

SOFTWARE

Open Access

# Using relative survival measures for cross-sectional and longitudinal benchmarks of countries, states, and districts: the BenchRelSurv- and BenchRelSurvPlot-macros

Christian O Jacke<sup>1\*</sup>, Iris Reinhard<sup>1</sup> and Ute S Albert<sup>2</sup>

## Abstract

**Background:** The objective of screening programs is to discover life threatening diseases in as many patients as early as possible and to increase the chance of survival. To be able to compare aspects of health care quality, methods are needed for benchmarking that allow comparisons on various health care levels (regional, national, and international).

**Objectives:** Applications and extensions of algorithms can be used to link the information on disease phases with relative survival rates and to consolidate them in composite measures. The application of the developed SAS-macros will give results for benchmarking of health care quality. Data examples for breast cancer care are given.

**Methods:** A reference scale (expected, E) must be defined at a time point at which all benchmark objects (observed, O) are measured. All indices are defined as O/E, whereby the extended standardized screening-index (eSSI), the standardized case-mix-index (SCI), the work-up-index (SWI), and the treatment-index (STI) address different health care aspects. The composite measures called overall-performance evaluation (OPE) and relative overall performance indices (ROPI) link the individual indices differently for cross-sectional or longitudinal analyses.

**Results:** Algorithms allow a time point and a time interval associated comparison of the benchmark objects in the indices eSSI, SCI, SWI, STI, OPE, and ROPI. Comparisons between countries, states and districts are possible. Exemplarily comparisons between two countries are made. The success of early detection and screening programs as well as clinical health care quality for breast cancer can be demonstrated while the population's background mortality is concerned.

**Conclusions:** If external quality assurance programs and benchmark objects are based on population-based and corresponding demographic data, information of disease phase and relative survival rates can be combined to indices which offer approaches for comparative analyses between benchmark objects. Conclusions on screening programs and health care quality are possible. The macros can be transferred to other diseases if a disease-specific phase scale of prognostic value (e.g. stage) exists.

**Keywords:** Benchmark\*, Prevention & control, Outcome assessment, Relative survival, Registries, Breast cancer

\* Correspondence: [Christian.jacke@zi-mannheim.de](mailto:Christian.jacke@zi-mannheim.de)

<sup>1</sup>Central Institute of Mental Health, Medical Faculty Mannheim/University Heidelberg, Square J5, 68159 Mannheim, Germany  
Full list of author information is available at the end of the article

## Background

In many diseases early detection and secondary prevention are of central importance [1-3]. Life-threatening malignant diseases are therefore divided into stages according to their disease phase. They are recorded using international nomenclatures (e.g. Union for International Cancer Control, American Joint Committee in Cancer) [4,5]. Thus, screening programs are aimed at discovering diseases in as many patients and as early as possible to increase their chance of survival.

First approaches on how the number of patients per disease phase/stage can be linked to the corresponding chance of survival are based on Beatty et al. [6,7]. They developed a series of benchmark-algorithms that address different aspects of health care. Beatty and colleagues described a screening index based upon the sum of the products of the stage number (0–4) and the number of cases at that stage divided by the total number of cases. Using the stage number in the calculation was considered arbitrary and it was replaced by the national 5-year mortality for that stage termed the case-mix index for the institution or region. This was then standardized by comparison to the national case-mix index and termed the standardized case-mix index (SCI). They also described a standardized work-up index (SWI) to address the issue of ‘upstaging’ of cases and a standardized treatment index (STI) to evaluate outcome using institutional or regional mortality (stage and overall) compared to national mortalities. The product of the SCI, SWI and STI defined the composite measure named overall performance evaluation (OPE). They used these indices and evaluations to compare different regions over the same time interval (cross-sectional analysis perspective) and the same institution across different time intervals (longitudinal analysis perspective).

However, the screening-index was not standardized, reflected a negative association with the observation period, and, thus, was not included in the OPE. In addition, the OPE had a logical weakness, because the standardized case-mix index was present in the denominator as well as the numerator and cancelled itself out. Finally, the approach was based on disease-specific cause of death statistics which required the modeling of competing risks according to the method of Gooley [8]. This method focuses on the mortality of one disease only and, therefore reduces the case numbers. This is of special importance for small geographic units that already have small case numbers as it is. This outcome-indicator also does not include background-mortality.

## Objectives

This contribution intends to overcome the existing shortcomings of the proposed algorithms and to extend the indices (see methods). Furthermore, the automation of

the benchmark-algorithms (SAS-macros) will standardize the screening-index, produce a positive association with the observation time after appropriate mathematical conversion, and will adequately assess a composite measure. Disease-specific fatality rates will be replaced by relative survival rates as outcome-indicators. The macro results will be demonstrated by exemplary applications. The development of the indices and the main working hypothesis is built on the assumption that comprehensive early detection programs are capable of detecting diseases early, leading to disease stage shifts and facilitates clinical work-up which is associated with increasing relative 5-year survival rates. Improvements of health care quality should be illustrated transparently.

## Implementations

### Data

In general, population based databases with demographic information such as registers (e.g. cancer register) may be used. Clinical registers, cohort studies, or health care network databases are also suited if a high data quality and epidemiologic relevance is ensured [9-11]. The application example is based on the assumption that data of the Surveillance Epidemiology and End Results (SEER) [12] and the Norwegian cancer register [13] fulfill these requirements.

