nicer
Foundation
National Institute for
Cancer Epidemiology
and Registration

Methods for the estimation of cancer survival

The proportion of people alive at a specified time following diagnosis with cancer is the **absolute or observed survival**:

$$S(t) = e^{-\int h(t)dt}$$

where *S* is the decaying proportion of patients alive after *t*, the time since diagnosis, and *h(t)* is the hazard function, or the death rate of patients having survived until *t*. The hazard is usually high shortly after diagnosis and smaller at later intervals, giving rise to more steeply decaying survival curves early after diagnosis.

S(t) and *h(t)* are estimated from Swiss data using the STATA™ command strs v1.4.2 in consecutive intervals after diagnosis of increasing length, assuming constant hazard within intervals [Dickman and Coviello 2015. Estimation and modeling relative survival. *The Stata Journal* **15**, 186-215]. Cut points were set to 0.1, 0.2, 0.6, 1, 2, 3, …, 10 years after diagnosis.

Absolute or observed survival proportions are not well suited for comparisons because they depend heavily on patient characteristics such as age at diagnosis or comorbidities, and on many other risk factors. It is more informing to assess the ability of patients to survive their cancer by considering only deaths due to the cancer in question, which is called **cause-specific survival**. Because it can be difficult to decide the underlying main cause of death, another approach is usually adopted: the **relative survival**. Relative survival is calculated as the ratio of observed survival of cancer patients and expected survival of a group of persons from the general population, who is identical with respect to the risk factors sex, age, calendar year and place of residence:

$$R(t) = \frac{S(t)}{S^*(t)}$$

with *R(t)* as relative survival, *S(t)* as the observed survival, *S*(t)* as the expected survival, and *t* as the time since diagnosis. The expected survival is derived from population lifetables according to the Ederer_II method [Ederer and Heise 1959. Instructions to IBM 650 programmers in processing survival computations, methodological note 10. End Results Evaluation Section, National Cancer Institute]. 50% relative survival signifies that cancer patients survived ½ as well as counterparts of the general population. The surplus mortality can be assigned specifically to the cancer in question. 100% relative survival signifies that deaths in the patient group were as frequent as deaths in the general population, in other words, the patients as a group may be regarded as cured from cancer.

The assumption that risk factors not specifically controlled for will be distributed similarly among patients and persons of the general population is arguably a strong one. But sensitivity analyses showed that even for lung cancer patients, who often smoke and carry a higher risk for other diseases than the general population, this did not have a concerning impact [Hincliffe, Rutherford, Crowther, Nelson and Lambert 2012. Should relative survival be used with lung cancer data? *BJC* **106**, 1854-1859].

In order to compare the Swiss survival estimates between sub-groups of different age-structure within Switzerland, or with other countries, and over time, they are **age-standardized** using the method proposed by Brenner et al. 2004 and cancer-specific weights suggested as International

Cancer Survival Standards (ICSS) [Corraziari, Quinn and Capocaccia 2004. Standard cancer patient population for age standardizing survival ratios. *EJC* **40**, 2307-2316]. Age-groups 15-44, 45-54, 55-64, 65-74, 75+ for each cancer type were created, and slightly adapted for prostate cancer patients, who are usually diagnosed at older ages (15-45, 55-64, 65-74, 75-84, 85+).

Assessment of vital status follow-up in Switzerland

For cantonal cancer registries to be included in the national survival statistics, the vital status of registered patients has to be systematically ascertained and at least five years of follow-up have to be available. These criteria included 9 of the 14 registries in existence as of 2016. Each Swiss cancer registry employs two forms of follow-up:

With passive vital status follow-up, dates of death received at regular intervals from the national or cantonal vital statistics division are linked to patient records. The quality of passive follow-up may be compromised by incomplete national or cantonal vital statistics, by false negative or false positive linkage errors, by linking only those deaths caused by cancer to patient records, by emigration of patients out of the registries catchment area or by registration of patients after the information on death was received.

