Cancer Registry of Norway
INSTITUTE OF POPULATION-BASED CANCER RESEARCH

Cancer in Norway 2017

Technical Supplement: Statistical Methods

Department of Registration
Section of Research

Cancer in Norway 2017

Technical Supplement: Statistical Methods

Cancer Registry of Norway

Department of Registration, Section of Research

October 26, 2018

1 Statistical methods

1.1 Target readership

The target readership for this technical supplement includes statisticians and cancer epidemiologists.

1.2 Incidence and mortality rates

Rates are used to measure the frequency with which an event occurs in a defined population in a defined time period. Rates facilitate comparisons across groups of different sizes. Let d and Y denote the number of events and the number of person-years in the population, respectively. In Cancer in Norway 2017 (CIN2017) Y, the mid-year population, is calculated as the simple arithmetic mean of the population at the start and end of each calendar year. If the interest lies in calculating a rate for a period of more than 1 year, one first calculates the annual mid-year population, and then aggregates these mid-year numbers to reach the total number of person-years.

Rates are reported both as age-specific rates and age-standardised rates per 100,000 person-years. The population is divided into 18 distinct five-year age groups. Let d_i and Y_i denote the number of events and the total number of person-years, respectively, for age group i. The age-specific rate r_i per 100,000 person-years, for age group i is then given by

$$r_i = \frac{d_i}{Y_i} \cdot 10^5. \tag{1}$$

The age-standardised rate (ASR) is calculated as

$$ASR = \frac{\sum_{i=1}^{18} w_i r_i}{\sum_{i=1}^{18} w_i}, \tag{2}$$

where w_i is a weight, for age group i, given by some reference population.

Typically the World Standard Population has been used (Doll & al, 1966). CIN2017 is using the age distribution of the Norwegian 2014 mid-year population as standard population. The population weights of the World Standard Population and the Norwegian 2014 population are given in the table below. One should notice that the World Standard Population upweights the younger age groups and downweights the older age groups compared to the 2014 Norwegian population.

i	Age	World (Doll/Segi)	Norway (2014)
1	0–4	12,000	6,039
2	5–9	10,000	6,102
3	10–14	9,000	5,993
4	15–19	9,000	6,349
5	20–24	8,000	6,681
6	25–29	8,000	6,770
7	30–34	6,000	6,690
8	35–39	6,000	6,670
9	40–44	6,000	7,296
10	45–49	6,000	7,207
11	50–54	5,000	6,492
12	55–59	4,000	6,108
13	60–64	4,000	5,575
14	65–69	3,000	5,369
15	70–74	2,000	3,702
16	75–79	1,000	2,663
17	80–84	500	2,063
18	85+	500	2,231
Sum		100,000	100,000

1.3 Cumulative risk

Cumulative risk (CR) is an estimate of the risk of developing a certain type of cancer by a given age. It is defined as

$$CR = 1 - e^{-\left(5 \sum_{i=1}^{N} r_i\right)}, \tag{3}$$

where i denotes age groups and N denotes the age group for which CR is estimated. In CiN CR is estimated up to the age of 74, so $N = 15$.

1.4 Prevalence

Prevalence is calculated as the number of people in the population that are alive and have been diagnosed with the cancer of interest at some point during their lifetime.

1.5 Relative survival (Net survival)

Net survival is estimated by the relative survival ratio, $R(t)$, defined by

$$R(t) = \frac{S_O(t)}{S_E(t)}, \tag{4}$$

where $S_O(t)$ is the observed survival at time t and $S_E(t)$ is the expected survival at time t. Observed survival is calculated using the actuarial method (also frequently named the life table method). The period of interest is divided into k time intervals, where interval $i \in [t_{i-1}, t_i)$. Let l_i, d_i and c_i denote the number of persons alive at the start of interval i, the number of deaths in interval i and the number of censored individuals during interval i, respectively. Assuming that censoring occurs uniformly throughout each time interval, the observed cumulative survival is calculated as

$$S_O(t) = \prod_{i=1}^{k} p_i, \tag{5}$$

where p_i denotes the interval-specific observed survival, given by $p_i = \left(1 - \frac{d_i}{l_i'}\right)$, and $l_i' = l_i - \frac{1}{2}c_i$ is the effective number at risk in interval i. When the period approach is used, the estimate of the interval-specific observed survival is calculated by transforming the estimated cumulative hazard, p_i =exp$\{b_i \cdot (-d_i/y_i)\}$. Here b_i is the width of the interval and y_i is the person-time at risk in the interval.