### Variables

The variables for the “best-performer“, that must be used to compare all others, define the reference object. All other comparative objects are designated as benchmark objects and must have identification numbers that can definitely be distinguished from one another. In addition, information on the year of first definite diagnosis (incidence year) and the disease phase (e.g. for cancer stage 0-IV) is necessary. The absolute and relative distribution of persons per disease phase (e.g. per stage) as well as the corresponding relative survival rates (e.g. 5 years) are needed. In addition, the total number of ill patients per time unit (e.g. year) and the relative overall survival are important. Table 1 serves as an example what specifically the SAS-Macro *BenchRelSurv* expects as a variable-set in the so-called long-format of SAS.

### Relative survival as outcome

If actually observed and expected chances of survival probabilities are related to each other, relative survival rates are obtained [14]. The former originate from empirical data (e.g. registers). But in contrast, the latter can be calculated from so-called “prospective probability of death” from period or cohort life tables stratified according to age, gender, calendar year and occasionally also ethnicity [15,16]. These life tables are available from the Federal Agency for Statistics or the Human Mortality

**Table 1 Configuration of Data and Variable Sets**

Group	Year	Identification	Stage	Number per Stage	Number Total	Relative Survival
RefGroup	RefYear	RefID	RefStage	RefNpStage	RefN	RefRelSurv
SEER17	1999-2003	1	0	167	191771	93.6
SEER17	1999-2003	1	1	83081	191771	100.00
SEER17	1999-2003	1	2	72195	191771	90.2
SEER17	1999-2003	1	3	12617	191771	61.3
SEER17	1999-2003	1	4	8449	191771	22.5
SEER17	1999-2003	1	5	15262	191771	77.3
SEER17	1999-2003	1	6	191771	191771	89.1

Legend: (Prefix Ref) Definition for reference object substitutable by prefix 'Bench' indicating benchmark object, (Stage) Stage of disease where 5='not known' and 6='total' overall respectively.

Source: New malignant breast cancer cases from Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER\*Stat.

Database: Incidence - SEER 17-Nov 2010 Sub (1973-2008 varying).

Database (HMD). Different methods for the handling of censored cases [17,18] or various observation periods [19-21] are possible. Non-parametric methods to derive relative survival rates are sufficient in this case and can be easily estimated using freely accessible software-solutions (e.g. [22]).

**Standardized indices**

All indices have the principle of construction in common. Time and factually-fixed reference objects define the comparative scale (here USA, 2003). This, so called reference object is the best possible expected result (expected, Exp\*). All other benchmark objects (observed, Obs) must be compared to this reference scale. Hence, all indices are defined as: index = Obs/Exp\*. From a cross-sectional analysis perspective, the benchmark objects (Obs) belong to the same time interval as the reference object (Exp\*). From a longitudinal analysis perspective, however, the benchmark objects can originate from different time intervals (Obs\_t), while the reference object is fixed in time. To simplify the matter, the time index t is omitted in the following.

**Extended standardized screening index (eSSI)**

The central idea of the eSSI is focused on the relative proportion of ill persons per disease stage (N\_i/N) weighted by the stage number itself. The products are then summed up for the benchmark object. The sum is then put into relation with the sum of the reference object which characterizes the standardization-process. The index is defined as:

$$eSSI = O/E* = \left[ \sum (N_i/N) \times i \right]_O / \left[ \left( \sum (N_i/N) \times i \right) E* \right]$$

**Standardized case-Mix index (SCI)**

The central idea of the SCI is to multiply the absolute number of ill persons per disease stage (N\_i) with the relative survival rate (RSR\_i). The products are then

summed up and divided by the total number of ill patients. The index is standardized by comparing benchmark to reference objects and is defined as:

$$SCI = O/E* = \left( \left( \sum N_i \times RSR_i \right) / N \right)_O / \left( \left( \sum N_i \times RSR_i \right) / N \right)_{E*}$$

**Standardized work-Up index (SWI)**

The central idea of the SWI is to relate the relative survival rate per stage (RSR\_i) of a benchmark object (Obs) with the RSR\_i of the reference object (E\*). The resulting proportions are then summarized. Finally, to get an idea of the average relative survival rate across the stages, the sum is divided by the number of represented disease stages i. The same is true for the reference object. This index is defined as:

$$SWI = O/E* = \left( \sum (RSR_i)_O \right) / \left( \sum (RSR_i)_{E*} \right) / N(i)$$

**Standardized treatment index (STI)**

The central idea of the STI is to set the overall relative survival rate (RSR) of benchmark objects (O) and reference objects (E\*) in relation to each other. However, since benchmark objects and reference objects can differ in their stage distribution, the SCI is needed as an indicator of risk adjustment. The index is defined as:

$$STI = O/E* = (RSR_{Obs} / RSR_{Exp*}) \times 1 / SCI_{obs}$$

**Composite measures**

According to Beatty et al. [6,7] SCI, SWI, and STI may be summarized to an overall performance evaluation (OPE): OPE= SCI × SWI × STI. As an alternative the relative overall performance index (ROPI) is suggested. The ROPI is defined as:

$$ROPI = 1/eSSI \times SWI \times STI$$

### Example of use

The data of new malignant breast cancer patients from the Surveillance Epidemiology and End Results (SEER17-Nov2010) [12] and the Norwegian cancer register [13] were used. The national SEER17-values from 1999–2003 served as reference objects in the cross-sectional analyses. Thus, these analyses were restricted to the period from 1999–2003 and the seventeen SEER-registers which define the benchmark objects. In the longitudinal analyses, the national SEER17-values from the last available year 2003 served as reference object. The national SEER17 data from 1990–2003 as well as the Norwegian data from the intervals 1969–73, 1974–78, 1979–1983, 1984–88, 1989–93, 1994–98, and 1999–03 [13] served as benchmark objects. The relative 5-year survival rate was calculated according to the Ederer II-method. The examples are available within the additional files (Additional file 1: Example1, Additional file 2: Example2) and can be downloaded from the project-homepage (<http://sourceforge.net/projects/benchrelsurv/>).

