Opening up Data Analysis for Medical Health Services: Cancer Survival Analysis with CARESS

David Korfkamp[1], Stefan Gudenkauf[1], Martin Rohde[1], Eunice Sirri[2], Joachim Kieschke[2], and H.-Jürgen Appelrath[1]

[1] OFFIS - Institute for Computer Science, Escherweg 2, Oldenburg, Germany
{david.korfkamp,stefan.gudenkauf,martin.rohde}@offis.de,
appelrath@informatik.uni-oldenburg.de

[2] Epidemiological Cancer Registry Lower Saxony, Oldenburg, Germany
{eunice.sirri,kieschke}@offis-care.de

Abstract. Dealing with cancer is one of the big challenges of the German healthcare system. Originally, efforts regarding the analysis of cancer data focused on the detection of spatial clusters of cancer incidences. Nowadays, the emphasis also incorporates complex health services research and quality assurance. In 2013, a law was enacted in Germany forcing the spatially all-encompassing expansion of clinical cancer registries, each of them covering a commuting area of about 1 to 2 million inhabitants [1]. Guidelines for a unified evaluation of data are currently in development, and it is very probable that these guidelines will demand the execution of comparative survival analyses.

In this paper, we present how the CARLOS Epidemiological and Statistical Data Exploration System (CARESS), a sophisticated data warehouse system that is used by epidemiological cancer registries (ECRs) in several German federal states, opens up data analysis for a wider audience. We show that by applying the principles of integration and abstraction, CARESS copes with the challenges posed by the diversity of the cancer registry landscape in Germany. Survival estimates are calculated by the software package periodR seamlessly integrated in CARESS. We also discuss several performance optimizations for survival estimation, and illustrate the feasibility of our approach by an experiment on cancer survival estimation performance and by an example on the application of cancer survival analysis with CARESS.

Keywords: Data analytics, cancer survival, CARESS, periodR.

1 Introduction

With an estimated annual number of 470,000 incident cases and nearly 215,000 deaths, dealing with cancer is one of the big challenges of the German healthcare system [2,3]. The analysis of cancer data can provide valuable insights on oncological care. Typical analyses of interest are, for example: detecting region-specific changes in the survival of cancer patients which may be attributable

L. Bellatreche and M.K. Mohania (Eds.): DaWaK 2014, LNCS 8646, pp. 382–393, 2014.

to improvements in diagnostics, therapeutics and secondary prevention, and detecting regional and international differences in the survival of cancer patients.

Originally, efforts regarding the analysis of cancer data focused on the detection of spatial clusters of cancer incidences, for example, finding bursts of leukemia in the proximity of nuclear power plants. In 1995, a national law was enacted directing the establishment of population-based cancer registries in all German federal states [4]. However, up until now, federal activities are still isolated and the landscape of cancer survival analysis is still diverse. This also applies to the regional level. For example, certified organ cancer centers,[1] oncology centers and clinics, which treat the majority of cancer patients in Germany, rely on heavily customized software systems with heterogeneous data storage systems, making it even harder to obtain comparable data bases for analysis. Although the Association of Population-based Cancer Registries in Germany (GEKID) provides a coordinated effort to harmonize cancer incidence data collection since 2006, their recommendations are still to be widely implemented and data acquisition, reporting activities, as well as the regulatory frameworks remain inconsistent.

Nowadays, the emphasis also incorporates complex health services research and quality assurance. Additionally, in 2013 a law was enacted forcing the spatially all-encompassing expansion of clinical cancer registries, each of them covering a commuting area of about 1 to 2 million inhabitants [1]. Guidelines for a unified evaluation of data are currently in development, and it is very probable that these guidelines will demand the execution of comparative survival analyses.

In this context of heterogeneity, cancer registries represent a necessity. As data warehouse (DWH) systems [5], they physically integrate cancer data of various formats and stemming from various sources into a single system. They provide an integrated view on population-based cancer data confined to a specific region and appropriate tools to enable their analysis.

