

A Practical Guide for Using Registry Data to Inform Decisions About the Cost Effectiveness of New Cancer Drugs: Lessons Learned from the PHAROS Registry

Hedwig M. Blommestein · Margreet G. Franken ·
Carin A. Uyl-de Groot

Published online: 3 February 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

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 generalizability of the outcomes and their value to decision makers. Multivariate regression modeling, 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 realize that real-world evidence provides extremely valuable and relevant policy information, but needs to be assessed differently compared with evidence derived from a randomized clinical trial.

H. M. Blommestein · M. G. Franken · C. A. Uyl-de Groot (✉)
Department of Health Policy and Management,
Institute for Medical Technology Assessment,
Erasmus University, P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands
e-mail: uyl@bmg.eur.nl

Key points for decision makers

Outcomes of economic evaluations based on registry data are to be assessed differently than economic evaluations based on trial data

Frequently encountered issues, such as confounding by indication, missing values, and insufficient number of comparable patients, need to be adequately addressed to maximize the internal validity

Real-world data provide generalizable outcomes and provide insights into a drug's value for money in daily practice

1 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) modeled from randomized clinical 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 generalizable (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 [1].

Guidelines on conducting and reporting economic evaluations are readily available [2–4], as well as questionnaires to assess the relevance and credibility of observational studies [5]. However, barriers still exist to use evidence from economic evaluations in actual decision making [6, 7]. This necessitates the evaluation of the strengths and limitations of different types of evidence [8]. 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.

2 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 [9, 10]. 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 hematologic malignancies (non-Hodgkin lymphoma, multiple myeloma, and chronic lymphatic leukemia) in three regions; these regions cover

40 % of the Netherlands [11]. PHAROS expanded over the years to other hematological malignancies (chronic myeloid leukemia, 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 [12] 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. [12].

3 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 [2–4] are presented in Table 1.

Table 1 Steps of an economic evaluation

Step	Description
Policy issue	Define the objective of the economic evaluation and ascertain its relevance for healthcare 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

3.1 The Policy Issue

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

3.2 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 regimes, 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 generalizability.

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 generalizable 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.

3.3 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.

3.4 Identify the Comparator(s)

Economic evaluations involve a “comparative analysis of alternative courses of action in terms of both their costs and

consequences” [2]. 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 [12]. The inclusion of control groups to a registry adds to its complexity, time, and costs [1], 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 unfavorable prognostic factors, i.e., carcinoembryonic antigen levels at baseline [13]. 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.

3.5 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 [2]. 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 [14].

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 [1]. 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 centralized treatment. Patients in the PHAROS registry were often discharged from hospital and referred to different rehabilitation centers. 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 [15]. 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 [16, 17].

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 hematologic patients [12, 18]. 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 [12].

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.

3.6 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 [14]. 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 [2]. 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 [19]. Moreover, physicians in daily practice often do not report using standardized response criteria [20], 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 hampered 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.

3.7 Calculate the ICER

This step usually involves modeling methods such as Markov modeling or patient-level simulation modeling [21]. It is important to carefully select the model that best fits the data from the registry [22]. 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 [19]. We used the methods as set forward by Dutch guidelines. Detailed information on the cost-effectiveness calculations performed with PHAROS data is reported elsewhere [12].

3.8 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 uncertainties. This information may be crucial for deciding on the adoption of a new drug. The uncertainty of input parameters can be analyzed by scenario analysis as well as probabilistic and univariate sensitivity analyses [2]. In PHAROS, we observed great patient heterogeneity which resulted, in combination with small numbers of eligible patients treated with the drug of

Table 2 Scenario analysis of the PHAROS economic evaluation [12]

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

Table adapted from Blommestein et al. [12]

ICER incremental cost-effectiveness ratio, QALY quality-adjusted life-year, RCT data from randomized clinical trials, RW data from real-world practice

interest, in wide confidence intervals. In addition, as presented in Table 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 [12]). 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.

3.9 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 [6]. 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 generalizability of the results. The latter two are extremely important to determine usefulness of the results for decision makers [2]. 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.

4 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.

4.1 Confounding by Indication

One of the main concerns about observational data raised in academic literature is the lack of a randomized controlled setting, which results in problems with internal validity [23–25]. 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 [8]. 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 randomization, 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 modeling, propensity score (PS) matching, and data synthesis. Multivariable regression modeling 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 [26]. This technique allows the calculation of the chance of receiving the treatment of

interest by using observed patient characteristics [27]. PS scores are then used to match a treatment group to a comparator group based on patients who have similar chances (PS scores) of receiving the treatment of interest. Other applications of the PS score matching technique include stratification, covariance adjustment, and weighting [27, 28].

Although PS matching techniques are increasingly and successfully used [26, 29], 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 modeling is still needed [26].

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 [30]. 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.

4.2 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 analyzing 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 [31]. 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) [32].

4.3 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 realize 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 [1]). 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 generalizability.

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 [12]. 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 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).