### Results

In the cross-sectional and longitudinal application examples, the reference object was also included as benchmark object. This leads to a special case because benchmark object and reference object are equal. Therefore, the corresponding indices are of value one in cross-sectional analysis. In longitudinal analyses, it leads to index values of one in the chosen reference year (here 2003). This logic allows for all the other benchmark objects that a health service quality gap – or advance – is identified by index values smaller or larger than one. If, for example,  $eSSI > 1$  is valid, more patients will be treated at a *later* time point in the benchmark object than in the reference object. If  $eSSI < 1$  is valid, more people will be treated at an *earlier* time point than in the reference object. The latter result might be interpreted as a more effective early detection and screening program than in the reference object.

In analogy, for  $SCI > 1$ , the survival conditions in the benchmark objects will adapt to the normal population's *faster* than in the reference object. If  $SCI < 1$  is measured, the survival conditions in the benchmark object will adapt more *slowly* to the conditions of the population than in the reference object. The latter result might be interpreted as a less effective early detection and screening program than in the reference object.

If  $SWI > 1$  is true, then the resumed average survival rates across the stages will be *higher* in the benchmark object than in the reference object. If  $SWI < 1$  is true, then the average survival rates across the stages will be *lower* in the benchmark objects than in the reference object. The latter result might be interpreted as a less effective clinical work-up than in the reference object.

If  $STI > 1$  is true, then the stage adjusted overall survival rates in the benchmark object will be *higher* than in the reference object. If  $STI < 1$  is true, then the stage adjusted overall survival rates in the benchmark object are *lower* than in the reference object. The latter result might be interpreted as a less effective overall treatment than in the reference object.

If OPE or  $ROPI > 1$  is valid, then patients in the benchmark object will have *earlier* treatment with *higher* relative survival rates on average than in the reference object. If OPE or  $ROPI < 1$  is true, then patients in the benchmark object will have *later* treatment with *lower* relative survival rates on average than in the reference object. Table 2 shows the cross-sectional results. Examples of longitudinal results for the eSSI and ROPI are depicted in the Figures 1 and 2. The interpretation follows the general instructions.

### Discussion

The existing, updated and extended standardized indices form an additional tool for the evaluation of health care services and quality. Reference objects can be defined and compared to benchmark objects such as countries, states, and districts. The indices offer a cross-sectional and longitudinal perspective on benchmark objects. Especially the latter offer the opportunity to demonstrate the relationship between disease stage and chance of survival during the course of time.

### Comparison with other benchmarking-algorithms

The performed cross sectional analysis detects the already observed high variability between urban-metropolitan areas and rural regions which have led to controversial discussions [23,24]. The longitudinal analyses, however, may show the growing influence of comprehensive early detection and screening-programs and -methods, guidelines as well as increasing utilization and participation rates that may lead to more favorable surrogate parameters such as absolute or relative stage distribution. In this respect the approach is very similar to purely descriptive benchmark projects [2,3,25-28], which also have to interpret obtained results within country- and time-specific conditions. From this perspective both approaches (presented benchmark, descriptives) are somewhat complementary because the latter may provide insights in explicitly measured and process-related quality indicators which are based on medical decisions. In exchange, the formulated benchmark algorithms reject the use of criteria-based approaches to estimate theoretically expected cases [1,5,29] in the indices' denominator. Therefore, the here presented algorithms cannot inform about insufficient respectively inappropriate health care. Furthermore, the proposed approach cannot revise short-term

**Table 2 Cross-Sectional Benchmarking of SEER17-Registers (1999–2003)**

	eSSI	SCI	SWI	STI	OPE	ROPI
Reference Object (1999–2003)						
SEER 17	1.000	1.000	1.000	1.000	1.000	1.000
Benchmark Objects						
Iowa	0.951	0.993	1.032	1.023	1.048	1.110
New Mexico	0.973	0.959	1.023	1.037	1.017	1.090
Seattle (Puget Sound)	0.960	1.035	1.033	1.004	1.074	1.080
New Jersey	0.991	0.938	0.997	1.052	0.983	1.059
Utah	0.991	0.987	1.026	1.010	1.022	1.045
San Francisco-Oakland (SF)	1.002	1.036	1.039	0.987	1.062	1.023
San Jose-Monterey (SJM)	0.983	1.049	1.009	0.991	1.049	1.018
Conneticut	0.973	1.020	0.985	1.000	1.005	1.013
California excl. SF/SJM/LA	0.994	1.018	1.006	0.990	1.014	1.002
Alaska Natives	0.981	0.932	0.956	0.980	0.874	0.955
Atlanta	1.024	1.001	0.982	0.982	0.965	0.941
Los Angeles (LA)	1.051	1.018	1.008	0.978	1.003	0.938
Kentucky	0.965	0.927	0.862	1.045	0.834	0.933
Detroit (Metropolitan)	1.029	0.981	0.935	0.995	0.913	0.905
Hawaii	0.975	1.078	0.926	0.948	0.947	0.900
Lousiana	1.074	0.960	0.952	0.975	0.891	0.865
Rural Georgia	0.947	0.884	0.709	1.076	0.674	0.805

