

Computing relative survival & Estimating net survival *What are the routinely use methods doing?* Jacques Estève

Ljubljana 10 February 2011

Survival Measures according to the NCI

<http://srab.cancer.gov/survival/measures.html>

- The table suggests that there are two methods of estimation of the *net* and *crude* measure

- The distinction is whether or not a cause of death has been assigned to each deceased patient and was known to the investigator.
- There is little discussion on what these *ad-hoc* calculations are estimating

Estimation Method	Measure	
	Net	Crude
Cause of death	Cause specific survival	Crude probability of death using cause of death information
Expected survival	Relative survival	Crude probability of death using expected survival

10 February 2011

HCL/UCB LYON

2

The widely discussed questions

- Cause specific survival
 - What is an objective definition of the cause of death?
 - Will two different observers assign the same cause of death to a given deceased patient? If not, what about survival comparability between region, countries, etc...
- Relative survival
 - How to define a similar group of patients free of the disease?
 - How to calculate the *expected survival*?
- It is widely believed that the two methods are estimating the "*net survival*"

10 February 2011

HCL/UCB LYON

3

Crude Risk, relative survival, net survival What do they mean?

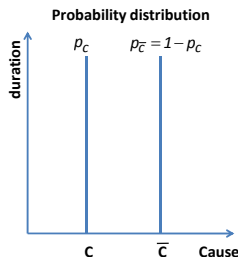
- The first and third concepts are usually defined in the framework of the *competing risk* theory.
- The second concept has never been clearly defined. Loosely speaking its purpose is to compare the observed survival to the survival which would have been observed in the absence of disease, the "*expected survival*".
- While the "*expected survival*" *of a person* is relatively easy to define, it is less simple to compare observed and expected survival *of a group* of subjects having very different expectation of remaining lifetime

10 February 2011

HCL/UCB LYON

4

The competing risk framework



- A patient may die
 - From his/her disease
 - From any other cause
- A bi-variate random description:
 - (C, T) :
C cause of death,
T survival duration
- The probability is distributed on two vertical lines

10 February 2011

HCL/UCB LYON

5

The death risks

- Risk of dying from one of two causes

$$R(t) = P(T \leq t) = P(C = c, T \leq t) + P(C = \bar{c}, T \leq t) \\ = R_c(t) + R_{\bar{c}}(t)$$

- $R(t)$ is a genuine probability distribution, i.e.: a function increasing from 0 to 1
- On the contrary $R_c(t)$ and $R_{\bar{c}}(t)$ are not: their respective maximum are p_c and $1 - p_c$
- $R_c(t)$ and $R_{\bar{c}}(t)$ are called the *crude risk of death* of the given cause

10 February 2011

HCL/UCB LYON

6

The crude survival

- It is defined by: $S_c(t) = P(C=c, T > t)$
- The relation $P(C=c, T \leq t) + P(C=c, T > t) = P(C=c)$ shows:

$$R_c(t) + S_c(t) = p_c$$
- And also: $S(t) = 1 - R(t) = S_c(t) + S_{\bar{c}}(t)$
- In words: the *survival probability* is the sum of the *crude survival probability*
- Note that the survival distribution for the cause "c" is not $S_c(t)$ but

$$P(T \leq t | C=c) = S_c(t)/p_c$$

10 February 2011

HCL/UCB LYON

7

The mortality rate for the cause "c"

- It is the probability that the subject dies from cause "c" on date t given that he is *still under observation* at this date:

$$\lambda_c(t) \Delta t = P(C = c, t < T < t + \Delta t | T > t)$$

$$\Rightarrow \lambda_c(t) = \frac{R'_c(t)}{S(t)}$$

$$\Rightarrow R_c(t) = \int_0^t S(u) \lambda_c(u) du$$

- A very intuitive relation between crude risk of dying from cause "c" and the survival probability