5 The Value of Real-World Economic Evaluations for Decision Makers

Decision makers often make limited use of evidence from economic evaluations [33, 34]. There is, however, a higher chance that decision makers use such evidence if the evidence is accessible (i.e., timeliness and understandability) and acceptable (i.e., accuracy and validity of research methods given institutional requirements) [7].

Above all, it is crucial that decision makers realize 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 generalizability and answer different questions.

The economic evaluation based on PHAROS data demonstrated these differences by calculating different scenarios. Table 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. Randomization 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 the understandability of the results, i.e., economic evaluations based on trial data are intuitive because conventional methods are used. This is, however, at the cost of generalizability, 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 generalizable 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

generalizability and internal validity. The estimated cost-effectiveness ratio (€11,245) also provides reassurance to decision makers that efficacy from the trial can be realized at favorable 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.

6 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 [35]. 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.

7 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 [12, 30, 36]. In PHAROS, both matched and unmatched outcomes seem favorable for the decision to adopt rituximab maintenance therapy. In other cases, the variation in outcomes can be much greater and less favorable 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 [19]. 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 [18, 19, 34]. 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 realize 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 generalizable 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 generalizability. Pragmatic trials have the major advantage of randomizing treatment but are on the other hand, however, associated with logistical, ethical, and sample size challenges as well as high resource investments [37].

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 The authors HB, MF, and CU have no conflict of interest to disclose. HB acts as the overall guarantor for the manuscript content. HB, MF, and CU have all contributed significantly to the design of the study, writing of the manuscript, and approved the final version of the manuscript.

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.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Gliklich R, Dreyer N, editors. Registries for evaluating patient outcomes: a user's guide (prepared by Outcome DEcIDE Center [Outcome Sciences, Inc. d/b/a Outcome] under Contract No. HHS290200500351 TO2). AHRQ Publication No. 10-EHC049. 2nd ed. Rockville: Agency for Healthcare Research and Quality; 2010.
2. 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 Inc.; 2005.
3. 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.
4. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, 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):231–50.
5. Berger ML, Martin BC, Huseareau D, Worley K, Allen JD, Yang W, 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):143–56.
 6. 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. doi:10.1007/s40258-014-0132-7.
 7. Williams I, Bryan S. Understanding the limited impact of economic evaluation in health care resource allocation: a conceptual framework. *Health Policy*. 2007;80(1):135–43.
 8. Dreyer NA, Garner S. Registries for robust evidence. *JAMA*. 2009;302(7):790–1.
 9. Delwel GO. Guidance for outcomes research. 2008;270. Available at <http://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/publicaties/publications-in-english/2008/0812-guidance-for-outcomes-research/0812-guidance-for-outcomes-research/Guidance+for+Outcomes+Research.pdf>.
 10. Kuijpers MR, Toenders WGM. Assessment procedure for inpatient drugs [in Dutch: procedure beoordeling intramurale geneesmiddelen], Report no. 26022597. 2006.
 11. 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;7:321–5.
 12. Blommestein HM, Issa DE, Pompen M, Ten Hoor G, Hogendoorn M, Joosten P, 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.
 13. van Gils CW, de Groot S, Tan SS, Redekop WK, Koopman M, Punt CJ, et al. Real-world resource use and costs of adjuvant treatment for stage III colon cancer. *Eur J Cancer Care (Engl)*. 2013. doi:10.1111/ecc.12154.
 14. Health Insurance Board [in Dutch College Voor Zorgverzekeringen]. Guidelines for Pharmacoeconomic research, updated version. 2006;2014.
 15. van de Poll-Franse LV, Horevoorts N, Eenbergen Mv, Denollet J, Roukema JA, Aaronson NK, 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):2188–94.
 16. 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):1029–36.
 17. Pettengell R, Donatti C, Hoskin P, Poynton C, Kettle PJ, Hancock B, et al. The impact of follicular lymphoma on health-related quality of life. *Ann Oncol*. 2008;19(3):570–6.
 18. Gaultney JG, Franken MG, Tan SS, Redekop WK, Huijgens PC, Sonneveld P, 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):41–7.
 19. Franken MG, van Gils CW, Gaultney JG, Delwel GO, Goettsch W, Huijgens PC, 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.
 20. Franken MG, Gaultney JG, Blommestein HM, Huijgens PC, Sonneveld P, Redekop WK, 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):245–53.
 21. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics*. 2008;26(2):131–48.
 22. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*. 2006;24(4):355–71.
 23. 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.
 24. 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):1053–61.
 25. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA*. 2010;304(8):897–8.
 26. 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):437–47.
 27. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.
 28. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. 2006;163(3):262–70.
 29. 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):819–28.
 30. 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):703–18.
 31. Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
 32. Little R, Rubin D. Statistical analysis with missing data. New York: Wiley; 1987.
 33. 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):358–9.
 34. Franken MG, Koopmanschap M, Steenhoek A. Health economic evaluations in reimbursement decision making in the Netherlands; time to take it seriously? *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2014;108(7):383–9.
 35. 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):411–21.
 36. 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):1673–8.
 37. Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation*. 2008;118(12):1294–303.