind het fijn dat we lekker over BMG onderwerpen kunnen kletsen maar nog leuker om met jullie feestjes te bouwen.

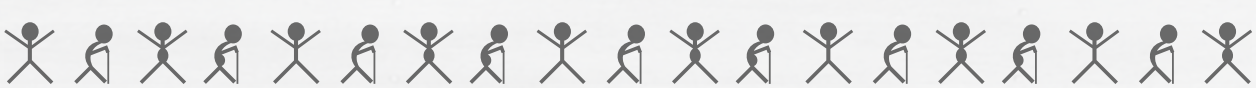
Van mijn familie wil ik in het bijzonder mijn broers, zwager, schoonzussen, opa's en oma's, schoonouders en ouders bedanken. Allemaal hebben jullie altijd, met alles wat mij bezighoudt, enorm meegeleefd. Echt uniek en wat fijn dat ik jullie dit nu kan laten zien!

Lieve broers, Arjan, wat is het fijn als jij een weekendje bij ons in Rotterdam bent. René, wat hebben we een fijne tijd gehad op de Blaardorpstraat tijdens onze studie en recent in Australië.

Lieve Ed en Marijke, ik kan me geen fijnere schoonouders wensen. Ed, nog voordat ik wist wat het was, zei jij dat ik “lekker moest gaan promoveren”. Dat zou goed voor mijn CV zijn. Bedankt voor de tip! Graag wil ik jullie allebei bedanken voor de noodzakelijke ontspanning tijdens heerlijke etentjes en fijne vakanties.

Lieve pap en mam, ik wil jullie bedanken voor alle steun en interesse. Wat is het fijn om zulke trotse ouders te hebben die altijd achter je staan. Niets is jullie ooit teveel en ik wil jullie bedanken voor echt alles wat jullie voor mij doen op ieder vlak.

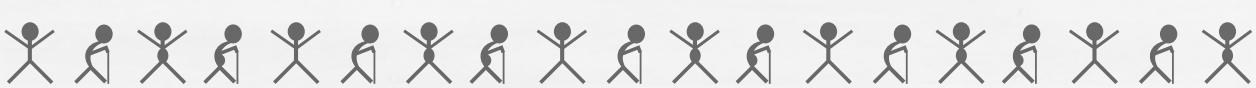
Allerliefste Michael, in al je bescheidenheid zal je onterecht jouw aandeel in dit proefschrift weten te bagatelliseren. Je heerlijke gevoel voor humor, relativeringsvermogen en liefde zijn voor mij echter onmisbaar, en waren dat ook bij het schrijven van mijn proefschrift. Ik word blij als ik aan je denk, intens gelukkig als ik bij je ben. Zonder jou had ik hier nooit op deze manier kunnen staan. Ik ben ontzettend blij om dit traject te kunnen afronden en kijk uit naar alle mooie dingen, groot en nu nog heel klein, die we samen (zullen) delen.



About the author

Hedwig Sluimer-Blommestein was born on March 13th 1987 in Zaandam. In 2009 she obtained her Master's degree in Health, Economics, Policy and Law at the Erasmus University Rotterdam. Since 2010 she has been working as a researcher at the institute of Health Care Policy and Management and institute for Medical Technology Assessment, Erasmus University Rotterdam. From 2014 onwards, she has also been working at the Netherlands Comprehensive Cancer Organisation.

Her research activities mainly focus on economic evaluations of expensive drugs in haematological diseases using real-world data. She was involved in multiple access and reimbursement dossiers for multiple myeloma and non-Hodgkin lymphoma in the Netherlands and was part of an evidence review committee for several single technology appraisals by the National Institute for Health and Care Excellence.



References

1. OECD Health Statistics 2014. Frequently requested data. http://www.oecd.org/els/health-systems/oecd-health-statistics-2014-frequently-requested-data.htm. Updated November 27th 2014. Accessed April 1, 2015.
2. De Meijer C, Wouterse B, Polder J, Koopmanschap M. The effect of population aging on health expenditure growth: A critical review. *European Journal of Ageing*. 2013;10(4):353-361.
3. Sullivan R, Peppercorn J, Sikora K, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncol*. 2011;12(10):933-980.
4. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the economic evaluation of health care programmes.* 3rd ed. New York: Oxford University Press; 2005.
5. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*. 1977;296(13):716-721.
6. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder M. International comparison of the definition and the practical application of health technology assessment. *Int J Technol Assess Health Care*. 2005;21(1):89-95.
7. 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;10(5):326-335.
8. Garrido M, Kristensen F, Nielsen C, Busse R. Health technology assessment and health policy-making in europe. current status, challenges and potential. observatory studies series. 2008; 14.
9. Krumholz HM. Outcomes research: Generating evidence for best practice and policies. *Circulation*. 2008;118(3):309-318.
10. Ray W, Griffin M, Avorn J. Evaluating drugs after their approval for clinical use. *The New England journal of medicine*. 1993;329(27):2029-2032.
11. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. effectiveness versus efficacy studies. *Pharmacoeconomics*. 1999;15(5):423-434.
12. Dreyer NA, Garner S. Registries for robust evidence. *JAMA*. 2009;302(7):790-791.
13. Gliklich R, Dreyer N, eds. Registries for evaluating patient outcomes: A user's guide (prepared fby outcome DEcIDE center [outcome sciences, inc. d/b/a outcome] under contract no. HHSA29020050035I TO3). AHRQ publication no.10-EHC049. Rockville, MD: Agency for Healthcare Research and Quality; September 2010.
14. Sullivan R, Purushotham A. Analysis of media reporting of the lancet oncology commission. *Lancet Oncol*. 2012;13(1):12-13.
15. Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the european union: A population-based cost analysis. *Lancet Oncol*. 2013;14(12):1165-1174.
16. Vanneman M, Dranoff G. Combining immunotherapy and targeted therapies in cancer treatment. *Nat Rev Cancer*. 2012;12(4):237-251.
17. Tax SEM, Scheres EAG, van der Hoeven JJM. Effective new anticancer agents, but the financing system creaks. obstacles and solutions in the utilisation of expensive cancer drugs. [in dutch: Effectieve nieuwe middelen tegen kanker, maar het financieringssysteem kraakt. belemmeringen en oplossingen bij de inzet van dure geneesmiddelen tegen kanker]. 2015;Juli.