There are several software tools for cancer survival analysis, for example SURV3/4 and periodR [6, p. 527ff]. Although proven to be practical regarding applicability [2], most lack in accessibility: The tools are isolated, meaning that the user must provide a prepared dataset of previously selected cancer data beforehand – a task that is notoriously time-consuming and error-prone, requires extensive technical skills, and represents a recurring discontinuity in the digital workflow. Moreover, none of them is particularly suited to generate and publish end-user-friendly reports on-the-fly.

In this paper, we show how specific data warehouse systems can open up data analysis for a wider audience. In an example we show survival analysis on cancer data with CARESS, an epidemiologic cancer registry (ECR) system that is utilized in several federal states in Germany. First, we introduce several methods for the computation of cancer survival estimates. Second, we introduce the CARESS system and its conceptual architecture, including the integration of cancer survival analysis. Next we highlight technical challenges of the implementation

[1] See `http://www.onkozert.de/` [last visited 2014/03/27]

and how we overcame them, especially regarding performance optimization, and present an example on the application of cancer survival analysis with CARESS.

2 Methods of Cancer Survival Analysis

Cancer survival analysis employs statistical methods to analyze cancer data by considering the time period between a defined starting point (e.g., the documented date of initial diagnosis) and the occurrence of an event of interest (e.g., patient death) [7]. Cancer survival estimates can be computed by a variety of methods, and the computation itself can be executed by a variety of software tools. This section presents an overview of the different types of cancer survival analyses and selected tools to perform such analyses.

The first dimension is whether the computation is cohort-based or period-based. The traditional cohort-based approach includes a group of patients in the analysis (i.e., the cohort) by considering a defined period of diagnosis (i.e., years of diagnosis), with all follow-up diagnoses within a defined timeframe [2,6]. Although this approach is considered limited regarding the reflection of recent progress made on cancer care, this shortcoming can be mitigated with complete analysis, a variant of the cohort-based analysis that additionally considers more recently diagnosed patients regardless of the length of follow-up [8]. Period-based analysis, in contrast, is an approach that focuses on information of recently departed patients by applying a survival function to the observed survival experience within a defined timeframe (i.e., the period) to estimate the survival of the patients within this timeframe of follow-up years [2]. In several experiments using historical data, period-based analysis has proven to be more accurate than cohort-based analysis in estimating the chance of survival of more recently diagnosed patients. [2,9,10,11]

The second dimension of cancer survival analysis is whether the computation is absolute or relative. According to Holleczek et al., absolute computation calculates survival in terms of proportions of patients still alive after a given time span after diagnosis, typically reported in 5 or 10-year survival [8]. Relative computation of cancer survival is instead calculated as the ratio of the observed survival in a group of patients and the expected survival of a comparable group considered not to have the cancer of interest in terms of age, sex and calendar period as obtained from population life tables [8]. Thus, the reported survival is corrected for other causes of death than the cancer without requiring detailed information on the exact cause of death.

There exist several methods and tools for estimating survival. Widely accepted methods are Ederer I, Ederer II, Hakulinen and Kaplan-Meier. Tool support can be differentiated into openly accessible software such as SURV-4, or periodR, and proprietary tooling that is directly integrated into ECR specific database systems. For example, Table 1 presents a categorization of the software tool periodR according to Holleczek et al., an open source add-on package to the R programming language and environment for statistical computation. As the table shows, periodR covers the whole range of cancer survival analyses and employs

Table 1. Categorization of the software tool periodR. X = supported. (a) = Ederer II, Hakulinen. (b) = Greenwoods method.

	Absolute	Relative
Period-based analysis	X	X (a)
Complete analysis	X	X (a)
Cohort-based analysis	X	X (a)
Standard-error detection	X (b)	X (b)

widely accepted methods to do so. As it provides an Application Programming Interface (API) naturally, we deem periodR a fit choice to integrate in an ECR system.

