Statistical methods for survival data: accounting for cured patients

Eni Musta

VU General Mathematics Colloquium, October 2020

University of Amsterdam

Outline

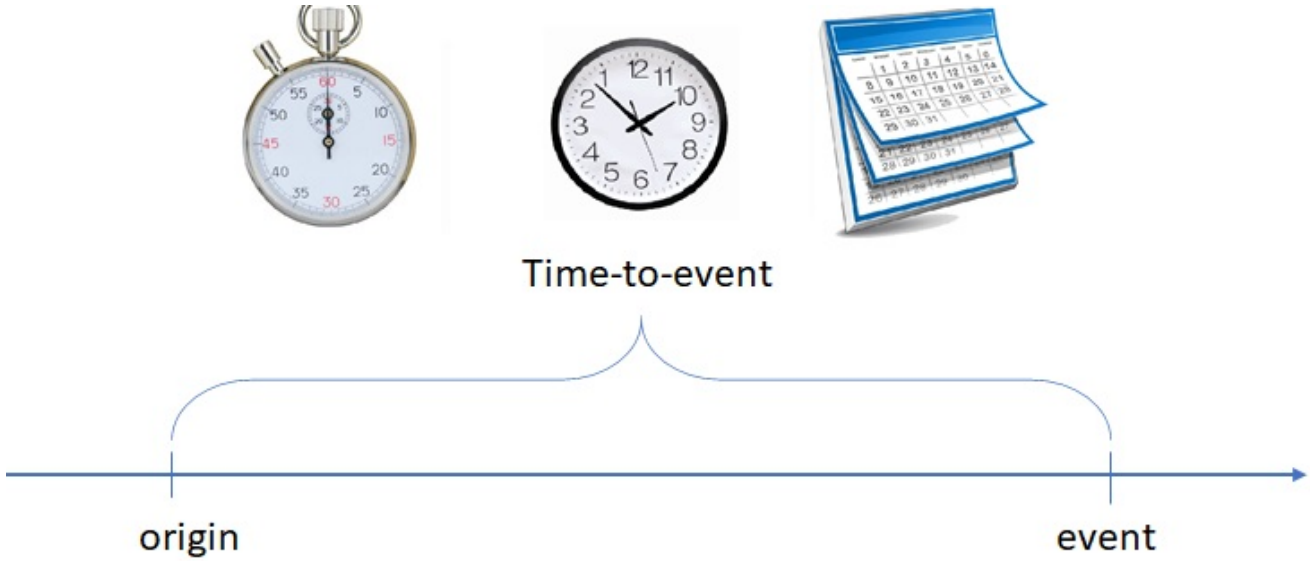
1. Basics of survival analysis
   - Time-to-event data
   - Estimation of the survival function
   - Cox proportional hazards model
2. Beyond conventional methods
3. Survival analysis with a cure fraction
   - Promotion time cure models
   - Mixture cure models

Part I

Basics of survival analysis

What is survival analysis?

A collection of statistical models and procedures for studying the time until an event of interest takes place.



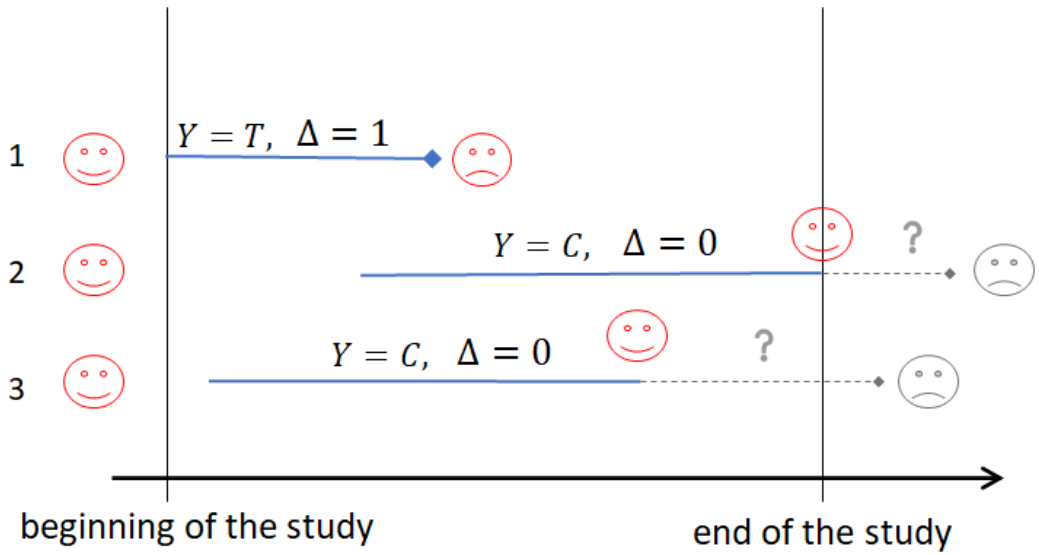
Other names

- Survival analysis (time until death or onset of disease)
- Reliability theory (time until equipment failure)
- Duration analysis (time until stock market crash)
- Event history analysis (time to first employment after graduation)