18. Smith A, Howell D, Patmore R, Jack A, Roman E. Incidence of haematological malignancy by sub-type: A report from the haematological malignancy research network. *Br J Cancer*. 2011; 105(11):1684-1692.

19. Bellera C, Praud D, Petit-Moneger A, McKelvie-Sebileau P, Soubeyran P, Mathoulin-Pelissier S. Barriers to inclusion of older adults in randomised controlled clinical trials on non-hodgkin's lymphoma: A systematic review. *Cancer Treat Rev*. 2013;39(7):812-817.

20. Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: Lessons learned from the PHAROS registry. *Pharmacoeconomics*. 2015.

21. Huijgens PC, Posthuma EFM, Coebergh JWW, van de Poll-Franse LV, Uyl-de Groot CA, Sonneveld P. A 'population based registry' for hemato-oncology. *Nederlands Tijdschrift voor Hematologie*. 2010;7:321-325.

22. Netherlands Cancer Registry. www.iknl.nl.

23. National Institute for health and Clinical Excellence. Guide to single technology appraisal process. london: NICE. http://www.nice.org.uk/media/42D/B3/STAGuideLrFinal.pdf. Updated 2010. Accessed September, 2013.

24. National Institute for health and Clinical Excellence. Myelodysplastic syndrome (deletion 5q) - lenalidomide [ID480]. http://guidance.nice.org.uk/TAG/291. Accessed January, 2014.

25. Wade R, Duarte A, Simmonds M, et al. The clinical and cost effectiveness of aflibercept in combination with irinotecan and fluorouracil-based therapy (FOLFIRI) for the treatment of metastatic colorectal cancer which has progressed following prior oxaliplatin-based chemotherapy: A critique of the evidence. *Pharmacoeconomics*. 2015;33(5):457-466.

26. Stevenson M, Pandor A, Stevens JW, et al. Nalmefene for reducing alcohol consumption in people with alcohol dependence: An evidence review group perspective of a NICE single technology appraisal. *Pharmacoeconomics*. 2015.

27. Holmes M, Davis S, Simpson E. Alteplase for the treatment of acute ischaemic stroke: A NICE single technology appraisal; an evidence review group perspective. *Pharmacoeconomics*. 2015;33(3):225-233.

28. Greenhalgh J, Bagust A, Boland A, et al. Erlotinib and gefitinib for treating non-small cell lung cancer that has progressed following prior chemotherapy (review of NICE technology appraisals 162 and 175): A systematic review and economic evaluation. *Health Technol Assess*. 2015; 19(47):1-134.

29. Latimer NR, Carroll C, Wong R, Tappenden P, Venning MC, Luqmani R. Rituximab in combination with corticosteroids for the treatment of anti-neutrophil cytoplasmic antibody-associated vasculitis: A NICE single technology appraisal. *Pharmacoeconomics*. 2014;32(12):1171-1183.

30. Simpson EL, Davis S, Thokala P, Breeze PR, Bryden P, Wong R. Sipuleucel-T for the treatment of metastatic hormone-relapsed prostate cancer: A NICE single technology appraisal; an evidence review group perspective. *Pharmacoeconomics*. 2015.

31. The Myelodysplastic Syndromes Foundation. Understanding myelodysplastic syndromes: A patient handbook. www.mds-foundation.org/pdf/handbook-english.pdf. Updated 2008. Accessed August 2nd, 2014.

32. Greenberg PL, Young NS, Gattermann N. Myelodysplastic syndromes. *Hematology Am Soc Hematol Educ Program*. 2002:136-161.

33. Bowen D, Culligan D, Jowitt S, et al. Guidelines for the diagnosis and therapy of adult myelodysplastic syndromes. *Br J Haematol*. 2003;120(2):187-200.

34. HESonline. Hospital episode statistics. http://www.hesonline.nhs.uk/. Accessed June, 2011.

35. Balducci L. Transfusion independence in patients with myelodysplastic syndromes: Impact on outcomes and quality of life. *Cancer*. 2006;106(10):2087-2094.

36. European Medicines Agency. Committee for medicinal products for human use. assessment report revlimid. procedure no. EMEA/H/C/000717/II/0056. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Assessment_Report_-_Variation/human/000717/WC500146340.pdf. Updated EMA/ 297642/2013. Accessed August, 2014.