3 The CARESS System

The *CARLOS Epidemiological and Statistical Data Exploration System* (CARESS) is a sophisticated data warehouse system that is used by the ECRs in several German federal states. Originally developed in 1993 in the pilot project Cancer Registry Lower-Saxony (CARLOS) serving as a geographic information system (GIS) tool for analyzing clusters of cancer incidences [12,13], the CARESS system was subsequently extended into a full-fledged ECR data warehouse system and adopted by the federal states of Hamburg, Schleswig-Holstein and North Rhine-Westphalia and the center for cancer registry data (ZFKD) at the Robert-Koch-Institute, which pools data from all federal states in Germany. The system supports several stakeholders in medical health services such as doctors and epidemiologists by providing sophisticated tools for data analysis in a highly accessible user interface, enabling them to carry out explorative analyses, ad-hoc queries and reporting activities without extensive technical skills.

CARESS consists of three layers. The *data source integration layer* provides a unified physical integration of various heterogeneous data sources. CARESS supports DWH products from different vendors such as Microsoft Analysis Services, Pentaho Mondrian and Jedox Palo.

The *component integration layer* provides a service facade to client applications in order to invoke the services supported by CARESS, see Figure 1. Complex statistical queries are executed by outsourcing requests, for example, to automate the calculation of cancer survival estimates using the R programming language and environment. Requests to the underlying data source integration layer are encapsulated as services as well. Additionally, the component integration layer provides access to the system's metadata repository containing complex analysis configurations. As the service factory only provide access to service interfaces instead of concrete services, the respective service implementations can be exchanged easily. For example, we introduced an optimized CachedSurvivalAnalysisDataService (see Section 4) that maintains an instance of the original SurvivalAnalysisDataService to forward any request not previously cached to this instance.

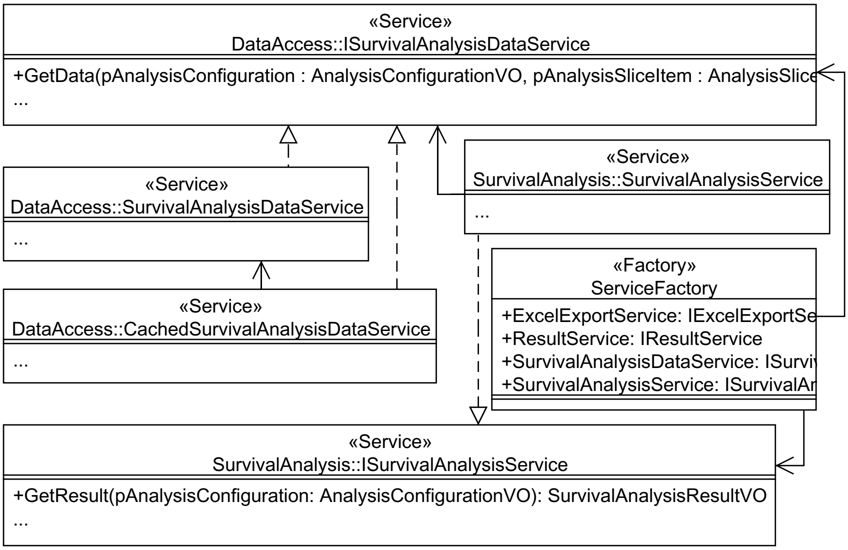

Fig. 1. Service classes of the CARESS component integration layer

The *client layer* provides a convenient graphical user interface (GUI) to access the service facade of the component integration layer. The GUI is realized using Microsoft Windows Presentation Foundation (WPF) and the Prism framework for modular applications. [14] Also, the GUI supports arbitrary inquiries (e.g., constructing a pivot table by combining arbitrary summary attributes with arbitrary compatible dimensions) as well as topic-oriented analyses with predefined statistical methods such as cancer survival analysis. This enables epidemiologists to carry out semi-automated explorative analyses and ad-hoc queries with minimal effort. In addition, CARESS offers a rich set of reporting functions for end users such as doctors and nurses, and supports different export formats such as XML, CSV, Excel and PNG image files.

In contrast to the use of individual tools for estimating survival analyses, requiring the user to provide a prepared dataset of previously selected cancer data beforehand, CARESS provides an experience that enables the user to focus on the task at hand without distraction. Setting up CARESS is a one-time-effort of DWH engineers, while operation is guaranteed by DWH and software engineers, depending on the question which architecture layer has to be adapted or extended with additional functionality.