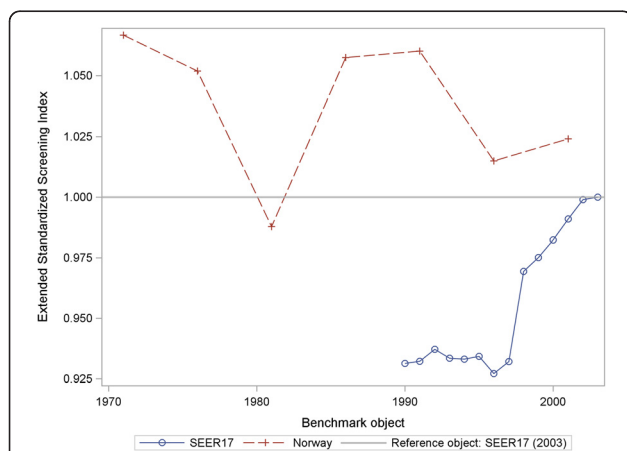
Legend: (eSSI) Extended Standardized Screening Index, (SCI) Standardized Case-Mix Index, (SWI) Standardized Work-Up Index, (STI) Standardized Treatment Index, (OPE) Overall Performance Evaluation, (ROPI) Relative Overall Performance Index.

Source: New malignant breast cancer cases from Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) SEER\*Stat.

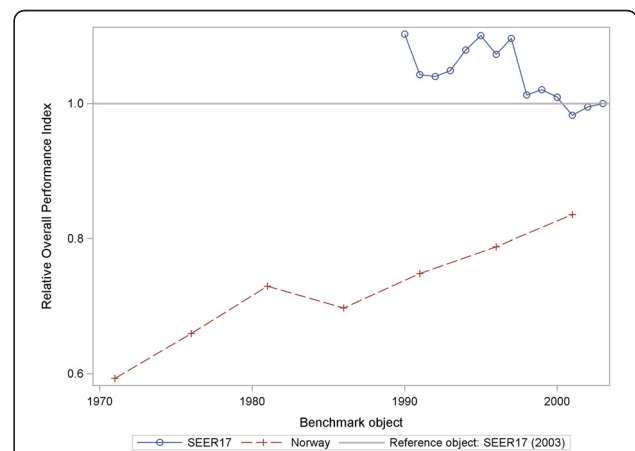
Database: Incidence - SEER 17-Nov 2010 Sub (1973–2008 varying).

outcomes such as morbidity or 30-day in-house mortality [4,30], because the chosen outcome parameters of relative survival are bound to annually presented life tables. These may also be calculated for shorter time intervals. In this case the estimation of country-specific or regional life tables is recommended [31] for

differentiation purposes. Finally, it should be stressed that the developed approach extends study methods that are aimed at comparing country-based relative 5-year survival rates [32-35], because information of absolute and relative stage distribution is linked to the corresponding survival rates.



**Figure 1** Extended Standardized Screening Index (eSSI) for new malignant breast cancer cases from SEER-17 (1990–2003), Norway (1969–2003) and SEER-17 (2003) as a reference object.



**Figure 2** Relative Overall Performance Index (ROPI) of new malignant breast cancer cases from SEER 17 (1990–2003), Norway (1969–2003) and SEER-17 (2003) as a reference object.

### Conceptualization and interpretation

The conceptualization and epidemiological interpretation or health policy conclusions should be drawn in the country and time-specific context. Therefore, the interpretation of the data examples remains restricted here to a.) great variations (cross-sectional) between benchmark objects and b.) convergence tendencies between the US and Norway over the course of time.

One contributing factor for this result is measured by the eSSI and SCI which are both conceptualized on the premise that the more effective early detection and screening programs, the more the distribution of cases will be skewed toward the earlier stages of disease. This characteristic reflects common knowledge regarding the coherence between early detection programs and shifts of stages [36-38] even if screening methods may be discussed as somewhat controversial in certain age groups [39]. Also the conceptualization of the SWI captures this stage migration effect directly. The SWI is based upon the premise that the more critical the work-up, the more upstaging will occur and the better the survival at each stage. However, unless the stage migration alters the treatments administered, it will not impact the overall survival, only the survival at each individual stage. The STI is based upon the premise that the better the overall survival corrected for the case-mix, the more effective the treatment being administered. The utilization of each of these indices as a benchmark provides a means of identifying specific areas of program strength and weakness. The combination of these indices to create ROPI provides a benchmark for assessment of overall program quality.

### Conceptual pitfalls

The suggested method allows the comparison of countries, states, and districts on a longitudinal scale. This perspective offers a high information grade in international comparisons. However, this means that an especially high data quality is necessary that must fulfill the minimum requirements of representative, accurate, complete, and comparable data [40,41]. Aside from these formal requirements, some important statistical details must be regarded which are related to stage distributions and survival rates.