37. Fenaux P, Giagounidis A, Selleslag D, et al. A randomized phase 3 study of lenalidomide versus placebo in RBC transfusion-dependent patients with low-/intermediate-1-risk myelodysplastic syndromes with del5q. *Blood*. 2011;118(14):3765-3776.

38. Jaeger M, Aul C, Sohngen D, Germing U, Schneider W. Secondary hemochromatosis in polytransfused patients with myelodysplastic syndromes. *Beitr Infusionsther*. 1992;30:464-468.

39. Kontoghiorghes GJ, Pattichi K, Hadjigavriel M, Kolnagou A. Transfusional iron overload and chelation therapy with deferoxamine and deferiprone (L1). *Transfus Sci*. 2000;23(3):211-223.

40. Wahlin A, Markevarn B, Golovleva I, Nilsson M. Prognostic significance of risk group stratification in elderly patients with acute myeloid leukaemia. *Br J Haematol*. 2001;115(1):25-33.

41. Szende A, Schaefer C, Goss TF, et al. Valuation of transfusion-free living in MDS: Results of health utility interviews with patients. *Health Qual Life Outcomes*. 2009;7:81-7525-7-81.

42. Szende A, Schaefer C, Goss TF, et al. Valuation of transfusion-free living in MDS: Results of health utility interviews with patients. *Health Qual Life Outcomes*. 2009;7:81-7525-7-81.

43. Dolan SP, Gudex C, Kind P, Williams A. A social tariff for EuroQol: Results from a UK general population survey. *The University of York. Discussion paper 138*. 1995;September.

44. Malcovati L, Porta MG, Pascutto C, et al. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: A basis for clinical decision making. *J Clin Oncol*. 2005;23(30):7594-7603.

45. Germing U, Lauseker M, Hildebrandt B, et al. Survival, prognostic factors and rates of leukemic transformation in 381 untreated patients with MDS and del(5q): A multicenter study. *Leukemia*. 2012;26(6):1286-1292.

46. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: A randomised, open-label, phase III study. *Lancet Oncol*. 2009;10(3):223-232.

47. Kuendgen A, Lauseker M, List AF, et al. Lenalidomide does not increase AML progression risk in RBC transfusion-dependent patients with low- or intermediate-1-risk MDS with del(5q): A comparative analysis. *Leukemia*. 2013;27(5):1072-1079.

48. Hutton J, Trueman P, Henshall C. Coverage with evidence development: An examination of conceptual and policy issues. *Int J Technol Assess Health Care*. 2007;23(4):425-432.

49. Trueman P, Grainger DL, Downs KE. Coverage with evidence development: Applications and issues. *Int J Technol Assess Health Care*. 2010;26(1):79-85.

50. BKV. Report on treatment of breast cancer patients with metastatic HER2-positive tumors *[in dutch: Rapport onderbehandeling van borstkankerpatiënten met uitgezaaide HER2-positieve tumoren]*. 2005;050603.

51. Delwel GO. Guidance for outcomes research for the assessment of the costeffectiveness of in-patient medicines. 2008;270: 28113706.
52. Health Insurance Board (College voor zorgverzekeringen). CFH-rapport 07/30 bortezomib(velcade®). 2007.
53. Jagannath S, Barlogie B, Berenson J, et al. A phase 2 study of two doses of bortezomib in relapsed or refractory myeloma. *Br J Haematol*. 2004;127(2):165-172.
54. Richardson PG, Barlogie B, Berenson J, et al. A phase 2 study of bortezomib in relapsed, refractory myeloma. *N Engl J Med*. 2003;348(26):2609-2617.
55. Richardson PG, Sonneveld P, Schuster MW, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *N Engl J Med*. 2005;352(24):2487-2498.
56. Kumar SK, Rajkumar SV, Dispenzieri A, et al. Improved survival in multiple myeloma and the impact of novel therapies. *Blood*. 2008;111(5):2516-2520.
57. Verelst SGR, Blommestein HM, Karim-Kos HE, Huijgens P,C, Sonneveld P. Trends in incidence and survival of multiple myeloma in the netherlands in the last two decades. results from a national population based study. 2011;118(Abstract 5071).
58. Sonneveld P, Zweegman S, Vellenga E, et al. Guidelines for treatment of plasma cell dyscrasias in 2010. *Nederlands Tijdschrift voor Hematologie*. 2010;7:84-95.
59. Groot MT, Huijgens PC, Uyl-de Groot CA. Introduction of expensive pharmaceuticals in haemato-oncology in the netherlands and throughout europe. 2006;12(6):30-36.
60. Wilking N, Jönsson B, Högberg D. Comparator report on patient access to cancer drugs in europe. 2009;January 15.
61. Cresswell JW. *Research design: Qualitative, quantitative, and mixed method approaches.* 2nd ed. California, United States of America: Sage Publication, Inc; 2003.
62. Franken MG, Gaultney J, Blommestein HM, et al. Pilot outcomes research: Effects and costs of bortezomib in relapsed or refractory multiple myeloma. 2010.
63. HOVON. HOVON 65 MM / GMMG HD4 clinical picture: MM (multiple myeloom) ISRCTN64455289. http://www.hovon.nl/trials/trials-by-type/mm.html?action=showstudie&studie_id=5&categorie_id=3. Accessed May, 2012.
64. HOVON. HOVON 86 MM clinical picture: MM (multiple myeloom) EudraCT number:2007-002533-37. http://www.hovon.nl/trials/trials-by-type/mm.html?action=showstudie&studie_id=60&categorie_id=3. Accessed May, 2012.
65. Rogers EM. *Diffusion of innovations.* 4th ed. New York: The Free Press; 1995.
66. Berwick DM. Disseminating innovations in health care. *JAMA*. 2003;289(15):1969-1975.
67. Wagelaar M, Lubberding S, Hugtenburg J, Wagner C. Accessibility of expensive drugs. an investigation into drug use in breast cancer and multiple myeloma. *[in dutch: Toegankelijkheid van dure geneesmiddelen. een onderzoek naar geneesmiddelgebruik bij mammacarcinoom en multipel myeloom.]*. 2010;ISBN 9789461220011.
68. Ljungman P, Bregni M, Brune M, et al. Allogeneic and autologous transplantation for haematological diseases, solid tumours and immune disorders: Current practice in europe 2009. *Bone Marrow Transplant*. 2010;45(2):219-234.
69. Lodewyck T, Oudshoorn M, van dH, et al. Predictive impact of allele-matching and EBMT risk score for outcome after T-cell depleted unrelated donor transplantation in poor-risk acute leukemia and myelodysplasia. *Leukemia*. 2011;25(10):1548-1554.