4 Optimizing CARESS for Survival Analyses

Survival analyses impose particular challenges regarding data acquisition. First, data required to run survival analyses should be highly detailed. For example, the diagnosis date of cancer cases should be available at least in terms of months to achieve a high precision of survival estimates.

Second, CARESS uses a single database as its *single point of truth* that conforms to the Online Analytical Processing (OLAP) cube paradigm. While the OLAP paradigm proves powerful when navigating through multidimensional data spaces, it is rather limited regarding the acquisition of heterogeneous data at high levels of detail compared to the conventional relational database model. For example, a typical 3-year period analysis with 5 years of follow-up diagnoses results in a 5-dimensional cube, consisting of approximately 1,400,000 cells (100 distinct age categories $\times(3+5)\times 12$ diagnosis months $\times 3\times 12$ death months $\times 2$ gender categories $\times 2$ vital status), resulting in comparatively large requests in contrast to an equivalent relational database request.[2] As a consequence, OLAP result sets for highly detailed survival analysis are typically large and require much time and memory for processing.

To address these challenges we enhanced CARESS' *SurvivalAnalysisDataService*, the component responsible for retrieving data from the underlying multidimensional database, and integrated the following adaptations for survival analyses in contrast to the regular DataService.

First, we optimized the Multidimensional Expressions (MDX) query used to retrieve data from the OLAP database by applying the *NON EMPTY* keyword to each axis. Axes tagged *NON EMPTY* contain only classification nodes with actual values with respect to the other axes. For example, if a query result contains no single cancer case for a patient who at the time of diagnosis was 40 years old, the result will not include the 40 years node from the age dimension at all, although this age was requested in the analysis. Depending on the actual distribution of cases this measure can significantly decrease the size of the returned result. However, the optimization is likely to go unnoticed for analyses that are performed on large areas, since most combinations of age, gender, date of diagnosis, date of death, and vital status include at least one case.

Second, we extended the SurvivalAnalysisDataService with functions to split, parallelize and merge query requests. The actual implementation operates as follows: (1) Candidate dimensions are identified for splitting the request into several smaller requests. In general, all classifying dimensions of an analysis (age, gender, vital status, date of diagnosis, and date of death) are considered as to be candidates. However, in certain situations some of those dimensions can not be used for splitting. For example, this is the case when a classifying dimension was selected as a *slicer dimension* (e.g., for retrieving age- or gender-specific survival estimates – in this case we only retrieve the data of the selected slice). (2) Of the remaining candidates up to two dimensions are selected automatically as *split dimensions*. In case of two split dimensions, partial cubes are retrieved based on the cross product of each classification node of the two dimensions.

[2] A relational database request would result in only a few hundred rows, when, for example, a rare diagnosis or a specific regional area is analyzed.

The resulting partial cubes are then being requested in parallel from the underlying multidimensional database in order to reduce the overall request time. Once all partial cubes are retrieved, they are merged into a single result cube available for further computations.

Third we introduced a caching algorithm in order to reduce the speed of subsequent survival analysis requests that are based on the same data. For example, these can occur when the statistical method for survival analyses is changed, the life table is exchanged, or a previously rendered survival analysis is exported to Microsoft Excel.[3] To do so, we used the Microsoft .Net Framework's native MemoryCache class to store results from the SurvivalAnalysisDataService, since it already provides functionalities for caching – for example, the ability to let cache values expire after a predefined period. To store and access the cache we derived a normalized key from the request object used to interact with the SurvivalAnalysisDataService. This key consists of sorted lists of the classifying and restricting dimensions containing the selected classification nodes (also sorted). As a result, small changes to queries such as the rearrangement of classification nodes or slicer axes do not result in new database requests.

In the following, we describe an experiment on the response-time of cancer survival analysis to illustrate the effectiveness of our optimizations.