With active vital status follow-up every patient registered with vital status alive is followed-up at regular intervals. This may be accomplished by repeated scrutiny of medical records in hospitals, enquiries with attending physicians, scanning the population registers, health registers of national health services, health insurance registers, electoral lists or postal/telephone enquiries. Patients for which the vital status could not be determined at the time of follow-up (e.g. because he/she has moved out of the registry population) are assigned to the category “lost to follow-up” and the last known date alive is retained.

Complete vital status follow-up is a prerequisite for valid survival statistics. Fig. 1 provides an overview of the situation of vital status follow-up information for the reference date 31.12.2013 in Switzerland. Complete information about follow-up is available for patients who have died (51%) or are alive with active follow-up date at least July 2013 (32% recent active FU). The follow-up information is incomplete in 2% of the patients because they are lost to follow-up, in 12% of patients because the vital status alive refers to dates in the past (i.e. before July 2013), in 2% of patients because they lack active follow-up, and in 1% of patients because they were registered via death certificate only (DCO) and the data of diagnosis is unknown. Patients registered as DCO or diagnosed at death were excluded from survival analysis.

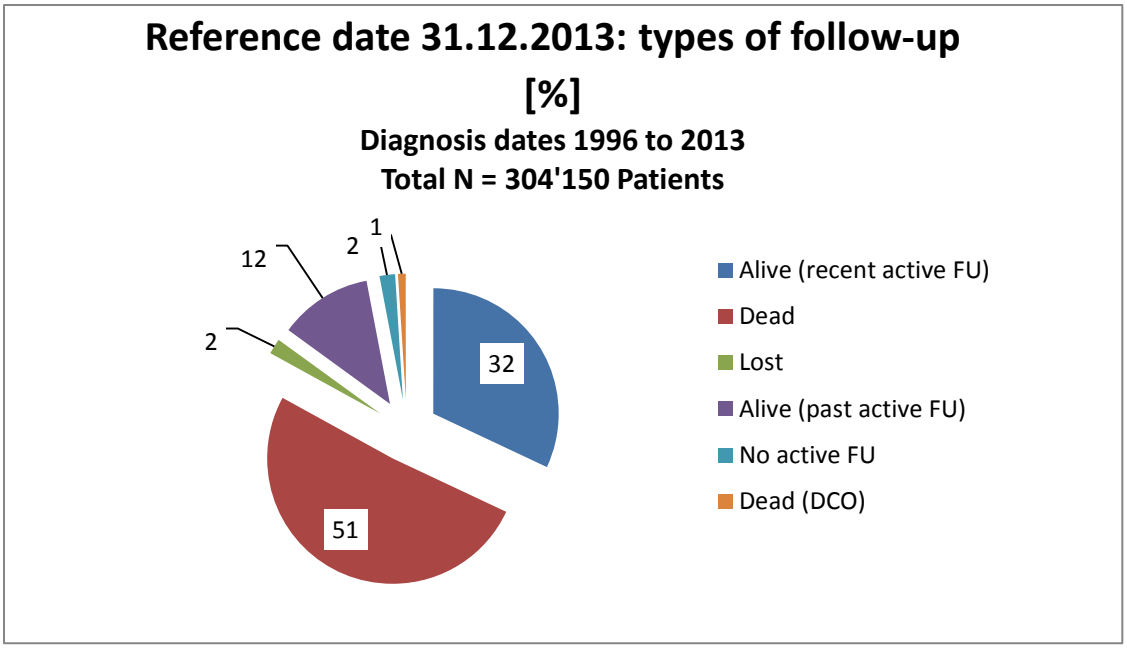

Fig. 1. Distribution of types of vital status and follow-up information

Assuming complete high quality passive follow-up, patients alive without active follow-up, or with active follow-up for past dates, could be treated as alive until the reference date 31.12.2013. Surveying registration practices in Switzerland revealed that not all cancer registries linked the national or cantonal official vital statistics to registered cases for every year of incidence, or that linkage was performed only for persons with cancer as the cause of death.