70. Lodewyck T, Cornelissen JJ. Allogeneic stem cell transplantation in acute myeloid leukemia: A risk-adapted approach. *Blood Rev*. 2008;22(6):293-302.

71. HOVON – the haemato oncology foundation for adults in the netherlands - HOVON stem cell working group and transplantation registry, part of the international registry of the european group for blood and marrow transplantation (EBMT). 2012.

72. van Agthoven M, Vellenga E, Fibbe WE, Kingma T, Uyl-de Groot CA. Cost analysis and quality of life assessment comparing patients undergoing autologous peripheral blood stem cell transplantation or autologous bone marrow transplantation for refractory or relapsed non-hodgkin's lymphoma or hodgkin's disease. a prospective randomised trial. *Eur J Cancer*. 2001; 37(14):1781-1789.

73. van Agthoven M, Segeren CM, Buijt I, et al. A cost-utility analysis comparing intensive chemotherapy alone to intensive chemotherapy followed by myeloablative chemotherapy with autologous stem-cell rescue in newly diagnosed patients with stage II/III multiple myeloma; a prospective randomised phase III study. *Eur J Cancer*. 2004;40(8):1159-1169.

74. Mishra V, Vaaler S, Brinch L. A prospective cost evaluation related to allogeneic haemopoietic stem cell transplantation including pretransplant procedures, transplantation and 1 year follow-up procedures. *Bone Marrow Transplant*. 2001;28(12):1111-1116.

75. Mishra V, Andresen S, Brinch L, et al. Cost of autologous peripheral blood stem cell transplantation: The norwegian experience from a multicenter cost study. *Bone Marrow Transplant*. 2005;35(12):1149-1153.

76. The Dutch Healthcare Authority. DBC-tariefapplicatie. http://dbc-tarieven.nza.nl/Nza-tarieven/top.do. Updated 31-12-2009. Accessed 2/12, 2010.

77. Gaultney JG, Franken MG, Tan SS, et al. Real-world health care costs of relapsed/refractory multiple myeloma during the era of novel cancer agents. *J Clin Pharm Ther*. 2013;38(1):41-47.

78. 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;13(6):712-719.

79. Majhail NS, Mothukuri JM, Brunstein CG, Weisdorf DJ. Costs of hematopoietic cell transplantation: Comparison of umbilical cord blood and matched related donor transplantation and the impact of posttransplant complications. *Biology of Blood and Marrow Transplantation*. 2009;15(5):564-573.

80. Saito AM, Cutler C, Zahrieh D, et al. Costs of allogeneic hematopoietic cell transplantation with high-dose regimens. *Biology of Blood and Marrow Transplantation*. 2008;14(2):197-207.

81. Lee SJ, Klar N, Weeks JC, Antin JH. Predicting costs of stem-cell transplantation. *J Clin Oncol*. 2000;18(1):64-71.

82. Esperou H, Brunot A, Roudot-Thoraval F, et al. Predicting the costs of allogeneic sibling stem-cell transplantation: Results from a prospective, multicenter, french study. *Transplantation*. 2004;77(12):1854-1858.

83. Dutch Healthcare Authority. Dutch healthcare authority; beleidsregel CI-1125 functiegerichte budgettering academische ziekenhuizen 2010. http://www.nza.nl/137706/145406/CI-1125.pdf. Updated 2010. Accessed May, 2012.

84. Fisher RI, LeBlanc M, Press OW, Maloney DG, Unger JM, Miller TP. New treatment options have changed the survival of patients with follicular lymphoma. *J Clin Oncol*. 2005;23(33): 8447-8452.

85. Liu Q, Fayad L, Cabanillas F, et al. Improvement of overall and failure-free survival in stage IV follicular lymphoma: 25 years of treatment experience at the university of texas M.D. anderson cancer center. *J Clin Oncol*. 2006;24(10):1582-1589.

86. Tan D, Horning SJ, Hoppe RT, et al. Improvements in observed and relative survival in follicular grade 1-2 lymphoma during 4 decades: The stanford university experience. *Blood*. 2013; 122(6):981-987.

87. Salles G, Mounier N, de Guibert S, et al. Rituximab combined with chemotherapy and interferon in follicular lymphoma patients: Results of the GELA-GOELAMS FL2000 study. *Blood*. 2008;112(13):4824-4831.