The comparison of the stage distribution may be distorted by the stage migration or the so called "Will-Rogers-Phenomenon" [42-44]. According to this phenomenon slow growing, "quiet", and not apparently discernible disease symptoms such as metastases are discovered earlier due to increasingly powerful imaging procedures (diagnostic imaging). This means these cases are no longer classified as early disease stages (0-II), but as later ones (III-IV). Thus, the chance of survival increases in the early disease stages, because

fewer patients with metastases and unfavorable prognoses are included. However, the chance of survival also increases in the later stages, because patients with metastases that are not apparent are detected earlier. Due to this effect also known as stage migration distorted stage-specific chances of survival result. The overall chance of survival, however, is not affected by this phenomenon [42].

On the contrary, comparison of SWI and STI facilitates a greater understanding of the contribution of this stage migration effect to the overall outcome. For example, if the stage-specific survival information (SWI) is approximately 1.0 and the overall survival information (STI) is substantially greater than 1.0, the improved overall outcome adjusted for the stage mix (SCI) appears to be primarily a treatment effect. On the other hand, if the SWI is substantially greater than 1.0 but the STI is approximately 1.0, there is a stage migration occurring that does not appear to have a major impact on administered treatments.

Alongside, the comparison of survival conditions in time may also be distorted by the so-called lead-time bias or zero-time shift [42,45]. Accordingly, screening-tests and diagnostic procedures can identify a disease even before the patient develops symptoms. This effect leads to increasing survival times without actually leading to prolongation of life, if the health care effectiveness remains constant.

### Strengths

Compared to the original indices according to Beatty et al. [6,7] an extended screening-index (eSSI) is included, which is standardized in the same logic as all the other indices and which is expressed as a reciprocal. Therefore a positive association between the eSS-Index and the outcome-indicator is established. The latter has been redefined by substituting the breast cancer-specific fatality rates by relative survival rates. This approach has several advantages:

- Overall more patients can be included in the analyses because survival of all patients is of concern; regardless of the cause of death. This leads to a higher statistical power.
- Other causes of death respective frequently reported misleading information [46,47] do not have to be modeled in a competing risk model following Gooley [8].
- Background mortality of the population can be considered in the model.
- The calculation of the non-parametrically estimated relative survival rates does not require distribution assumptions. They can be estimated with existing software solutions appropriately.

- One step forward, relative survival rates may be standardized by variables such as age, gender, ethnicity etc. limited only by the parameters available in life tables.
- Benchmark-algorithms and the outcome-indicator relative survival may be extended to any other disease as long as a classification in subsequent disease phases is possible.

The outcome-indicator relative survival can also be substituted by probabilities. For example, logistic regressions can be calculated to quantify readmission probabilities after inpatient treatment, if the corresponding benchmark parameter exists. Correspondingly, disease phases can be substituted by information on disease severity (e.g. Charlson-score, Elixhauser-index [48,49]) as long as they have an ordinal order.

#### **Weakness**

The proposed eSS-Index serves as a level-parameter (“intercept”) which defines the general premise of the benchmark respective reference object. Its weakness is based on the arbitrary weights provided by the stage numbers which devalue the earliest disease identified (stage 0) and emphasize the “not known” (stage 5, see Table 1). This arbitrariness might decrease the clinical value of both, the eSSI and ROPI. However, eSSI informs about stage distributions without any survival information and obtains the logical consistency of ROPI.

The conceptual pitfalls of the Will-Roger phenomenon, stage migration and lead-time bias have already been mentioned (see above). The former is explored by comparing SWI- and STI, but lead-time bias would distort that assessment and cannot be distinguished from an apparent increase in the incidence of the disease. In addition, it is highly recommended from a methodological point of view to estimate outcome-parameters using the same method but with different regional life tables. Therefore, population-based data is a crucial prerequisite, i.e. the catchment area of integrated networks or new organization forms in general should be clearly defined by their landmarks in order to obtain crucial demographic information from local or regional statistical authorities. In addition, if possible and if these cannot be avoided, benchmark objects should have the same structural breaks in their nomenclature of disease phase (e.g. AJCC, UICC). For example, the UICC Tumor-Node-Metastasis version 5 (1997–2001) was valid until version 6 (2002–2008) and version 7 (since 2009) became effective. The use of the same nomenclature should be assured for reference and benchmark objects. But from a practice point of view, this is difficult to achieve due to time lags in implementation and documentation. Furthermore a fair benchmark has to be assured which

means that a comparison between “equivalent” comparison objects should be achieved. Thus, homogenous benchmark objects in terms of “peer-groups” should be identified [50]. This is recommended since most health care systems have evolved historically and, thus, infrastructure characteristics and innovations can be implemented 1:1 from one country, region or district to another under certain restrictions only [51,52]. However, the identification of “peer groups” can either be based on content-related considerations (e.g. countries with national health care services), on statistically chosen disease-specific parameters (e.g. distribution of risk, prognosis, and predictive factors) or both [9,53]. Overall incidence-based factors must be differentiated from patient-, disease-, and health care system-centered factors [45] which are responsible for statistical distortion associated with survival analyses.