10 February 2011

HCL/UCB LYON

8

The net survival probability

- From the previous results:

$$\lambda_c(t) + \lambda_{\bar{c}}(t) = \frac{R'_c(t)}{S(t)} + \frac{R'_{\bar{c}}(t)}{S(t)} = \frac{R'(t)}{S(t)} = \lambda(t) \left(= -\frac{S'(t)}{S(t)} \right)$$

- Therefore $S(t)$ may be written:

$$S(t) = \exp\left[-\int_0^t \lambda(u) du\right] = \exp\left[-\int_0^t \lambda_c(u) du\right] \times \exp\left[-\int_0^t \lambda_{\bar{c}}(u) du\right] \\ = \tilde{S}_c(t) \times \tilde{S}_{\bar{c}}(t)$$

- i.e.: as the product of **net survival probabilities**, which are defined by cause specific mortality rates acting alone (i.e.: *in the absence of the other*)

10 February 2011

HCL/UCB LYON

9

Why net survival?

- The net survival is an imaginary survival function,
 - Meaning that it is not observable and will never happen, but
- it is what would be observed in the fictitious situation where
 - the patients would only be subject to the mortality from the disease of interest and
 - the mortality rate of this disease would remain the same as it was in the context of competing risk, which is the only situation which can be observed
- Its importance stems from its independence from other causes of death and its ability to serve for comparison

10 February 2011

HCL/UCB LYON

10

The relative survival and the estimation of net survival

- If we supposed that the probability of surviving "other causes" of death is given by the life table, i.e.: $\tilde{S}_{\bar{c}}(t) = S_p(t)$
- The relative survival**, $S_R(t) = \frac{S(t)}{S_p(t)}$ seems to be equal to the net survival. (c.f.: slide #9)
- Therefore it was thought that estimating $S(t)$ with the usual methods (Kaplan-Meier, actuarial...) would provide an estimate of the *net survival*...
- We shall see below that the *relative survival* is an estimator of the *net survival* in particular situations only.

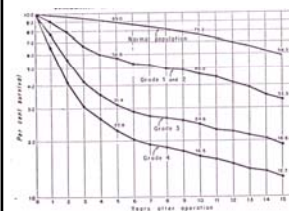
10 February 2011

HCL/UCB LYON

11

An historical publication

- Cancer of the stomach
(2682 cases of the mayo clinics)
- Joseph BERKSON & Robert GAGE (JASA 1952)
- The idea of the relative survival was to compare the "normal population" survival to the observed survival.
- For these authors, the main measure was the death rates associated with the observed and expected survival curves



Note that the vertical axis has a logarithmic scale

10 February 2011

HCL/UCB LYON

12

Two main approaches

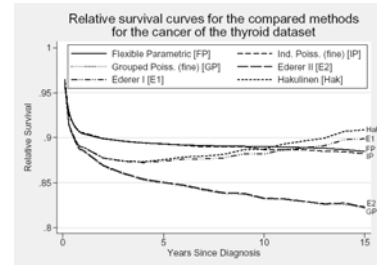
- The relative survival
 - $S_R(t) = S_O(t) / S_P(t)$ seen initially as the percentage of the “normal” survival
 - ⇒ EDERER I or HAKULINEN
- The additive model for the death rates
 - $\lambda_{OI}(t) = \lambda_{PI}(t) + \lambda_C(t)$
 - ⇒ EDERER II, Cause-specific, Maximum likelihood
- If $\lambda_C(t)$ is the same for all subjects in the group, the two approaches are equivalent and estimate the net survival. If it is not the case, they may be very different

10 February 2011

HCL/UCB LYON

13

Behavior of the usual estimators



Source :M.J. Rutherford, P.C. Lambert and P.W. Dickman, poster presented at a workshop on population survival (Frascati sept 2010)

