Web-Assisted Estimation of Relative Survival (WAERS) tool methodology

1. Observed Survival (OS)

This tool uses the Kaplan-Meier method [1,2] to calculate the observed survival.

Kaplan-Meier method (KM):

Let us define:

- \( t_i \) is the \( i \)-th death time observed, \( 1 \leq i \leq I \)
- \( d_i \) is the number of deaths observed in time \( t_i \)
- \( r_i \) is the number of censored observations on the interval \([t_i, t_{i+1})\)
- \( n_i = n_{i-1} - d_{i-1} - r_{i-1} \) is the number of individuals at risk just before deaths have occurred in time \( t_i \). To obtain \( n_i \), the deaths that occur in time \( t_{i-1} \) are extracted from \( n_{i-1} \) and also all other cases for which they have finished the follow-up time in the interval \([t_{i-1}, t_i)\).

Then, the conditional probability of being alive in the interval \((t_i, t_{i+1}]\) given that the subject is at risk at time \( t_i \) is: \( s_i = 1 - (d_i/n_i) \)

And, therefore, the probability of surviving after time \( t \) is:

\[
OS_i = \prod_{j=1}^{i} s_j
\]

This probability remains constant throughout the period \( t_{i+1} \).

2. Expected Survival (ES)

The Expected Survival (ES)[3] is the survival that we expect to have in a cohort according to the population mortality from which the individuals come from and regardless of cause of death.

We define \( \lambda_{ij} \) as the overall mortality rate for the individuals in the \( i \)-th age group and \( j \)-th year in the residence area from the cohort individuals (see section 4). Then, \( \mu_{ij} \) is the population survival rate in the \( j \)-th year for the individuals from the \( i \)-th age group, and it is estimated as:

\[
\hat{\mu}_{ij} = e^{-\lambda_{ij}}
\]
Once the survival rates are determined, these are applied at the cohort. From the above formula, ES are calculated by Hakulien method[4].

**Hakulinen Method:**

Hakulinen Method, also called long-term[3,5], estimates ES by using a similar estimator to KM but considering censorships. The main concept is that the number of individuals at risk for each interval is calculated taking into account the expected number of dropouts during that time.

Let us define:

- $r_i$: the number of individuals at risk in time $t_i$
- $r_i^*$: the estimator of the number of individuals at risk. It is calculated as the sum of population survival rates (formula 1) from each risk individuals ($r_i$).

Finally, the ES estimator by Hakulinen method is calculated as:

$$ES(t_i) = \prod_{j|t_j \leq t_i} \left(1 - \frac{d_j^*}{r_j^* - \frac{1}{2} w_j}\right)$$

Where:

- $d_i^*$ is the expected number of deaths: $d_i^* = r_i - r_i^*$
- $w_i$ is the time interval (range)
- The expected number of dropouts in the interval $w_i$ is calculated applying a correction to the sum of population survival rates (1) from the risk individuals in this interval whose follow up doesn’t exceed this interval.

**3. Relative Survival (RS)**

The relative survival (RS)[1,3] is interpreted as the proportion of individuals that has survived under the hypothesis that the illness of study was the unique and possible cause of death. For the calculation of RS it is necessary the calculation of OS and ES and it is calculated as:

$$RS = \frac{OS}{ES}$$

Considering that RS is a ratio, we could obtain higher values than 1, meaning the study individuals’ survival is higher than the reference general population.
Confidence interval for RS(t_i):

First, the variance of OS(t_i) is calculated, then, considering its lower and upper limit and ES(t_i) as a constant, RS(t_i) is calculated.

Applying the Greenwood formula[6,7], the OS(t) variance is approximated as:

\[
\text{Var}[\text{OS}(t_i)] \approx [\text{OS}(t_i)]^2 \sum_{j=1}^{i} \left( \frac{d_j}{n_j(n_j - d_j)} \right)
\]

Where \(j\) indicates all the previous events in the time (or intervals) until the “i” moment, \(d\) is the number of deaths in each moment and \(n\) is the number of individuals at risk in each moment. So, the standard error is:

\[
\text{EE}[\text{OS}(t_i)] \approx \text{OS}(t_i) \sqrt{\sum_{j=1}^{i} \left( \frac{d_j}{n_j(n_j - d_j)} \right)}
\]

Thus, the 95% confidence interval for the OS in the t_i time is:

\[
\text{LL(OS(t_i))} = \text{OS}(t_i) - 1.96 \times \text{EE}[\text{OS}(t_i)] \quad ; \quad \text{UL(OS(t_i))} = \text{OS}(t_i) + 1.96 \times \text{EE}[\text{OS}(t_i)]
\]

And considering ES(t_i) as a constant, the RS(t_i) 95% confidence interval is obtained as:

\[
\text{LL(RS(t_i))} = \frac{\text{LL(OS(t_i))}}{\text{ES(t_i)}} \quad ; \quad \text{UL(RS(t_i))} = \frac{\text{UL(OS(t_i))}}{\text{ES(t_i)}}
\]
4. Information sources to calculate mortality rates

Specific rates by age group and year from the area you want to evaluate are needed to calculate the ES. This tool provides specific rates for the autonomous communities and provinces of Spain, as well as for whole Spain.

The national WAERS is updated based on data from the Instituto Nacional de Estadística (INE) available from [http://www.ine.es/](http://www.ine.es/). The last updated of WAERS contains mortality specific rates from 1991 to 2013 by 21 age groups (0-1 year, 1-4 years, …, 90-94 years, >95 years).

The information is processed by R free software to obtain a file structure to be included in the WAERS tool. The user only has to select the population area of which wants to calculate the relative survival.
REFERENCES