#### **Perspectives**

Benchmark-algorithms that compare countries, states, and districts are highly complex and require great attention to research details [40,54]. To be able to take these into consideration, high quality data is necessary [9,53,55]. But high data quality in terms of accuracy and completeness are hard to achieve. In case of the SEER register for example, some concerns have been documented [56] which should be thoroughly considered when results are interpreted for health care decision making. This approach is even more recommended instead of the growing number of health care providers who seek to leave the data gathering process due to cost reductions and missing benefits [57]. However, it is these data that form the basis to achieve a higher transparency of efficiency and health care quality which became a crucial competition parameter in a growing health care industry. Therefore it is crucial for the next step in quality assurance to demonstrate how these data may achieve clues of evidence for further improvements. The algorithms proposed here may serve as first identifier of infrastructural differences in screening programs and compare these with alternated consequences for the clinical work-up in countries, states and districts. However from a methodical point of view, the development of benchmark algorithms is not complete. Corresponding tests are missing to generate p-values for which distribution assumptions are requested. If and under what circumstances certain distributions are given, will be the task of future developments. Finally, the clinical meaning and interpretation of index differences between reference- and benchmark-objects has to be explored in future applications.

#### **Conclusions**

To measure international, national, and regional health care quality, the suggested algorithms and freely accessible SAS-macros BenchRelSurv and BenchRelSurvPlot

offer an additional tool to evaluate screening programs, the clinical work-up and effectiveness in general. An effectiveness comparison is sought that links the earliest possible time point of a progressive disease with the time point of an absorbing result after the onset of the primary disease considering the background mortality (relative survival). This is especially relevant for diseases (e.g. breast cancer) where the etiology and disease causes remain unclear. However, this concept can also be transferred to preventable diseases or avoidable mortalities which have clearly defined disease courses (e.g. cardiovascular disease, diabetes mellitus II) and which are clearly avoidable by (behavioral) interventions. The software allows the identification of performance measurement in relation to comparative regions. It offers a first step towards an in depth research analysis.

### Availability and requirements

Project name: Benchmarking relative survival (BenchRelSurv, BenchRelSurvPlot)

Project home page: <http://sourceforge.net/projects/benchrelsurv/> delivers files and examples as well as Technical Reports in German and English

Operating system(s): Platform dependency of SAS 9.2 or higher

Programming language: SAS 9.2 and higher

Other requirements: None

License: None

Any restrictions to use by non-academics: SAS 9.2 license or higher

### Additional files

**Additional file 1: Example1.sas: Cross-sectional benchmark for new malignant breast cancer cases from SEER-17 (1999–2003) registers with BenchRelSurv and BenchRelSurvPlot.**

**Additional file 2: Example2.sas: Longitudinal benchmark for new malignant breast cancer cases from SEER-17 (1990–2003), Norway (1969–2003) with BenchRelSurv and BenchRelSurvPlot.**

### Competing interests

Authors declare no competing interest.

### Authors' contributions

COJ contributed substantially to the conception, acquisition of data, design and programming of the source code, analysis and interpretation of data, drafting the manuscript, and gave final approval. IR was involved in the design, programming, and validation of the source code and revised and approved the final draft. USA was involved in the conception, analysis, and interpretation of data, revised critically for important intellectual content, and gave approval for the final draft. All authors read and approved the final manuscript.

### Acknowledgement

This work was supported by a grant of the German Ministry of Health (Fö.Kz.: FB 2-43332-70/6) and the Ministry of Education and Research within the program: Health research to benefit the people - Guideline implementation for early diagnosis and treatment of breast cancer (Fö.-KZ: GFZPO1119302).

The authors wish to thank the reviewers for their constructive comments and suggestions.

### Author details

<sup>1</sup>Central Institute of Mental Health, Medical Faculty Mannheim/University Heidelberg, Square J5, 68159 Mannheim, Germany. <sup>2</sup>Department of Gynecology, Gynecological Endocrinology and Oncology, Breast Center Regio, University of Marburg, Marburg, Germany.