We carried out a sensitivity analysis whether the assumption of continued survival in the absence of reported death could lead to implausible high estimates of survival:
Link to posters "Swiss cancer survival statistics: quality of vital status follow-up (2014)" and "Funnel plots to explore the quality of vital status follow-up in Switzerland (2014)"
The assumption of continued survival in the absence of reported death caused unexpected high survival in cancer registries with less than optimal passive follow-up procedures. Thus, the conservative approach of censoring all survival times to the last known date was adopted for national survival statistics in Switzerland.

Data selection to assess cancer survival trends

Fig. 2 shows how cancer cases were selected to assess temporal changes in cancer survival.

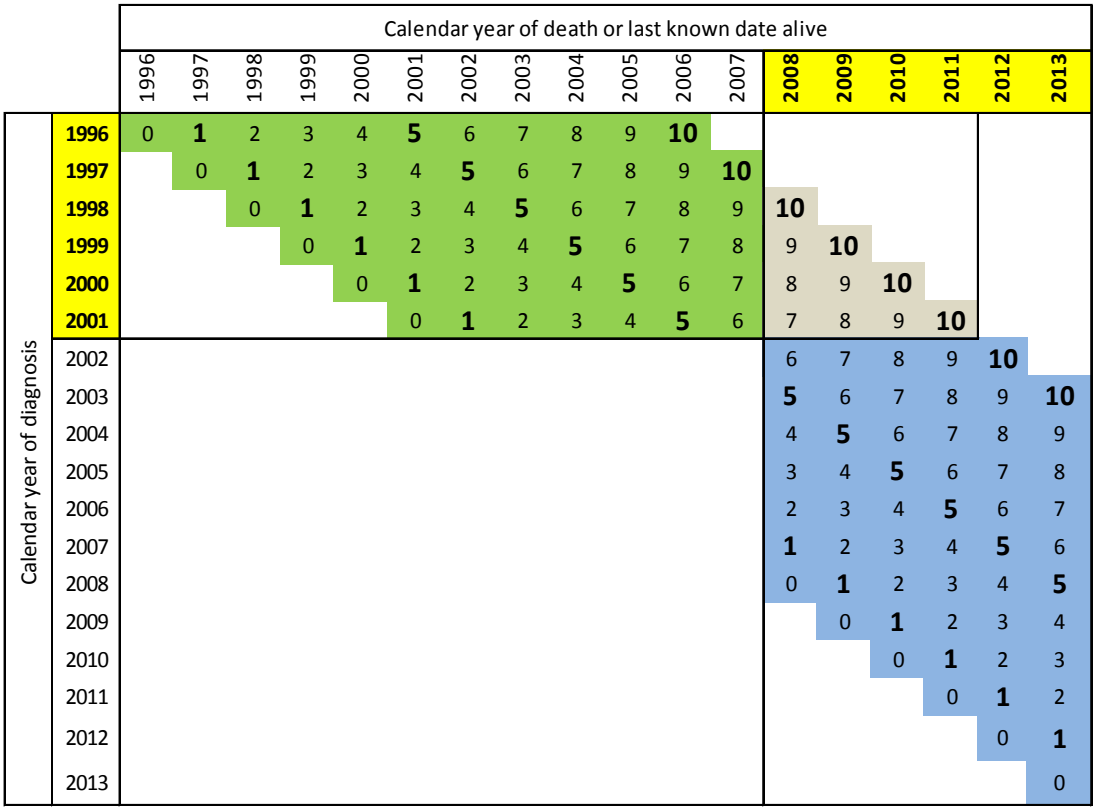

Fig. 2. Case selection for comparing the survival of patients diagnosed 1996 to 2001 with the survival of patients diagnosed more recently.

The numbers in the cells of Fig. 2 indicate the *minimum* number of complete years of follow-up available for patients diagnosed between 1996 and 2013 (vertical axis) and who survived to the end of a given year up to 2013 (horizontal axis).

