Department of Registration Section of research
Cancer in Norway 2015

Technical Supplement: Statistical Methods

Department of Registration, Section of research

October 21, 2016
1 Statistical methods

1.1 Target readership

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

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 2015 (CiN) $Y$, 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. $d$ is simply the total number of events in the time period of interest.

Rates are reported both as age-specific rates and age-standardised rates per 100,000 person-years. The population is divided into 18 distinct 5-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

$$ r_i = \frac{d_i}{10^5}. $$

The age-standardised rate (ASR) is calculated as

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

where $w_i$ is the weight assigned to age group $i$. The weights are typically designed to reflect the distribution of the population across different age groups.
where $w_i$ is a weight given by some reference population. Typically the World Standard Population has been used (Doll & al, 1966). Cancer in Norway 2015 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 mid-year 2014 population are given in the table below. One should be aware that the world standard upweights the younger age groups and downweights the older age groups compared to the most recent Norwegian population.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Age</th>
<th>World (1960)</th>
<th>Norway (2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-4</td>
<td>12,000</td>
<td>6,039</td>
</tr>
<tr>
<td>2</td>
<td>5-9</td>
<td>10,000</td>
<td>6,102</td>
</tr>
<tr>
<td>3</td>
<td>10-14</td>
<td>9,000</td>
<td>5,993</td>
</tr>
<tr>
<td>4</td>
<td>15-19</td>
<td>9,000</td>
<td>6,349</td>
</tr>
<tr>
<td>5</td>
<td>20-24</td>
<td>8,000</td>
<td>6,681</td>
</tr>
<tr>
<td>6</td>
<td>25-29</td>
<td>8,000</td>
<td>6,770</td>
</tr>
<tr>
<td>7</td>
<td>30-34</td>
<td>6,000</td>
<td>6,690</td>
</tr>
<tr>
<td>8</td>
<td>35-39</td>
<td>6,000</td>
<td>6,670</td>
</tr>
<tr>
<td>9</td>
<td>40-44</td>
<td>6,000</td>
<td>7,296</td>
</tr>
<tr>
<td>10</td>
<td>45-49</td>
<td>6,000</td>
<td>7,207</td>
</tr>
<tr>
<td>11</td>
<td>50-54</td>
<td>5,000</td>
<td>6,492</td>
</tr>
<tr>
<td>12</td>
<td>55-59</td>
<td>4,000</td>
<td>6,108</td>
</tr>
<tr>
<td>13</td>
<td>60-64</td>
<td>4,000</td>
<td>5,575</td>
</tr>
<tr>
<td>14</td>
<td>65-69</td>
<td>3,000</td>
<td>5,369</td>
</tr>
<tr>
<td>15</td>
<td>70-74</td>
<td>2,000</td>
<td>3,702</td>
</tr>
<tr>
<td>16</td>
<td>75-79</td>
<td>1,000</td>
<td>2,663</td>
</tr>
<tr>
<td>17</td>
<td>80-84</td>
<td>500</td>
<td>2,063</td>
</tr>
<tr>
<td>18</td>
<td>85+</td>
<td>500</td>
<td>2,231</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>100,000</td>
<td>100,000</td>
</tr>
</tbody>
</table>
1.3 Cumulative risk and prevalence

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

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

(3)

where \(N\) is the age group corresponding to the age of interest. In CIIC CR is estimated up to the age of 74, so \(N = 15\).

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.

1.4 Relative survival (Net survival)

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

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

(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). Here 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 survival is calculated as

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

(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.

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

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

where

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

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 population life tables matched on gender, age, and calendar year. We have used unsmoothed lifetables.

To reduce the potential for bias and to facilitate comparisons over time the relative 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. The age-standardised relative survival $R_s(t)$, the estimate of net survival, is given by

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

where $w_j$ is the weight for age group $j$, determined by the age distribution of patients diagnosed during the most recent 5-year period.

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 smaller and more lethal cancer sites we still cannot avoid sparse data, particularly when estimating 15-year net survival. When the standard weighting cannot be used due to sparse data, we obtain age-standardised estimates by using the alternative weighting method proposed by Brenner (Brenner, 2004). This method assigns individual weights to all patients, and therefore does not depend on obtaining estimates for net survival in each age group. The method works by upweighting patients that are under-represented in the study population compared with the reference population and downweighting patients in age groups that are over-represented. As an example, if the weight for the youngest age group is 0.24 in the reference population, and 0.15 in the study population. Each patient in this age group would then be assigned a weight of $\frac{0.24}{0.15} = 1.6$.

As mentioned in the Methods-section in Cancer in Norway 2015, 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 2011-2014 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 2011 based on a 5-year period window from 2007-2011 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 2007-2011, complete 5-year follow-up is available for patients diagnosed 2007-2010, whereas only 4 year follow-up is available for patients diagnosed 2011. To make up for this lack of follow-up patients diagnosed before 2007 is considered at risk from January 1st 2011. This means that only the survival experience between year 4 and 5 for
patients diagnosed 2006 is used when estimating the 5-year relative survival for the 2007-2011 cohort, exactly making up for the last year of follow-up lacking for the 2011 patients.

All relative survival analysis were performed using the the Stata program strs.

2 References

Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age ad-
2322.

Dickman P W, Coviello E. Estimating and modeling relative survival. The Stata

Doll R, Payne P, Waterhouse J. (Eds) Cancer Incidence in Five Continents: A

Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival
computations, methodological note 10. End Results Evaluation Section, Na-
tional 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 Methodol-
ogy 2015. http://www.biomedcentral.com/1471-2288/15/64

StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp
LP 2015