Relative Survival Estimators

Karri Seppä

Finnish Cancer Registry, Helsinki, Finland

State of Art of Methods for the Analysis of Population-Based Cancer Data

Ispra, 22–23 January 2014
Introduction

Relative survival ratio

- A basic step for describing cancer survival in population-based data
- Definition: ratio between the observed survival proportion in a patient group and the expected survival proportion
  - Expected survival is calculated from a reference population that is sufficiently similar to the patient group with respect to characteristics affecting the mortality due to other causes than the cancer, but free of the cancer under study.
  - No information on the actual causes of death is needed.
- Three classical estimators: Ederer I and II, and Hakulinen
  - Biased estimators of net survival
Interpretations of relative survival ratio:

1. Compares patients’ cumulative survival against a theoretical maximum
   - An observable quantity that has a clear interpretation, but depends on expected survival.

2. Net survival probability i.e. hypothetical survival probability, when the cancer is the only cause of death and all other causes could be eliminated
   - A useful concept, because net survival is independent of the competing mortality
   - The time to death from the cancer must be independent of the competing mortality
   - Stratified analyses, e.g. age standardisation of relative survival ratios, have been used to reduce bias of the classical estimators of net survival
     - Ederer II method was recommended (Hakulinen et al. 2011), because it is often close to the age-standardised relative survival ratio
Relative survival estimators

Relative survival ratio $S_R(t)$ in a group of $n$ patients is the ratio between the averages of the patient-specific observed and expected survival probabilities $S_{Oi}(t)$ and $S_{Pi}(t)$, respectively:

$$S_R(t) = \frac{S_O(t)}{S_P(t)} = \frac{\frac{1}{n} \sum_{i=1}^{n} S_{Oi}(t)}{\frac{1}{n} \sum_{i=1}^{n} S_{Pi}(t)}.$$

The classical estimators of expected survival can be written as $S_P(t) = \exp \left( - \int_0^t h_P(u) \, du \right)$ where the expected hazard of death $h_P(t)$ is estimated as a weighted average of patient-specific expected hazards $h_{Pi}(t)$:

$$h_P(t) = \sum_{i=1}^{n} \frac{W_i(t)h_{Pi}(t)}{\sum_{i=1}^{n} W_i(t)}.$$
- Ederer I method (Ederer et al. 1961)
  - \( W_i(t) = S_{Pi}(t) \)
  - Unbiased estimator of expected survival
  - Biased estimator of relative survival ratio, if censoring mechanism is informative

- Ederer II method (Ederer and Heise 1959)
  - \( W_i(t) = \begin{cases} 
  1 & \text{if } t < \text{observed follow-up time of patient } i \\
  0 & \text{otherwise} 
\end{cases} \)
  - Biased estimator of relative survival ratio as the expected survival depends on the observed survival
  - Estimates the ‘observable net survival’ (similar to cause-specific survival)
Relative survival estimators

Example data

Female patients diagnosed with colon cancer in Finland in 1970–1989.

- 8544 patients

Two different closing years for assessing the effect of administrative censoring of observed survival

- 1989 (incomplete follow-up)
- 2011 (complete follow-up for the first 22 years)

Five age groups for age standardisation and for estimation of distribution of potential follow-up times

- 0–44, 45–54, 55–64, 65–74 and 75+ years at diagnosis
Ederer I and II estimators in incomplete follow-up
Informative censoring of observed survival

Observed survival $S_O(t)$ is often estimated using the Kaplan-Meier or the actuarial method.

- Biased estimator of observed survival, if censoring mechanism is informative

Censoring is called informative if it is associated with the outcome event.

- Administrative censoring can be informative, because patients’ potential follow-up times (i.e. time from diagnosis to a common closing date) are heterogeneous
Two sources of bias due to informative censoring

1. Distributions of variables that relate to survival have changed during the period of diagnosis
   - Typically old patients have shorter potential follow-up times, because relatively more old patients are diagnosed in the end of the diagnosis period due to aging of population
     - Observed survival is overestimated
   - Corrections: stratified analyses, Hakulinen method, inverse probability weighting (Rebolj Kodre and Pohar Perme 2013)

2. Prognosis has changed during diagnosis period
   - Typically prognosis has improved over time
     - Observed survival is underestimated
   - Corrections are subject to pure guessing
Relative survival estimators