10 February 2011

HCL/UCB LYON

14

The population net survival

- When the death rate from the disease is specific to each patient (i indexes λ_C):

$$\lambda_{OI}(t) = \lambda_{CI}(t) + \lambda_{\bar{C}i}(t)$$

- The net survival (non “observable”) of the i^{th} subject

$$S_{CI}(t) = \exp\left(-\int_0^t \lambda_{CI}(t) dt\right)$$

- The net survival of the population (N subjects)

$$S_C(t) = \frac{1}{N} \sum_{i=1}^N S_{CI}(t)$$

- The net sub-hazard of the population (what we wish to know):

$$\lambda_{C_0}(t) = \frac{\sum_i S_{CI}(t) \lambda_{CI}(t)}{\sum_i S_{CI}(t)}$$

10 February 2011

HCL/UCB LYON

15

The population value in the cause specific setting

- If a cause of death is assigned to each subject and we know it,

- We observe the bi-variate distribution $[T, CA]$, with the hazard

$$\lambda_{C_1}(t) dt = \Pr(T < t + dt, CA = C | T > t) = \frac{\sum_i S_{OI}(t) \lambda_{CI}(t) dt}{\sum_i S_{OI}(t)}$$

- This differs from $\lambda_{C_0}(t)$ and is clearly the hazard associated to the cause-specific survival, it may be written:

$$\lambda_{C_1}(t) = \lambda_{OI}(t) - \frac{\sum_i S_{OI}(t) \lambda_{\bar{C}i}(t)}{\sum_i S_{OI}(t)}$$

10 February 2011

HCL/UCB LYON

16

The population value in the cause specific setting, when the cause of death is unknown

- The principle of the relative survival is to consider *the death rate of other causes as known* and given by the life table of the population in which the patients were diagnosed: $\lambda_{\bar{C}_i}(t) = \lambda_{p_i}(t)$

$$\lambda_{C_2}(t) = \lambda_O(t) - \frac{\sum_i S_{O_i}(t) \lambda_{p_i}(t)}{\sum_i S_{O_i}(t)}$$

- The second term is the average of life table rates for those who are still alive at time t .
- This is clearly the *population value* estimated by the EDERER II method

10 February 2011

HCL/UCB LYON

17

The population value in the relative survival setting

- The population value of the relative survival is

$$S_R(t) = \frac{\sum_i S_{O_i}(t)}{\sum_i S_{p_i}(t)}$$

- Where the sum is over the N subjects of the patient population
- denoting $S_{R_i}(t) = S_{O_i}(t)/S_{p_i}(t)$ the “individual” relative survival the relative survival is written

$$S_R(t) = \frac{\sum_i S_{p_i}(t) \times S_{R_i}(t)}{\sum_i S_{p_i}(t)}$$

–

10 February 2011

HCL/UCB LYON

18

... ..

- If the subjects have the same expected survival or if the subjects have the same relative survival
 - The relative survival is the average of the individual relative survival
 - Otherwise the relative survival is closer to that of the patients who have the best “population survival”
- Even if $S_R(t)$ is not necessarily a survival distribution the derivative of $-\log(S_R(t))$ is still of interest to evaluate the dynamic of the relative survival, it is given by:

$$\lambda_R(t) = -\frac{S_R'(t)}{S_R(t)} = \lambda_O(t) - \frac{\sum_i S_{p_i}(t) \lambda_{p_i}(t)}{\sum_i S_{p_i}(t)}$$

10 February 2011

HCL/UCB LYON

19

Where are we?

Population value

$$\lambda_{C_0}(t) = \frac{\sum_i S_{C_i}(t) \lambda_{C_i}(t)}{\sum_i S_{C_i}(t)}$$

$$\left\{ \begin{array}{l} \lambda_{C_1}(t) = \lambda_O(t) - \frac{\sum_i S_{O_i}(t) \lambda_{\bar{C}_i}(t)}{\sum_i S_{O_i}(t)} \\ \lambda_{C_2}(t) = \lambda_O(t) - \frac{\sum_i S_{O_i}(t) \lambda_{p_i}(t)}{\sum_i S_{O_i}(t)} \end{array} \right.$$

$$\lambda_R(t) = \lambda_O(t) - \frac{\sum_i S_{p_i}(t) \lambda_{p_i}(t)}{\sum_i S_{p_i}(t)}$$

Estimators

- ???