4.1 Experimental Setup

The experiment was conducted on a single machine running the complete CARESS stack, including the database server, to minimize external effects. The machine was equipped with a Dual-Core Opteron 2220 processor clocked at 2,6 GHz and with 8 GB RAM. Microsoft SQL Server 2012 Analysis Services were used as the database backend. We ran several tests in which always the same realistic example configuration was computed. After every computation we restarted CARESS to avoid interfering effects introduced by caching. For each run we recorded the response-time using CARESS' builtin logging mechanism. Thereby, the measured response-time represents the effects notable by the user. It includes the time required to retrieve the data and subsequent activities such as statistical calculations in the R component and client-side rendering.

We conducted 15 survival analyses for the original and for the optimized survival analysis each. We used the arithmetic mean and the 95% confidence interval (CI) for the results of the analyses. The CIs allow us to address two questions: (1) Do the different implementations show significantly different behavior or not? (2) How large is the performance variation of the individual measurements for a single implementation? The CIs are computed using the Students t-distribution, as the number of measurements is small ($n < 30$). We refer to the work of Georges et al. as an excellent reading on the importance of confidence intervals for statistically rigorous performance evaluation. [15]

[3] Exporting a survival analysis to Excel is performed by the *ExcelExportService*, which uses the *CachedSurvivalAnalysisDataService* via the*ResultService*.

4.2 Results

As Figure 2 shows, the mean response-times of both implementations differ and the 95% confidence intervals (CI) do not overlap, showing that the difference between the measurements is statistically significant (optimized analysis: CI 95% max. 10.493, original analysis: CI 95% min. 381.567), indicating that the optimized version is much faster. Furthermore, the CI show that the variance of the measurements of the original implementations is much greater than the variance of the optimized version, indicating a more stable behavior.

However, any empirical study like ours is vulnerable to certain threats to validity. For example, the experiment has only been executed on a single machine and wider applicability is yet to prove. Also, the optimized survival analysis is only compared to its functionally equivalent original version. Comparative analyses to other survival analysis tools are desirable. On the other hand, to the best of our knowledge, there are currently no other implementations for survival analysis that comprise both the actual analysis as well as automatized data preparation and retrieval. We leave these topics amongst others for future work (i.e., more experiments).

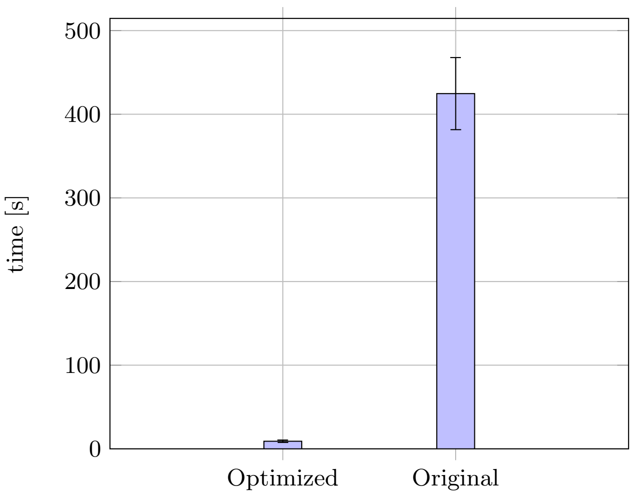

Fig. 2. Response-time of the optimized and the original survival analysis implementation. The optimized implementation required 9.07 seconds in average (CI 95% 1.427), the original approach required 424.67 seconds in average (CI 95% 43.100).

5 Application of Cancer Survival Analysis with CARESS

In the following, we describe how cancer survival estimates are computed with CARESS by the example of the ECR in Lower Saxony (Epidemiologisches Krebsregister Niedersachsen (EKN)). The system integrates data from various data sources into a data warehouse, including cancer incidence data along with date of diagnosis, date of death (if applicable), vital status, gender, diagnosis, death certificate only (DCO) cases etc. gathered regularly at EKN. Typical stakeholders to the system are epidemiologists that prepare mandatory annual reports on cancer and ad-hoc reports for governmental requests and journalistic inquiries. Since 2003, EKN has comprehensively covered the state of Lower Saxony in Germany, inhabited by approximately eight million people. The completeness of registration was reported to be over 95% in 2010 [16].