Received: 27 July 2012 Accepted: 19 December 2012

Published: 14 January 2013

### References

1. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, Carney PA, Kerlikowske K, Buist DSM, Weaver DL, Barlow WE, Ballard-Barbash R: **Performance benchmarks for screening mammography.** *Radiology* 2006, **241**:55–66.
2. Sickles EA, Miglioretti DL, Ballard-Barbash R, Geller BM, Leung JWT, Rosenberg RD, Smith-Bindman R, Yankaskas BC: **Performance benchmarks for diagnostic mammography.** *Radiology* 2005, **235**:775–790.
3. Pierce LJ, Moughan J, White J, Winchester DP, Owen J, Wilson JF: **1998–1999 patterns of care study process survey of national practice patterns using breast-conserving surgery and radiotherapy in the management of stage I-II breast cancer.** *Int J Radiat Oncol Biol Phys* 2005, **62**:183–192.
4. O'Brien MER, Borthwick A, Rigg A, Leary A, Assersohn L, Last K, Tan S, Milan S, Tait D, Smith IE: **Mortality within 30 days of chemotherapy: a clinical governance benchmarking issue for oncology patients.** *Br J Cancer* 2006, **95**:1632–1636.
5. Kerba M, Miao Q, Zhang-Salomons J, Mackillop W: **Defining the need for breast cancer radiotherapy in the general population: a criterion-based benchmarking approach.** *Clin Oncol (R Coll Radiol)* 2007, **19**:481–489.
6. Beatty J, Rees J, Atwood M, Pugliese M, Bolejack V: **Standardized evaluation of regional and institutional breast cancer outcomes.** [http://www.facs.org/cancer/coc/pdf/beatty\\_breast-outcomes.pdf?bcsi\\_scan\\_58A52892D4FFF3EB=0&bcsi\\_scan\\_filename=beatty\\_breast-outcomes.pdf](http://www.facs.org/cancer/coc/pdf/beatty_breast-outcomes.pdf?bcsi_scan_58A52892D4FFF3EB=0&bcsi_scan_filename=beatty_breast-outcomes.pdf).
7. Beatty J, Rees J, Atwood M, Pugliese M, Bolejack V: **Standardized evaluation of regional and institutional breast cancer outcomes.** *Am J Surg* 2008, **195**:636–640.
8. Gooley TA, Leisenring W, Crowley J, Storer BA: **Estimation of failure probabilities in the presence of competing risks. New representations of old estimators.** *Statist Med* 1999, **18**:695–706.
9. Jacke C, Kalder M, Koller M, Wagner U, Albert U: **Systematic assessment and improvement of medical data quality.** *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 2012, **55**:1495–1503.
10. Stausberg J, Nonnemacher M, Weiland D, Antony G, Neuhäuser M: **Management of Data Quality. Development of a Computer-Mediated Guideline.** *Stud Health Technol Inform*, 2006:477–482.
11. Nonnemacher M, Weiland D, Stausberg J: **Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern.** Berlin: Med.-Wiss. Verl.-Ges; 2007.
12. Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) SEER\*Stat Database: *Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2011 Sub (1973–2009 varying). Linked To County Attributes - Total U.S., 1969–2010 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012, based on the November 2011; submission.*
13. Cancer Registry of Norway: **Cancer in Norway 2008: Cancer Incidence, Mortality, Survival and Prevalence in Norway.** Oslo: Cancer Registry of Norway; 2009.
14. Therneau T, Grambsch P: *Modeling Survival Data: Extending the Cox Model.* New York: Springer; 2001.
15. Dickman P, Sloggett A, Hills M, Hakulinen T: **Regression models for relative survival.** *Stat Med* 2004, **23**:51–64.
16. Pohar M, Stare J: **Making relative survival analysis relatively easy.** *Comput Biol Med* 2007, **37**:1741–1749.
17. Hakulinen T: **Cancer survival corrected for heterogeneity in patient withdrawal.** *Biometrics* 1982, **38**:933–942.
18. Hakulinen T, Tenkanen L: **Regression analysis of relative survival rates.** *J R Stat Soc Ser C Appl Stat* 1987, **36**:309–317.