88. Herold M, Haas A, Srock S, et al. Rituximab added to first-line mitoxantrone, chlorambucil, and prednisolone chemotherapy followed by interferon maintenance prolongs survival in patients with advanced follicular lymphoma: An east german study group hematology and oncology study. *J Clin Oncol*. 2007;25(15):1986-1992.

89. Hershman DL, McBride RB, Eisenberger A, Tsai WY, Grann VR, Jacobson JS. Doxorubicin, cardiac risk factors, and cardiac toxicity in elderly patients with diffuse B-cell non-hodgkin's lymphoma. *J Clin Oncol*. 2008;26(19):3159-3165.

90. Marcus R, Imrie K, Belch A, et al. CVP chemotherapy plus rituximab compared with CVP as first-line treatment for advanced follicular lymphoma. *Blood*. 2005;105(4):1417-1423.

91. van Oers MH, Klasa R, Marcus RE, et al. Rituximab maintenance improves clinical outcome of relapsed/resistant follicular non-hodgkin lymphoma in patients both with and without rituximab during induction: Results of a prospective randomized phase 3 intergroup trial. *Blood*. 2006;108(10):3295-3301.

92. van Oers MH, Van Glabbeke M, Giurgea L, et al. Rituximab maintenance treatment of relapsed/resistant follicular non-hodgkin's lymphoma: Long-term outcome of the EORTC 20981 phase III randomized intergroup study. *J Clin Oncol*. 2010;28(17):2853-2858.

93. Franken MG, van Gils CW, Gaultney JG, 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;49(1):8-16.

94. Hoogendoorn M, Joosten P, Storm H, Kibbelaar RE. HemoBase: An intelligent electronic patient file as aid for hemato-onocology care. *Nederlands Tijdschrift voor Hematologie*. 2009; 6(3):104-109.

95. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.

96. Tan S, Bouwmans C, Hakkaart-van Roijen L. *Handleiding voor kostenonderzoek: Methoden en referentieprijzen voor economische evaluaties in de gezondheidszorg.* Diemen: College voor Zorgverzekeringen; 2010. http://www.cvz.nl/binaries/content/documents/zinl-www/documenten/publicaties/overige-publicaties/1007-handleiding-voor-kostenonderzoek/Handleiding+voor+kostenonderzoek.pdf.

97. Wild D, Walker M, Pettengell R, Lewis G. Utility elicitation in patients with follicular lymphoma. [abstract no. PCN62]. *Value in Health*. 2006;9:A294.

98. Bouwmans C, Janssen J, Huijgens P, Uyl-de Groot C. Costs of haematological adverse events in chronic myeloid leukaemia patients: A retrospective cost analysis of the treatment of anaemia, neutropenia and thrombocytopenia in patients with chronic myeloid leukaemia. *J Med Econ*. 2009;12(2):164-169.

99. Deconinck E, Miadi-Fargier H, Pen CL, Brice P. Cost effectiveness of rituximab maintenance therapy in follicular lymphoma: Long-term economic evaluation. *Pharmacoeconomics*. 2010; 28(1):35-46.

100. Kasteng F, Erlanson M, Hagberg H, Kimby E, Relander T, Lundkvist J. Cost-effectiveness of maintenance rituximab treatment after second line therapy in patients with follicular lymphoma in sweden. *Acta Oncol*. 2008;47(6):1029-1036.

101. Demousis F, Athanasiadis A. Cost-effectiveness analysis of maintenance treatment with rituximab (MabThera) for non-hodgkins lymphoma patients in greece, a local adaptation. *Value in Health*. 2012;15:A421 (PCN 69).

102. Rutten-van Mölken, MPMH., Uyl-de Groot C, Rutten F. *Costs and effects. A guideline for economic evaluations in health care.* 2nd ed. Amsterdam: Elsevier Gezondheidszorg; 2010.

103. Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS)-explanation and elaboration: A report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. *Value Health*. 2013; 16(2):231-250.

104. Berger ML, Martin BC, Husereau D, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value Health*. 2014;17(2):143-156.

105. Merlo G, Page K, Ratcliffe J, Halton K, Graves N. Bridging the gap: Exploring the barriers to using economic evidence in healthcare decision making and strategies for improving uptake. *Appl Health Econ Health Policy*. 2014;Oct 7 [Epub ahead of print].

106. Williams I, Bryan S. Understanding the limited impact of economic evaluation in health care resource allocation: A conceptual framework. *Health Policy*. 2007;80(1):135-143.

107. Kuijpers MR, Toenders WGM. Assessment procedure for inpatient drugs [in dutch: Procedure beoordeling intramurale geneesmiddelen]. 2006;26022597.

108. Blommestein HM, Issa DE, Pompen M, et al. Cost-effectiveness of rituximab as maintenance treatment for relapsed follicular lymphoma: Results of a population-based study. *Eur J Haematol*. 2014;92(5):398-406.

109. van Gils CW, de Groot S, Tan SS, et al. Real-world resource use and costs of adjuvant treatment for stage III colon cancer. *Eur J Cancer Care (Engl)*. 2013.

110. Health Insurance Board [in Dutch College Voor Zorgverzekeringen]. Guidelines for pharmacoeconomic research, updated version. http://www.ispor.org/PEguidelines/source/HTAGuidelinesNLupdated2006.pdf. Updated 2006. Accessed October 15th, 2014.