Cumulative expected survival is calculated using the Ederer II estimator (Ederer, 1959)

$$S_E(t) = \prod_{i=1}^{k} p_i^E, \tag{6}$$

where

$$p_i^E = \sum_{h=1}^{l_i} \frac{p_i(h)}{l_i}$$

denotes the interval-specific expected survival, obtained by averaging the annual expected survival probabilities $p_i(h)$ of the patients alive at the start of interval i.

The individual expected survival is obtained from national unsmoothed population life tables matched on sex, age, and calendar year.

To reduce the potential for bias and to facilitate comparisons over time, the net survival estimate must be age-standardised. Let $R_j(t)$ denote the relative survival of patients in age group j, and assume we have s distinct age groups, then an age-standardised estimate of net survival, denoted $R_s(t)$ is given by

$$R_s(t) = \sum_{j=1}^{s} w_j R_j(t), \tag{7}$$

where w_j is the weight for age group j. In time the these weights are determined, for each group of diagnosis and sex, by the age distribution of the patients diagnosed during the most recent 5-year period, 2013–2017.

Ideally it is better to have more age groups than fewer, and to use more narrow age groups for older patients. However, in practice, when splitting data in many age groups we will run out of patients during the follow-up period in one or more age groups, and an age-standardised estimate will be unobtainable. We calculate age-standardised estimates of net survival by dividing patients into three distinct age groups, defined by the tertiles of the age distribution in the most recent 5-year period of diagnosis. This approach reduces the probability of age-standardised estimates being unobtainable. For less frequent and more lethal cancer sites we still cannot avoid sparse data, particularly when estimating 15-year net survival.

As mentioned in Statistical Methods in CIN2017, the cohort method was used when follow-up was complete. The period approach was used to obtain estimates for the most recent year when analysing trends, as well as, for the most recent 5-year period in other analyses. When analysing trends; the relative survival estimates for the years 2013–2016 is obtained using a mixture of the cohort and period approach. This is done to avoid artifical changes in the trend curves when switching from the cohort

approach to the period approach. An estimate for the year 2013 based on a 5-year period window from 2009–2013 would for many cancer sites cause a drop in the trend curve since the 5-year estimate is largely affected by survival experience from patients diagnosed several years ago. To avoid this, the time at risk is conditioned on the year of diagnoses. As an example, for the cohort of patients diagnosed in the period 2009–2013, complete 5-year follow-up is available for patients diagnosed 2009–2012, whereas only 4 years of follow-up is available for patients diagnosed in 2013. To make up for this lack of follow-up, patients diagnosed before 2009 is considered at risk from January 1, 2013. This means that only the survival experience between year 4 and 5 for patients diagnosed 2008 is used when estimating the 5-year relative survival for the 2009–2013 cohort, exactly making up for the last year of follow-up lacking for the 2013 patients.

All survival analyses were performed using the Stata (Stata/MP 15.1 Revision 06 Jun 2018) program `strs` (version 1.4.2.2).

2 References

Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age adjustment of cancer survival rates. European Journal of Cancer 2004; 40: 2317–2322.

Dickman P W, Coviello E. Estimating and modeling relative survival. The Stata Journal 15, Number 1, pp. 186–215. 2015 (`strs`, version 1.4.2. May 7th 2015).

Doll R, Payne P, Waterhouse J. (Eds) Cancer Incidence in Five Continents: A Technical Report. Springer-Verlag (for UICC), Berlin, 1966.

Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival computations, methodological note 10. End Results Evaluation Section, National Cancer Institute, 1959.

Lambert P C, Dickman P W, Rutherford M J. Comparison of Different Approaches to Estimating Age Standardized Net Survival. BMC Medical Research Methodology 2015. http://www.biomedcentral.com/1471-2288/15/64

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC

Statistical Methods in Cancer Research. Volume IV - Descriptive Epidemiology IARC Scientific Publications No. 128 Edited by Jacques Esteve, Ellen Benhamou, Luc Raymond. Lyon, 1994. http://www.iarc.fr/en/publications/pdfs-online/stat/sp128/SP128-V4.pdf.

Statistics Norway, Statistics Norway provides statistics on population size each year, www.ssb.no/befolkning/statistikker/folkemengde.

Return Adress:
Kreftregisteret
P.O. box 5313 Majorstuen
N-0304 Oslo
Norway

Cancer Registry of Norway
Institute of Population-based Cancer Research

https://www.kreftregisteret.no/cin-2017
https://www.kreftregisteret.no/globalassets/cancer-in-norway/2017/cin-2017supmeth.pdf