19. Brenner H: Long-term survival rates of cancer patients achieved by the end of the 20th century. A period analysis. *Lancet* 2002, **360**:1131–1135.
20. Brenner H, Gefeller O: An alternative approach to monitoring cancer patient survival. *Cancer* 1996, **78**:2004–2010.
21. Brenner H, Hakulinen T, Gefeller O: Computational realization of period analysis for monitoring cancer patient survival. *Epidemiology* 2002, **13**:611–612.
22. Geiss K, Meyer M, Padespiel-Tröger M, Gefeller O: SURVSOFT-Software for nonparametric survival analysis. *Comput Methods Programs Biomed* 2009, **96**:63–71.
23. Sarioego J: Regional variation in breast cancer treatment throughout the United States. *Am J Surg* 2008, **196**:572–574.
24. Sarioego J: Patterns of breast cancer presentation in the United States: does geography matter? *Am Surg* 2009, **75**:545–9.
25. Brucker SY, Schumacher C, Sohn C, Rezaei M, Bamberg M, Wallwiener D: Benchmarking the quality of breast cancer care in a nationwide voluntary system: the first five-year results (2003–2007) from Germany as a proof of concept. *BMC Cancer* 2008, **8**:358.
26. Brucker SY, Wallwiener M, Kreienberg R, Jonat W, Beckmann MW, Bamberg M, Wallwiener D, Souchon R: Optimizing the quality of breast cancer care at certified German breast centers: a benchmarking analysis for 2003–2009 with a particular focus on the interdisciplinary specialty of radiation oncology. *Strahlenther Onkol* 2011, **187**:89–99.
27. Wallwiener M, Brucker SY, Wallwiener D: Multidisciplinary breast centres in Germany: a review and update of quality assurance through benchmarking and certification. *Arch Gynecol Obstet* 2012, **285**:1671–83.
28. Daroui P, Gabel M, Khan AJ, Haffty BG, Goyal S: Utilization of breast conserving therapy in stages 0, I, and II breast cancer patients in New Jersey: an American College of Surgeons National Cancer Data Base (NCDB) analysis. *Am J Clin Oncol* 2012, **35**:130–135.
29. Ng W, Delaney GP, Jacob S, Barton MB: Estimation of an optimal chemotherapy utilisation rate for breast cancer: setting an evidence-based benchmark for the best-quality cancer care. *Eur J Cancer* 2010, **46**:703–712.
30. El-Tamer MB, Ward BM, Schiffner T, Neumayer L, Khuri S, Henderson W: Morbidity and mortality following breast cancer surgery in women. National benchmarks for standards of care. *Ann Surg* 2007, **245**:665–671.
31. Dinkel R: Die Berechnung des Parameters „Relative Survival“ für ein Tumoregister mit regionalem Einzugsbereich. <http://www.krebsregister.uni-rostock.de/docs/Relative%20Survival.pdf>.
32. Coleman M, Babb P, Damięcki P, Grosclaude P, Honjo S, Jones J, Kerner G, Pitard A, Quinn M, Sloggett A, de Stavola B: *Cancer survival trends in England and Wales 1971–1995: deprivation and NHS Region*. London: The Stationery Office; 1999 [Series SMPS, vol. 61].
33. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan CE, Turner D, Marrett L, Gjerstorff ML, Johannesen TB, Adoffson J, Lambe M, Lawrence G, Meechan D, Morris EJ, Middleton R, Steward J, Richards MA: Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011, **377**:127–138.
34. Coleman M, Quaresma M, Berrino F, Lutz J, de Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, Micheli A, Sant M, Weir H, Elwood J, Tsukuma H, Koifman S, E-Silva G, Francisci S, Santaquilani M, Verdecchia A, Storm H, Young J: Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008, **9**:730–756.
35. de Blacam C, Gray J, Boyle T, Kennedy MJ, Hollywood D, Butt J, Griffin M, Nicholson S, Dunne B, Wilson G, McDermott R, Murphy P, Short I, Rowley S, Connolly E, Reynolds JV: Breast cancer outcomes following a national initiative in Ireland to restructure delivery of services for symptomatic disease. *Breast* 2008, **17**:412–417.
36. Katalinic A, Bartel C, Raspe H, Schreer I: Beyond mammography screening. Quality assurance in breast cancer diagnosis (The QuaMaDi project). *BJC* 2007, **96**:157–161.
37. Ferroni E, Camilloni L, Jimenez B, Furnari G, Borgia P, Guasticchi G, Rossi PG: How to increase uptake in oncologic screening: a systematic review of studies comparing population-based screening programs and spontaneous access. *Prev Med* 2012, in press.
38. Puliti D, Zappa M: Breast cancer screening: are we seeing the benefit? *BMC Med* 2012, **10**:106.
39. Amir E, Bedard PL, Ocaña A, Seruga B: Benefits and harms of detecting clinically occult breast cancer. *J Natl Cancer Inst* 2012, in press.
40. Mainz J, Hjulsgaard M, Og MTE, Burgaard J: National benchmarking between the Nordic countries on the quality of care. *J Surg Oncol* 2009, **99**:505–507.
41. Mainz J, Bartels P, Rutberg H, Kelley E: International benchmarking. Option or illusion? *Int J Qual Health Care* 2009, **21**:151–152.
42. Feinstein AR, Sosin DM, Wells CK: The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* 1985, **312**:1604–1608.
43. Golder WA: Das Will-Rogers-Phänomen und seine Bedeutung für die bildgebende Diagnostik. *Radiologe* 2009, **49**:348–354.
44. Spratt JS: Will Rogers phenomenon. *Arch Surg* 1992, **127**:868.
45. Autier P, Boniol M: Caution needed for country-specific cancer survival. *Lancet* 2011, **377**:99–101.
46. Modelmog D, Goertchen R, Steinhard K, Sinn HP, Stahr H: Vergleich der Mortalitätsstatistik einer Stadt bei unterschiedlicher Obduktionsquote (Görlitzer Studie). *Pathologe* 1991, **12**:191–195.
47. Schelhase T, Weber S: Die Todesursachenstatistik in Deutschland. Probleme und Perspektiven. *Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz* 2007, **50**:969–976.
48. Lieffers JR, Baracos VE, Winget M, Fassbender K: A comparison of Charlson and Elixhauser comorbidity measures to predict colorectal cancer survival using administrative health data. *Cancer* 2011, **117**:1957–1965.
49. Li B, Evans D, Faris P, Dean S, Quan H: Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC Health Serv Res* 2008, **8**:12.
50. Walshe K: International comparisons of the quality of health care. What do they tell us? *Qual Saf Health Care* 2003, **12**:4–5.
51. Walshe K: Understanding what works-and why-in quality improvement: the need for theory-driven evaluation. *Int J Qual Health Care* 2007, **19**:57–59.
52. Nolte E, McKee M: Measuring the health of nations. Analysis of mortality amenable to health care. *BMJ* 2003, **327**:1129.
53. Jacke C, Kalder M, Wagner U, Albert U: High-quality data for valid comparisons and decisions. Assessing the accuracy and completeness of medical data. *BMC Research Notes*, in press.
54. Mainz J: Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003, **15**:523–530.
55. Nonnemacher M, Weiland D, Neuhäuser M, Stausberg J: Adaptive management of data quality in cohort studies and registers. Proposal for a guideline. *Acta Informatica Medica* 2007, **15**:225–230.
56. Beatty JD, Adachi M, Bonham C, Atwood M, Potts MS, Haferson JL, Aye RW: Utilization of cancer registry data for monitoring quality of care. *Am J Surg* 2011, **201**:645–649.
57. Greene FL, Gilkerson S, Tedder P, Smith K: The role of the hospital registry in achieving outcome benchmarks in cancer care. *J Surg Oncol* 2009, **99**:497–499.

doi:10.1186/1471-2458-13-34

Cite this article as: Jacke et al.: Using relative survival measures for cross-sectional and longitudinal benchmarks of countries, states, and districts: the BenchRelSurv- and BenchRelSurvPlot-macros. *BMC Public Health* 2013 **13**:34.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