111. van de Poll-Franse LV, Horevoorts N, Eenbergen Mv, et al. The patient reported outcomes following initial treatment and long term evaluation of survivorship registry: Scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. *Eur J Cancer*. 2011;47(14):2188-2194.

112. Franken MG, Gaultney JG, Blommestein HM, et al. Policymaker, please consider your needs carefully: Does outcomes research in relapsed or refractory multiple myeloma reduce policymaker uncertainty regarding value for money of bortezomib? *Value Health*. 2014;17(2): 245-253.

113. Concato J, Lawler EV, Lew RA, Gaziano JM, Aslan M, Huang GD. Observational methods in comparative effectiveness research. *Am J Med*. 2010;123(12 Suppl 1):e16-23.

114. 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;12(8):1053-1061.

115. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: The importance of restriction. *JAMA*. 2010;304(8):897-898.

116. 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;59(5):437-447.

117. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of non-uniform effect. *Am J Epidemiol*. 2006;163(3):262-270.

118. Glynn RJ, Schneeweiss S, Wang PS, Levin R, Avorn J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. *J Clin Epidemiol*. 2006;59(8):819-828.

119. 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;31(8):703-718.

120. Rubin D. *Multiple imputation for nonresponse in surveys.* New York: Wiley; 1987.

121. Little R, Rubin D. *Statistical analysis with missing data.* New York: John Wiley & Sons; 1987.

122. Gerber-Grote A, Windeler J. What is the contribution of health economic evaluations to decision-making in health care? experiences from 7 selected countries. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2014;108(7):358-359.

123. Franken MG, Koopmanschap M, Steenhoek A. Health economic evaluations in reimbursement decision making in the netherlands; time to take it seriously? *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen*. 2014;108(7):383-389.

124. Gaultney JG, Sanhueza E, Janssen JJ, Redekop WK, Uyl-de Groot CA. Application of cost-effectiveness analysis to demonstrate the potential value of companion diagnostics in chronic myeloid leukemia. *Pharmacogenomics*. 2011;12(3):411-421.

125. 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*. 2014;271(6):1673-1678.

126. Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: The role of effectiveness studies in evaluating cardiovascular therapies. *Circulation*. 2008;118(12):1294-1303.

127. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010;60(5):277-300.

128. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in europe in 2008. *Eur J Cancer*. 2010;46(4):765-781.

129. Mateos MV, San Miguel JF. How should we treat newly diagnosed multiple myeloma patients? *Hematology Am Soc Hematol Educ Program*. 2013;2013:488-495.

130. Palumbo A, Bringhen S, Ludwig H, et al. Personalized therapy in multiple myeloma according to patient age and vulnerability: A report of the european myeloma network (EMN). *Blood*. 2011;118(17):4519-4529.

131. Kyle RA. Prognostic factors in multiple myeloma. *Stem Cells*. 1995;13 Suppl 2:56-63.

132. Ludwig H, Bolejack V, Crowley J, et al. Survival and years of life lost in different age cohorts of patients with multiple myeloma. *J Clin Oncol*. 2010;28(9):1599-1605.

133. Ludwig H, Sonneveld P. Disease control in patients with relapsed and/or refractory multiple myeloma: What is the optimal duration of therapy? *Leuk Res*. 2012;36 Suppl 1:S27-34.

134. Kristinsson SY, Landgren O, Dickman PW, Derolf AR, Bjorkholm M. Patterns of survival in multiple myeloma: A population-based study of patients diagnosed in sweden from 1973 to 2003. *J Clin Oncol*. 2007;25(15):1993-1999.

135. Kumar SK, Dispenzieri A, Lacy MQ, et al. Continued improvement in survival in multiple myeloma: Changes in early mortality and outcomes in older patients. *Leukemia*. 2014;28(5): 1122-1128.

136. San Miguel JF, Garcia-Sanz R. Prognostic features of multiple myeloma. *Best Pract Res Clin Haematol*. 2005;18(4):569-583.

137. Fonseca R, Stewart AK. Targeted therapeutics for multiple myeloma: The arrival of a risk-stratified approach. *Mol Cancer Ther*. 2007;6(3):802-810.

138. Bang SM, Kyle RA, Rajkumar SV, Kumar S. Treatment patterns and outcomes in elderly patients with multiple myeloma. *Leukemia*. 2013;27(4):971-974.

139. Fayers PM, Palumbo A, Hulin C, et al. Thalidomide for previously untreated elderly patients with multiple myeloma: Meta-analysis of 1685 individual patient data from 6 randomized clinical trials. *Blood*. 2011;118(5):1239-1247.

140. San Miguel JF, Schlag R, Khuageva NK, et al. Persistent overall survival benefit and no increased risk of second malignancies with bortezomib-melphalan-prednisone versus melphalan-prednisone in patients with previously untreated multiple myeloma. *J Clin Oncol*. 2013;31(4): 448-455.

141. Palumbo A, Hajek R, Delforge M, et al. Continuous lenalidomide treatment for newly diagnosed multiple myeloma. *N Engl J Med*. 2012;366(19):1759-1769.

142. Gaultney JG, Redekop WK, Sonneveld P, Uyl-de Groot CA. Novel anticancer agents for multiple myeloma: A review of the evidence for their therapeutic and economic value. *Expert Rev Anticancer Ther*. 2012;12(6):839-854.

