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 λ_{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, μ_{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:

$$\hat{\mu}_{ij} = e^{-\lambda_{ij}} \tag{1}$$

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:

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

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$_i$) 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:

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

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

$$\text{LL}(\text{RS}(t_i)) = \frac{\text{LL}(\text{OS}(t_i))}{\text{ES}(t_i)} \quad ; \quad \text{UL}(\text{RS}(t_i)) = \frac{\text{UL}(\text{OS}(t_i))}{\text{ES}(t_i)}$$

4. Information sources to calculate mortality rates

To calculate the ES specific rates by age group and year are needed from the area you want to evaluate. This tool provides specific mortality rates for some countries since 1994 (Table 1) in 18 age groups (0-4, 5-9, …, 80-84, +85).

The international WAERS data is updated from World Health Organization (WHO) (http://www.who.int/healthinfo/statistics/mortality_rawdata/en/) and from United Nations (UN)[8] (https://esa.un.org/unpd/wpp/).

WHO provides mortality and population data, but for each country this data could be structured in different age groups and years. In addition, for one particular year, we can find registered mortality data but not population data or vice versa. Therefore, the data must be first structured and completed and then the mortality rates can be calculated and included in WAERS tool.

To solve the problem of having different age groups between mortality and population, among different years or among different countries, data is transformed in 18 age groups (0-4, 5-9, ..., 80-84, +85).

The process of data completion is not the same for deaths and population. For mortality data, a generalized linear model is fit, adjusted for each country, gender and age group and considering a Poisson distribution [63] for deaths. With this model the mortality data is estimated for empty years. As a WHO recommendation, the UN population estimations are used to complete empty years from WHO population.

Likewise, mortality and population gaps are filled for each country since 1994 (at least) until maximum data available in each country.

The available mortality and population periods depend on each country. Table 1 shows available data and countries included in WAERS.

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

Table 1. Countries which mortality rates are included in WAERS and its availability per year

País	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Argentina																					
Australia																					
Austria																					
Belgium																					
Bulgaria																					
Canada																					
Croatia																					
Cuba																					
Czech R.																					
Denmark																					
Estonia																					
Finland																					
France																					
Germany																					
Greece																					
Hungary																					
Iceland																					
Ireland																					
Italy																					
Japan																					
Latvia																					
Lithuania																					
Luxembourg																					
Malta																					
Netherlands																					
New Zealand																					
Norway																					
Poland																					
Portugal																					
Rep. Moldova																					
Romania																					
Slovakia																					
Slovenia																					
Spain																					
Sweden																					
Macedonia																					
UK																					
USA																					

Czech R.: Czech Republic**; Rep. Moldova:** Republic of Moldova**; UK:** United Kindom**; USA:** United States of America

- WHO population
- UN estimated population
- WHO mortality
- Own estimated mortality

REFERENCES

1. Esteve J, Benhamou E, Raymond L. Statistical Methods in Cancer Research. Volume. IV. Descriptive Epidemiology. IARC Sci Publ. Lyon; 1994;128:1–302.

2. Llorca J, Delgado-Rodríguez M. Análisis de supervivencia en presencia de riesgos competitivos: estimadores de la probabilidad de suceso. Gac Sanit. 2004;18(5):391–7.

3. Cleries R, Ribes J, Moreno V, Esteban L, Pareja L, Galvez J, et al. Cálculo de la supervivencia relativa. Comparación de métodos de estimación de la supervivencia esperada. Gac Sanit. 2006 Jul;20(4):325–31.

4. Capocaccia R, Gatta G, Roazzi P, Carrani E, Santaquilani M, De Angelis R, et al. The EUROCARE-3 database: Methodology of data collection, standardisation, quality control and statistical analysis. Ann Oncol. 2003;14(Supplement 5):14–27.

5. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. Biometrics. 1982;38(4):933–42.

6. Szklo M, Nieto J. Epidemiología intermedia. Conceptos y aplicaciones. Díaz de Sa. Madrid; 2003. 435 p.

7. Clèries R, Valls J, Esteban L, Gálvez J, Pareja L, Sanz X, et al. WAERS: an application for Web-assisted estimation of relative survival. Med Inform Internet Med. 2007 Sep;32(3):169–75.

8. United Nations. Department of Economic and Social Affairs. Population Division. World Population Prospects, the 2015 Revision [Internet]. [cited 2016 Nov 6]. Available from: https://esa.un.org/unpd/wpp/