The **cohort approach** was chosen for survival of patients diagnosed 1996 to 2001 (vertical axis marked yellow in Fig. 2). Each patient diagnosed 1996-2001 has at least 10 years of follow-up by the end of 2011 (green and grey areas in Fig. 2). This approach, based on the follow-up of a clearly-defined group of patients, is attractive because of its conceptual simplicity. However, the necessity to collect 10-years follow-up data on every patient means that cohort survival estimates summarise more historic information on patients than what the registry currently holds.

The **period approach** is able to include the most recently diagnosed patients [Brenner and Gefeller 1996. An alternative approach to monitoring cancer patient survival. *Cancer* **78**, 2004-2010]. It estimates survival from patients selected by their time of death or last known vital status within the calendar period 2008-2013 (horizontal axis marked yellow in Fig. 2), for patients who were diagnosed during the period 1998 to 2013 (blue and grey areas in Fig. 2). Short-term survival is estimated from recently diagnosed patients (e.g. 1-year survival is estimated from patients diagnosed 2007-2013). Long-term survival is estimated from patients diagnosed earlier (e.g. 10-year survival is estimated from patients diagnosed 1998-2003). Thus, period survival includes more up-to-date data than cohort analysis. However, it still may not capture recent changes to survival that occur late after diagnosis as exemplified by the grey area in Fig. 1, were period and cohort analysis are based on the

same data. Period survival analysis is conceptually more difficult to explain than cohort analysis. An alternative way to show the different case selection for cohort and period analysis is provided in Fig. 3. The left panel in Fig. 3 shows that in cohort analysis, short- and long-term survival probabilities are estimated from the same patients. The right panel in Fig. 3 shows that in period analysis, short term survival probabilities are estimated only from recently diagnosed patients and that patients diagnosed earlier contribute only for the estimation of long survival probabilities.

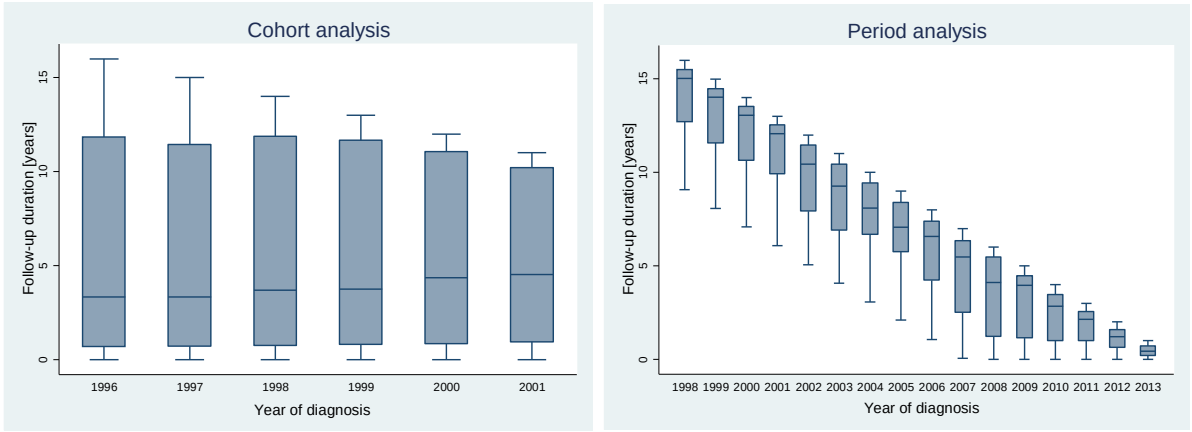

Fig. 3. Follow-up durations included for survival analysis in the cohort (left panel) and the period approach (right panel). Distribution of follow-up durations per year of diagnosis are shown as box plots, with the width of the box signifying the 25th to 75th percentile, the horizontal line the median, and the outer caps the minimum and maximum.