143. Garrison LP,Jr, Wang ST, Huang H, et al. The cost-effectiveness of initial treatment of multiple myeloma in the U.S. with bortezomib plus melphalan and prednisone versus thalidomide plus melphalan and prednisone or lenalidomide plus melphalan and prednisone with continuous lenalidomide maintenance treatment. *Oncologist*. 2013;18(1):27-36.

144. Hornberger J, Rickert J, Dhawan R, Liwing J, Aschan J, Lothgren M. The cost-effectiveness of bortezomib in relapsed/refractory multiple myeloma: Swedish perspective. *Eur J Haematol*. 2010;85(6):484-491.

145. Moller J, Nicklasson L, Murthy A. Cost-effectiveness of novel relapsed-refractory multiple myeloma therapies in norway: Lenalidomide plus dexamethasone vs bortezomib. *J Med Econ*. 2011;14(6):690-697.

146. Brown RE, Stern S, Dhanasiri S, Schey S. Lenalidomide for multiple myeloma: Cost-effectiveness in patients with one prior therapy in england and wales. *Eur J Health Econ*. 2013;14(3): 507-514.

147. Zweegman S, Palumbo A, Bringhen S, Sonneveld P. Age and aging in blood disorders: Multiple myeloma. *Haematologica*. 2014;99(7):1133-1137.

148. Blommestein HM, Verelst SG, de Groot S, Huijgens PC, Sonneveld P, Uyl-de Groot CA. A cost-effectiveness analysis of real-world treatment for elderly patients with multiple myeloma using a full disease model. *Eur J Haematol*. 2015.

149. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999;18(6):681-694.

150. Kleinbaum D, Klein M. Parametric survival models, chapter 7. In: Gail M, Krickberg K, Samet J, Tsiatis A, Wong W, eds. *Statistics for biology and health. survival analysis A self-llearning text.* 2nd ed. New York, United States of America: Springer Science+Business Media, Inc.; 2005: 257-329.

151. Population based HAematological registry for observational studies (PHAROS). http://www.pharosregistry.nl/. Updated 2015.

152. Caro JJ. Pharmacoeconomic analyses using discrete event simulation. *Pharmacoeconomics*. 2005;23(4):323-332.

153. EuroQol Group. EQ-5D-5L value sets. http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html. Updated 2014. Accessed July 2014, 2014.

154. Medicijnkosten. Zorginstituut Nederland. http://www.medicijnkosten.nl/. Updated 2014. Accessed June 26th, 2014.

155. Armoiry X, Fagnani F, Benboubker L, et al. Management of relapsed or refractory multiple myeloma in french hospitals and estimation of associated direct costs: A multi-centre retrospective cohort study. *J Clin Pharm Ther*. 2011;36(1):19-26.

156. Hakkaart-van Roijen L, Tan S, Bouwmans C. *Guidelines for research on costs.* Diemen: College van Zorgverzekeraars; 2010.

157. Briggs A, Claxton K, Sculpher M. Correlation parameters. cholesky decomposition for multivariate normal distribution. In: *Decision modelling for health economic evaluation.* New York: Oxford University Press Inc; 2006:95-97.

158. Zweegman S, Lokhorst HM, Levin MD, et al. Guidelines for the treatment of multiple myeloma 2012. *Nederlands Tijdschrift voor Hematologie*. 2012;9(8):300.

159. Liwing J, Uttervall K, Lund J, et al. Improved survival in myeloma patients: Starting to close in on the gap between elderly patients and a matched normal population. *Br J Haematol*. 2014; 164(5):684-693.

160. Gaultney JG, Uyl-de Groot CA. Efficient allocation of novel agents in multiple myeloma: A work in progress. *Oncologist*. 2013;18(1):5-7.

161. Sonneveld P, Verelst SG, Lewis P, et al. Review of health-related quality of life data in multiple myeloma patients treated with novel agents. *Leukemia*. 2013;27(10):1959-1969.

162. Tosh J, Archer R, Davis S, Stevenson M, Stevens JW. Golimumab for the treatment of rheumatoid arthritis after the failure of previous disease-modifying antirheumatic drugs: A NICE single technology appraisal. *Pharmacoeconomics*. 2013;31(8):653-661.

163. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices--overview: A report of the ISPOR-SMDM modeling good research practices task force-1. *Med Decis Making*. 2012;32(5):667-677.

164. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in europe: Estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49(6):1374-1403.

165. Escudier B, Porta C, Schmidinger M, et al. Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2014;25 Suppl 3:iii49-56.

166. Motzer RJ, Escudier B, Bukowski R, et al. Prognostic factors for survival in 1059 patients treated with sunitinib for metastatic renal cell carcinoma. *Br J Cancer*. 2013;108(12):2470-2477.

167. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med*. 2007;356(2):115-124.

168. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: A double-blind, randomised, placebo-controlled phase III trial. *Lancet*. 2008;372(9637):449-456.

169. Motzer RJ, Hutson TE, Tomczak P, et al. Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2009;27(22):3584-3590.

170. Motzer RJ, Escudier B, Oudard S, et al. Phase 3 trial of everolimus for metastatic renal cell carcinoma : Final results and analysis of prognostic factors. *Cancer*. 2010;116(18):4256-4265.

171. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med*. 2007;356(2):125-134.

172. Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol*. 2009;27(20):3312-3318.