The minimum data requirements for survival analysis include sex, month and year of diagnosis (*dm* and *dy*), age at diagnosis, month and year of end of follow-up (*fm* and *fy*) and vital status at the end of follow-up. A detailed specification of the minimum data requirements and the concrete periodR functions are described by Holleczek et al. [8]

CARESS calculates cancer survival estimates by performing a three-stage procedure. The software component that manages these three stages was extensively empirically evaluated to guarantee that survival estimates and the corresponding plotted survival curves are correct. Stage one includes querying the required data from the data warehouse by narrowing the data space according to the user input. In general, cancer survival estimation requires particular attention to define the temporal dimensions date of diagnosis (*dm* and *dy*) and end of follow-up (*fm* and *fy*). CARESS reduces the effort needed by considering a higher level of abstraction: the user defines a period (period approach) or a cohort (cohort approach) of interest, and a number of follow-up years. The parameters *dm* and *dy* are then constrained to a range based on either the cohort or the period, while *fm* and *fy* are calculated by CARESS based on the corresponding date of death, including all cases that either have no date of death at all (representing patients that were still alive at the end of follow-up) or died in the course of the follow-up years.

The second stage performs a transformation of the retrieved data to meet the aforementioned minimum data requirements for periodR. For example, cases that are still alive at the end of the follow-up period are assigned the end of follow-up dates (i.e., fm and fy). The actual transformation is executed by a software component within CARESS. Inconclusive data is excluded, for example, cases with unknown month of diagnosis and unknown month of death, or implausible dates. Excluded datasets are logged in a separate file for later examination.

In the final stage the prepared dataset is handed to the periodR component. The results returned include a chart showing the absolute and relative survival rates by follow-up years as well as tables that show the survival estimates along with 95% confidence intervals and standard errors. Both are visualized directly in the CARESS client. The chart is illustrated in Figure 3. For convenient reporting, results can be exported to different formats such as Excel and PNG image files.

As an example, we illustrate cancer survival analysis with CARESS using a dataset that includes 33,611 records of lung cancer patients aged 15-99 years, diagnosed in 2003-2010 with passive mortality follow-up until December 2010 and for the period 2008-2010. The event of interest considered for survival estimates