- “cause-specific”
- EDERER II

- EDERER I or HAKULINEN

10 February 2011

HCL/UCB LYON

20

The construction of the estimators

- The population values of the death rates of interest are similarly defined. They differ by the quantity which is subtracted to the *observable* death rate
- Their estimators can be constructed in the same way using the counting process toolkit
 - For example for λ_{C_1} we get the estimator:

$$\hat{\lambda}_{C_1}(t) = \int_0^t \frac{dN(u)}{Y(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i(u) d\Lambda_{pi}}{Y(u)}$$
 - That is the *Nelson estimate* minus the cumulative average population death rate of those *still at risk at each death time*, in other word the EDERER II estimate

10 February 2011

HCL/UCB LYON

21

Do Relative and net survival agree?

- As estimator of *net survival*, all the above estimators are biased!
- This is fairly well understood in the cause specific setting, since the censoring mechanism (other causes of death) is “informative”: the cause of interest and the other causes share usually the influence of the same covariates.
- The biases of EDERER II and of the cause specific estimate are identical if the relation $\lambda_{\bar{C}_i}(t) = \lambda_{pi}(t)$ is true.
- It is possible to correct these two estimates using multivariate models adjusting for “demographic” covariates

10 February 2011

HCL/UCB LYON

22

An extreme example

- The following example is intentionally extreme and will not be met in practice.
- Its purpose is to separate clearly *relative* and *net* survival
- It is based on competing exponentials and a mixture of two groups with different rates and a positive association between cancer and population rates
 - Group 1 $\lambda_c = 0.01$ $\lambda_p = 0.05$
 - Group 2 $\lambda_c = 0.4$ $\lambda_p = 0.3$
 - *No censoring*

10 February 2011

HCL/UCB LYON

23

The example set-up

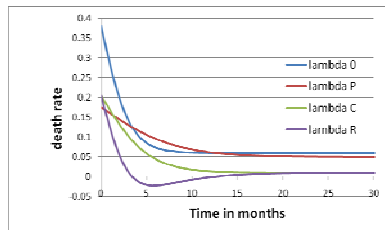
- Since S_o and S_p are known in each group it is possible to calculate the true population values (relative survival, net survival, and the associated rates)
- We calculated EDERER I & II and specific survival estimates for one simulation of this model
- Since there is no censoring HAKULINEN and EDERER I estimates are identical
- At “long term” follow-up only are group 1 subjects surviving: relative and net death rates converge both to group 1 net death rate 0.01

10 February 2011

HCL/UCB LYON

24

Population values of the death rates

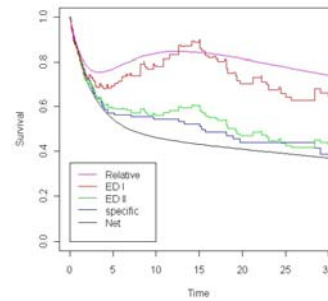


10 February 2011

HCL/UCB LYON

25

The survival "Curves" & their estimates



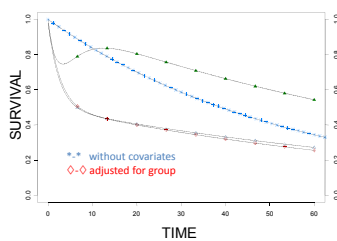
- As expected EDERER I estimates relative survival
- Specific and EDERER II survival are close together and are a biased estimate of net survival
- This bias can be corrected to obtain unbiased estimates of net survival
- As shown in the following slides, this correction may be obtained from a multivariate model adjusted for "demographic" variables