- Ederer I method (Ederer et al. 1961)
  - \( W_i(t) = S_{Pi}(t) \)
  - Unbiased estimator of expected survival
  - Biased estimator of relative survival ratio, if censoring mechanism is informative

- Ederer II method (Ederer and Heise 1959)
  - \( W_i(t) = \begin{cases} 
  1 & \text{if } t < \text{observed follow-up time of patient } i \\
  0 & \text{otherwise}
  \end{cases} \)
  - Biased estimator of relative survival ratio as the expected survival depends on the observed survival
  - Estimates the ‘observable net survival’ (similar to cause-specific survival)

- Hakulinen method (1982)
  - \( W_i(t) = \begin{cases} 
  S_{Pi}(t) & \text{if } t < \text{potential follow-up time of patient } i \\
  0 & \text{otherwise}
  \end{cases} \)
  - Reduces bias of the Ederer I estimator by introducing a similar bias to the expected survival
Reducing bias due to informative censoring

Relative survival estimators

- Ederer I
- Ederer II
- Hakulinen
Inverse probability weighting

A correction for informative censoring (Rebolj Kodre and Pohar Perme 2013)

- The weights are proportional to the inverse probability of censoring $1/S_{Gi}(t-)$
- The distribution of potential follow-up times $G_i$ has to be estimated from a sample.

- Weighted Ederer I estimator
  - Observed survival is estimated as follows
    \[
    S^w_O(t) = \exp \left( - \sum_t \sum_{ij} \frac{d_{ij}}{S_{Gi}(t^-)} \right)
    \]

- Weighted Ederer II estimator
  - The same estimator $S^w_O(t)$ for observed survival
  - In expected survival, the patient-specific expected hazards $h_{Pi}(t)$ are weighted as follows
    \[
    W_i(t) = \begin{cases} 
    \frac{1}{S_{Gi}(t^-)} & \text{if } t < \text{observed follow-up time of patient } i \\
    0 & \text{otherwise}
    \end{cases}
    \]
Reducing bias due to informative censoring

Relative survival estimators

Time from diagnosis (years)

Relative survival ratio (%)

Ederer I
Ederer II
Hakulinen
Weighted Ederer I
Weighted Ederer II
Relative survival estimators

Incomplete vs. complete follow-up

Relative survival ratio (%)

Ederer I
Ederer II
2 5 10 15 20
35
40
45
50
55
Time from diagnosis (years)

Hakulinien
Weighted Ederer I
Weighted Ederer II
Ederer I, complete follow-up
Ederer II, complete follow-up
Age-standardised relative survival estimators

The hazard of death due to cancer and due to other causes depend on the same demographic covariates (censoring due to competing mortality is informative).

- The independence assumption of net survival may be more reasonable if it is made conditional on important determinants of survival.

- The internal age standardisation:

  \[ S_E(t) = \sum_{a=1}^{k} \frac{W_a S_{Ra}(t)}{\sum_a W_a} \]

  where \( W_a \) is the number of patients diagnosed and \( S_{Ra}(t) \) is the relative survival ratio in age group \( a \).

  - Unbiased estimator of net survival if the time to death from the cancer is independent of the competing mortality within each age group.
Age-standardised relative survival ratios

Relative survival estimators

Ederer I
Ederer II
Hakulinen
Age-standardised Ederer I
Age-standardised Ederer II
Age-standardised Hakulinen

Time from diagnosis (years)
Relative survival ratio (%)

Ederer I
Hakulinen
Ederer II
Relative survival estimators

Age-standardised relative survival ratios

Complete follow-up
- Ederer I
- Age-standardised Ederer I
- Age-standardised Ederer II
- Ederer II

Relative survival ratio (%)

Time from diagnosis (years)
Conclusions

- The weighted Ederer I estimator proposed by Rebolj Kodre and Pohar Perme (2013) is the preferrable estimator of relative survival ratio under informative censoring.

- In practical applications, estimation of net survival is a more central goal than comparison of patients’ survival against a theoretical maximum.

- Estimation of long-term net survival is problematic: the few observations in old age groups get large weights, because the competing mortality does not leave for older ages sufficiently sizable materials on which to base reliable estimation.
Conclusions

- The non-standardised Ederer II estimator often slightly overestimates the net survival but the magnitude of bias is much smaller than in the other classical relative survival estimators.
  - A useful index number of cancer survival that summarises the excess mortality observed in a patient population.
References