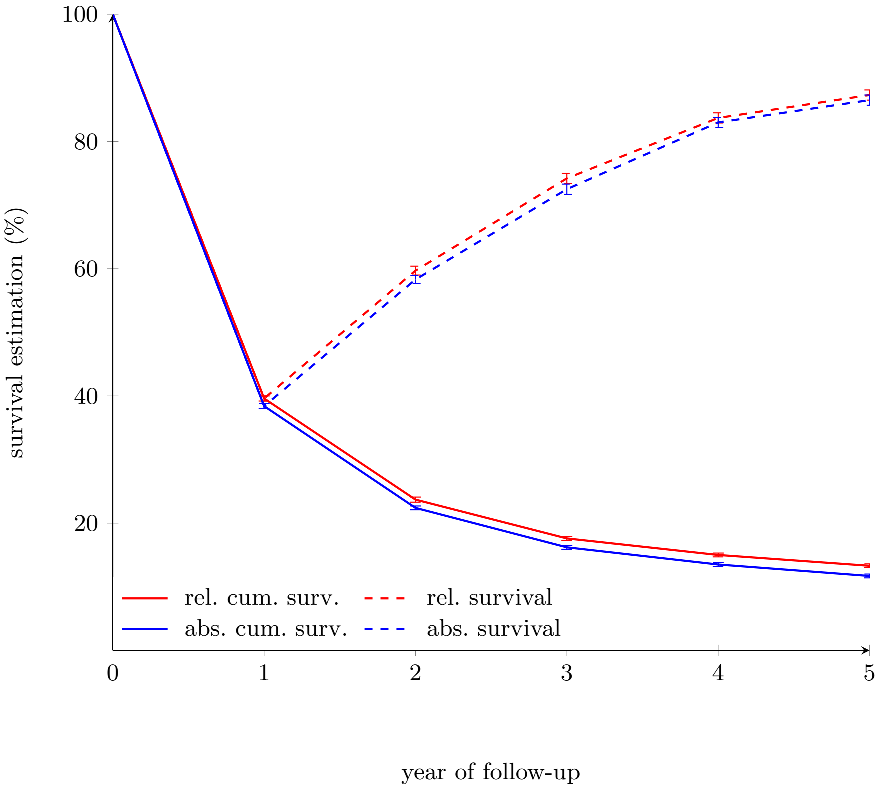

Fig. 3. CARESS analysis report on cumulative absolute and relative survival estimates of breast cancer patients grouped into three age categories for five years of follow-up for patients diagnosed in 2003-2010, with mortality follow-up until 2010 and for period 2008-2010.

was the death of the patient. Therefore, patients still alive at the end of 2010 were right censored: For these patients, fm and fy were imputed as 12 and 2010, respectively, automatically by the CARESS system. Addressing further data quality concerns such as the exclusion of DCO cases can be performed by the user by simply removing the corresponding data entries from a filter view within the user interface of CARESS. The user interface also lets the user choose the actual analysis methods supported by periodR (see Figure 4).

A particular advantage of CARESS for survival analysis is the high degree of automation. For example, computing survival estimates by prognostic variables (e.g., sex, age groups, stage of disease at diagnosis, histology, anatomic subsite) is executed automatically once the user has selected the respective variables in the user interface. Table 2 illustrates the results for absolute and relative survival estimates of the cohort of patients diagnosed in 2003-2010 and of the period 2008-2010, as analyzed with CARESS. The estimation is ordered by sex.

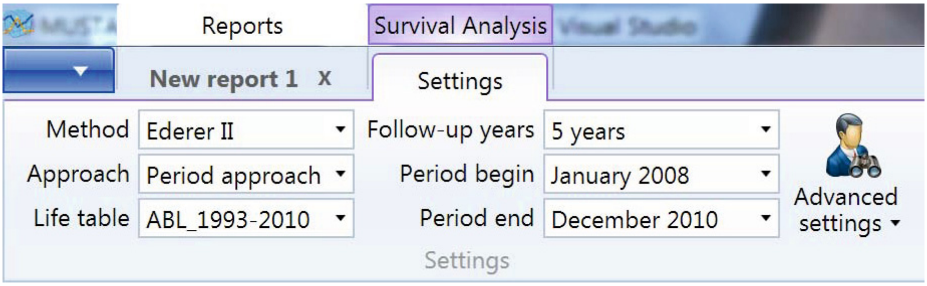

Fig. 4. Graphical user interface for survival computation in CARESS at the ECR of Lower Saxony, Germany

Table 2. Absolute and relative survival estimates by sex for lung cancer patients diagnosed in 2003-2010 in EKN, period 2008-2010

Follow-up year	Male		Female	
	absolute (SE*)	relative (SE*)	absolute (SE*)	relative (SE*)
1	42.9 (0.5)	44.1 (0.5)	48.4 (0.8)	49.4 (0.8)
2	24.5 (0.5)	25.9 (0.5)	29.5 (0.7)	30.6 (0.7)
3	17.7 (0.4)	19.1 (0.4)	22.0 (0.7)	23.2 (0.7)
4	14.4 (0.4)	16.2 (0.4)	18.4 (0.6)	19.7 (0.7)
5	12.3 (0.3)	14.2 (0.4)	16.5 (0.6)	18.0 (0.7)

6 Conclusions and Outlook

Cancer epidemiology is an explorative art on the one hand and uses complex statistical methods like survival analyses on the other hand. A sophisticated multidimensional data model for data warehouse systems in health care must provide integration of statistical methods and definition of ad-hoc aggregations at run-time. In contrast to standard OLAP tools and standard statistical tools the CARESS system provides sophisticated mechanisms to integrate domain-specific statistical methods into the multidimensional data model and makes them available for epidemiologists and scientists via a convenient graphical user interface. Additionally, CARESS is specifically optimized for OLAP-based survival analysis. This is illustrated by experiments on the response-time of cancer survival analyses.