10 February 2011

HCL/UCB LYON

26

ML correction of bias



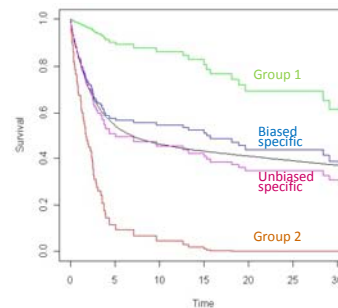
- ML estimate using Remontet&Estève model. Baseline rate estimated with a cubic spline.
- As expected the model without covariates gives a strongly biased net survival estimate.
- The bias is perfectly corrected, when adjusting the model for "demographic" variables (here the group)

10 February 2011

HCL/UCB LYON

27

Correction of bias in the cause specific setting



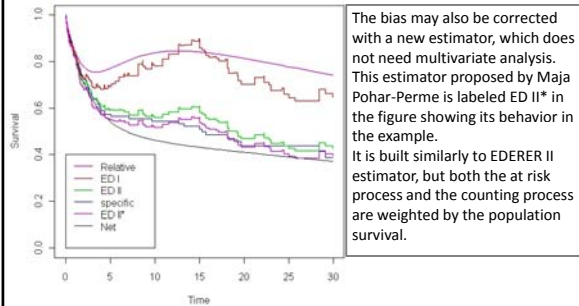
- Estimation in the cause specific setting using the Cox model adjusted for group.
- The classical approach considering other causes of death as leading to censoring events is biased
- The average of group 1&2 cox estimates is unbiased as an estimate of net survival

10 February 2011

HCL/UCB LYON

28

A new unbiased estimator



10 February 2011

HCL/UCB LYON

29

The new estimator in detail (I)

- The original idea is due to Robins and Satten* : if the censoring process shares covariates with the survival process, but is conditionally independent of the survival process, the survival can be estimated with the usual approach, after a simple modification of the Counting and at Risk processes.
- The modification is a *weighting of the two processes* by the conditional probability of surviving the censoring process given the covariates (*or at least something similar in case of time-dependant covariates*):

*SATTEN *et al. Statistics & probability letters* 54 (2001) 397-403

10 February 2011

HCL/UCB LYON

30

The new estimator in detail (II)

- In the general case the weight has to be estimated. In the case of **relative** survival the weight is known and the weighting is trivial.
- It provides an unbiased estimator for the cause specific estimator if the survival from other causes of deaths are given by the life table.
- And a new estimator for the *net survival*, when the cause of death is not available, by weighting the EDERER II estimate.

10 February 2011

HCL/UCB LYON

31

The new estimator in detail (III)

- Define the two weighted processes
 - $N_i^*(t) = N_i(t)/S_{pi}(t)$ et $Y_i^*(t) = Y_i(t)/S_{pi}(t)$
 - $N^*(t) = \sum N_i^*(t)$; $Y^*(t) = \sum Y_i^*(t)$
- Calculate the weighted estimate in introducing this weighted process in the formula of the EDERER II estimator.

$$\hat{\Lambda}(t) = \int_0^t \frac{dN^*(u)}{Y^*(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i^*(u) \lambda_{pi}(u)}{Y^*(u)} du$$

10 February 2011

HCL/UCB LYON

32

The new estimator intuitively

- The composition of the “at risk” set $\{Y_i(t), i=1..N\}$ is more and more “biased” as time elapses, since those subjects with low values of $S_{pi}(t)$ tend to disappear first from the *risk set* and had also a greater chance to die from their disease.
- To make it unbiased each $Y_i(t)$ is increased in such a way that, the smaller the probability of remaining at risk at time t ($S_{pi}(t)$), the larger the increase ($*1/S_{pi}(t)$).
- Obviously we do the same with the step of the counting process at time t

10 February 2011

HCL/UCB LYON

33

Systematic evaluation of the estimators (Danieli *et al.*)