173. de Groot S, Blommestein H, Redekop W, Oosterwijk E, Kiemeney L, Uyl-de Groot C. The cost-effectiveness of sequential first- and second-line treatments in metastatic renal cell carcinoma using real-world data and a patient-level simulation model. *Value in health*. 2013;16(7):A587.

174. Greipp PR, San Miguel J, Durie BG, et al. International staging system for multiple myeloma. *J Clin Oncol*. 2005;23(15):3412-3420.

175. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J. Modeling using discrete event simulation: A report of the ISPOR-SMDM modeling good research practices task force-4. *Med Decis Making*. 2012;32(5):701-711.

176. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: A report of the ISPOR-SMDM modeling good research practices task force--3. *Value Health*. 2012;15(6):812-820.

177. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: An overview and guide. *Pharmacoeconomics*. 2008;26(2):131-148.

178. Latimer N. NICE DSU technical support document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011.

179. Leunis A, Redekop WK, van Montfort KA, Lowenberg B, Uyl-de Groot CA. The development and validation of a decision-analytic model representing the full disease course of acute myeloid leukemia. *Pharmacoeconomics*. 2013;31(7):605-621.

180. Davis S, Stevenson M, Tappenden P, Wailoo A. NICE DSU technical support document 15: Cost-effectiveness modelling using patient-level simulation. 2014.

181. Dutch National Health Care Institute (ZIN) (formerly named CVZ). Guidelines for pharmacoeconomic research [in dutch: Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie]. 2006;25001605.

182. Goossens L. *Underestimated uncertainties. hospital-at-home for COPD exacerbations and methodological issues in the economic evaluation of healthcare.* Erasmus University Rotterdam; 2013.

183. Heeg BM, Damen J, Buskens E, Caleo S, de Charro F, van Hout BA. Modelling approaches: The case of schizophrenia. *Pharmacoeconomics*. 2008;26(8):633-648.

184. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: The ISPOR RCT-CEA task force report. *Value Health*. 2005;8(5):521-533.

185. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess*. 2004;8(36):iii-iv, ix-xi, 1-158.

186. O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: Is there a future for piggyback evaluations? *Value Health*. 2005;8(1):67-79.

187. Hershman DL, Wright JD. Comparative effectiveness research in oncology methodology: Observational data. *J Clin Oncol*. 2012;30(34):4215-4222.

188. Blommestein HM, Franken MG, Verelst SG, van Agthoven M, Huijgens PC, Uyl-de Groot CA. Access to expensive cancer drugs in dutch daily practice: Should we be concerned? *Neth J Med*. 2014;72(4):235-241.

189. Ludwig H, Pohl G, Osterborg A. Anemia in multiple myeloma. *Clin Adv Hematol Oncol*. 2004; 2(4):233-241.

190. Rovithis D. Do health economic evaluations using observational data provide reliable assessment of treatment effects? *Health Econ Rev*. 2013;3(1):21-1991-3-21.

191. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiol Drug Saf*. 2004; 13(12):841-853.

192. Oerlemans S, Husson O, Mols F, et al. Perceived information provision and satisfaction among lymphoma and multiple myeloma survivors--results from a dutch population-based study. *Ann Hematol*. 2012;91(10):1587-1595.

193. Oerlemans S, Mols F, Issa DE, et al. A high level of fatigue among long-term survivors of non-hodgkin's lymphoma: Results from the longitudinal population-based PROFILES registry in the south of the netherlands. *Haematologica*. 2013;98(3):479-486.

194. Oerlemans S, Issa DE, van den Broek EC, et al. Health-related quality of life and persistent symptoms in relation to (R-)CHOP14, (R-)CHOP21, and other therapies among patients with diffuse large B-cell lymphoma: Results of the population-based PHAROS-registry. *Ann Hematol*. 2014;93(10):1705-1715.

195. Oerlemans S, Issa DE, van den Broek EC, et al. Impact of therapy and disease-related symptoms on health-related quality of life in patients with follicular lymphoma: Results of the population-based PHAROS-registry. *Eur J Haematol*. 2014.

196. Oerlemans S, Mols F, Nijziel MR, Zijlstra WP, Coebergh JW, van de Poll-Franse LV. The course of anxiety and depression for patients with hodgkin's lymphoma or diffuse large B cell lymphoma: A longitudinal study of the PROFILES registry. *J Cancer Surviv*. 2014.

197. van den Broek EC, Oerlemans S, Nijziel MR, Posthuma EF, Coebergh JW, van de Poll-Franse LV. Impact of active surveillance, chlorambucil, and other therapy on health-related quality of life in patients with CLL/SLL in the netherlands. *Ann Hematol*. 2014.

198. Stryer D, Tunis S, Hubbard H, Clancy C. The outcomes of outcomes and effectiveness research: Impacts and lessons from the first decade. *Health Serv Res*. 2000;35(5 Pt 1):977-993.

199. Dutch Institute for Clinical Auditing (DICA). https://www.clinicalaudit.nl/. Accessed June 29, 2015.

200. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: More than meets the eye. *Lancet Oncol*. 2015;16(4):e173-80.

201. Wang SJ, Lemieux A, Kalpathy-Cramer J, et al. Nomogram for predicting the benefit of adjuvant chemoradiotherapy for resected gallbladder cancer. *J Clin Oncol*. 2011;29(35):4627-4632.

202. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19(4):980-991.