Further extension of CARESS comprise the implementation of Cox regression methods and more experiments. In addition, the EKN reckons with being assigned the task of analyzing data from all future clinical cancer registries in Lower Saxony, comprising a population of about 8 million people. An corresponding extension of the CARESS system is subject of our current efforts in anticipation of this task. With this extension, we deem CARESS an appropriate candidate as the DWH system of choice for regional clinical cancer registries.

References

1. Gesetz zur Weiterentwicklung der Krebsfrüherkennung und zur Qualitätssicherung durch klinische Krebsregister (Krebsfrüherkennungs- und -registergesetz – KFRG) of 2013, BGBl. I, pp. 617–623 (April 3, 2013) (German)
2. Nennecke, A., Brenner, H., Eberle, A., Geiss, K., Holleczek, B., Kieschke, J., Kraywinkel, K.: Cancer survival analysis in Germany – Heading towards representative and comparable findings. Gesundheitswesen 72, 692–699 (2010) (German)
3. Robert Koch-Institut und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e. V., editors. Krebs in Deutschland 2007/2008, Berlin (2012) (German)
4. Gesetz über Krebsregister (Krebsregistergesetz – KRG) of 1994, BGBl. I pp. 3351–3355 (November 4, 1994) (German)
5. Bauer, A., Günzel, H.: Data Warehouse Systeme, 3rd edn. dpunkt.verlag, Heidelberg (2009) (German)
6. Kleinbaum, D.G., Klein, M.: Survival Analysis, 3rd edn. Springer, New York (2012)
7. Altman, D.G., Bland, J.M.: Time to event (survival) data. British Medical Journal 317, 468–469 (1998)
8. Holleczek, B., Gondos, A., Brenner, H.: periodR – an R Package to Calculate Long-term Cancer Survival Estimates Using Period Analysis. Methods Inf. Med. 48, 123–128 (2009)
9. Brenner, H., Söderman, B., Hakulinen, T.: Use of period analysis for providing more up-to-date estimates of long-term survival rates: Empirical evaluation among 370 000 cancer patients in Finland. Int. J. Epidemiol. 31, 456–462 (2002)
10. Talbäck, M., Stenbeck, M., Rosén, M.: Up-to-date long-term survival of cancer patients: An evaluation of period analysis on Swedish Cancer Registry data. Eur. J. Cancer 40, 1361–1372 (2004)
11. Ellison, L.F.: An empirical evaluation of period survival analysis using data from the Canadian Cancer Registry. Ann. Epidemiol. 16, 191–196 (2006)
12. Wietek, F.: Spatial Statistics for Cancer Epidemiology – the Cancer Registry's Epidemiological and Statistical Data Exploration System (CARESS). In: Fehr, R., Berger, J., Ranft, U. (eds.) Environmental Health Surveillance. Fortschritte in der Umweltmedizin, pp. 157–171. ecomed-Verlag, Landsberg (1999)
13. Kamp, V., Sitzmann, L., Wietek, F.: A spatial data cube concept to support data analysis in environmental epidemiology. In: Proceedings of the 9th International Conference on Scientific and Statistical Database Management, August 11-13. IEEE, Olympia (1997)
14. Brumfield, B., Cox, G., Hill, D., Noyes, B., Puleo, M., Shifflet, K.: Developer's Guide to Microsoft Prism 4: Building Modulare MVVM Applications with Windows Presentation Foundation and Microsoft Silverlight. Microsoft Press (2010)
15. Georges, A., Buytaert, D., Eeckhout, L.: Statistically Rigorous Java Performance Evaluation. ACM SIGPLAN Notices - Proceedings of the 2007 OOPSLA Conference 42(10), 57–76 (2007)
16. Registerstelle des Epidemiologischen Krebsregisters Niedersachsen, editor. Krebs in Niedersachsen, Oldenburg, 117 p. (2010) (German)