- A simulation study was conducted in Lyon to assess the performance of the estimators.
- We report below the result of two scenarios:
 - The net survival does not depend on the demographic variable (homogeneous)
 - The net survival depend only on age through a proportional hazard model with a RR of 1.05 per year of age.
- The age distribution is similar to the observed incidence distribution of most adult cancers
- The sample size is 500

10 February 2011

HCL/UCB LYON

34

The estimates in the homogeneous situation

Net survival does not depend on age					
time net Survival probability (%)	Bias (Mean estimate-theoretical net survival)				
	ED I	ED II	Mult. model	Pohar	specific survival
5 29.60	-0.18	-0.17	0.26	-0.13	-0.16
10 14.87	-0.05	-0.13	0.04	0.13	-0.13

With a coverage rate very close to 95% for all of them

⇒ *All estimators estimate net survival*

10 February 2011

HCL/UCB LYON

35

The estimates in the age-dependent situation

Net survival depends on age					
time net Survival probability (%)	Bias (Mean estimate-theoretical net survival)				
	ED I	ED II	Mult. model	Pohar	specific survival
5 32.44	3.31	1.21	0.23	-0.05	1.24
10 20.39	7.04	2.00	0.06	0.19	2.01

The coverage rate is very poor for ED I (19.8% at 10 years) and not very good for ED II and specific survival (88% at 10 years), while it is good for the two other estimators.

⇒ *All estimators except the modeling and Pohar estimator showed a positive bias*

10 February 2011

HCL/UCB LYON

36

The “real life” situation

- The French cancer registries will publish soon an extension of their survival study, conducted in collaboration with the biostatistic group in LYON
- It was decided to use the Pohar estimator for this updating. A comparison was made with the previous approach.
- We show below the result for the thyroid cancer, which is one site for which the bias is large as shown earlier.

10 February 2011

HCL/UCB LYON

37

Thyroid cancer, French registries

age group	Pohar estimate	5 year survival		
		multivariate model		
		age,sex,year	age only	no covariable
All age	0.95	0.94	0.95	0.97
15-45	0.99	1.00	1.00	0.99
45-55	0.99	0.99	0.99	0.99
55-65	0.96	0.96	0.96	0.96
65-75	0.87	0.87	0.88	0.88
75+	0.64	0.63	0.65	0.69
standardized	91.10	90.90	91.50	92.00

10 February 2011

HCL/UCB LYON

38

Thyroid cancer, French registries

age group	Pohar estimate	10 year survival		
		multivariate model		
		age,sex,year	age only	no covariable
All age	0.92	0.93	0.93	0.96
15-45	0.99	0.99	0.99	0.99
45-55	0.97	0.98	0.98	0.98
55-65	0.94	0.95	0.95	0.96
65-75	0.83	0.83	0.84	0.88
75+	0.49	0.56	0.59	0.69
standardized	87.50	89.00	89.90	91.80

⇒ Not only is the “all age” survival biased, but also the standardized survival

10 February 2011

HCL/UCB LYON

39

Conclusion

- None of the classical estimators estimate the *net survival*,
- BUT if corrected for their biases, EDERER II and CAUSE-SPECIFIC could do it.
- EDERER I & HAKULINEN estimate *relative survival*
 - Contrary to *net survival*, *relative survival* is influenced by population mortality and is therefore less adequate than *net survival* for international comparisons.
- In addition net survival of the group being the average of individual net survival the “problem” of lack of consistency of standardized rates does not exist any more.

10 February 2011

HCL/UCB LYON

40

Acknowledgment

- The work presented here is a collaborative work between Ljubljana and Lyon. The new estimator was constructed by Maja POHAR-PERME.
- The discussion benefited from the work in the group of biostatistics in Lyon, where Janez STARE was in sabbatical leave in 2010.
- The Lyon working group (Nadine Bossard, Aurelien Belot, Coraline Danieli, Laurent Remontet, Laurent Roche) was in charge of the comparison of the estimators in the framework of the "*MESURE Project*"