Right censoring

In practice, for some subjects the event of interest cannot be observed:

- the study ends before the event takes place
- the subjects drops out of the study
- another event happens before the event of interest



Survival time: T
Censoring time: C

Observed data:

Follow-up time: $Y = \min(T, C)$
Censoring indicator: $\Delta = \mathbb{1}_{\{T \leq C\}}$
Covariate vector: $Z \in \mathbb{R}^q$

*We assume T and C are independent

Common functions in survival analysis

- Survival function

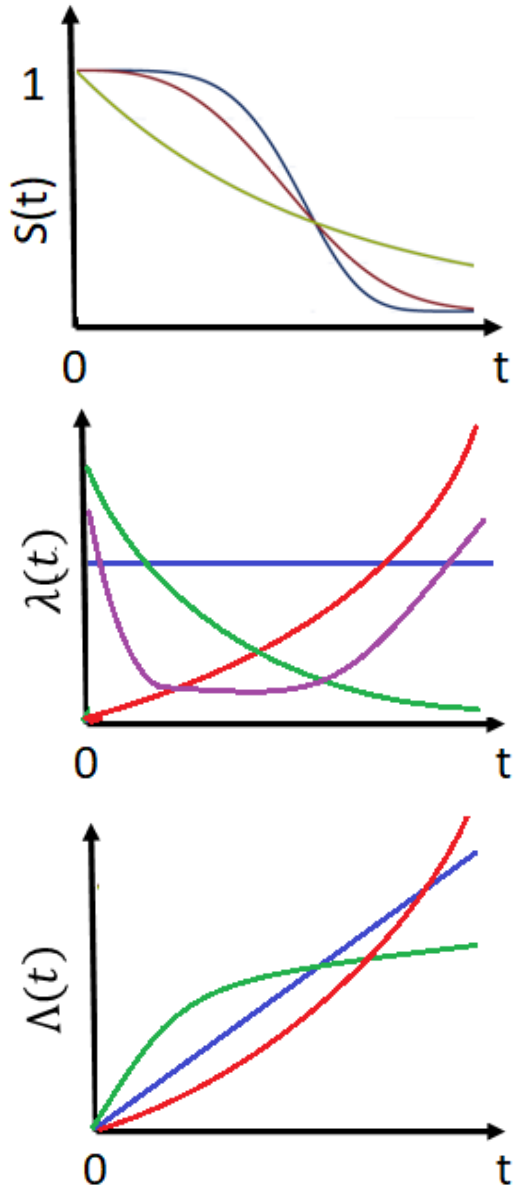
$$S(t) = \mathbb{P}(T > t) = 1 - F(t)$$

- Hazard function

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

- Cumulative hazard function

$$\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u = -\log S(t)$$



Nonparametric estimation

Given i.i.d. observations $(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)$ we want to estimate the survival function S of the survival time T.

If there was no censoring, we could use the empirical survival function

$$\hat{S}_n(t) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{T_i > t\}}}{n}$$

In the presence of censoring, we don't always know whether $T_i > t$!

Nonparametric estimation

Given i.i.d. observations $(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)$ we want to estimate the survival function S of the survival time T.

Likelihood for n i.i.d. observations

$$L = \prod_{i=1}^{n} \{f(Y_i)[1 - G(Y_i)]\}^{\Delta_i} \{g(Y_i)S(Y_i)\}^{1-\Delta_i}$$

where g and G denote the density and distribution function of C.

*The factors $[1 - G(Y_i)]^{\Delta_i}$, $g(Y_i)^{1-\Delta_i}$ are uninformative.

Aim: find S that maximizes

$$\sum_{i=1}^{n} [\Delta_i \log f(Y_i) + (1 - \Delta_i) \log S(Y_i)]$$

Kaplan-Meier estimator (1958)

Notations:

- ordered distinct observed times $Y_{(1)}, Y_{(2)}, \ldots, Y_{(r)}$ $(r \leq n)$
- number of events at time $Y_{(j)}$ is $d_{(j)}$, $j = 1, \ldots, r$
- size of risk set at time Y_j is $R_{(j)} = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i \geq Y_{(j)}\}}$

The maximizer of $\log L$ is

$$\hat{S}_n(t) = \prod_{j: Y_{(j)} \leq t} \left(1 - \frac{d_{(j)}}{R_{(j)}}\right),$$

$$\hat{S}_n(t) = \begin{cases} 1 & \text{for } t < Y_{(1)} \\ 1 - \frac{d_{(1)}}{R_{(1)}} & \text{for } Y_{(1)} \leq t < Y_{(2)} \\ \left(1 - \frac{d_{(1)}}{R_{(1)}}\right)\left(1 - \frac{d_{(2)}}{R_{(2)}}\right) & \text{for } Y_{(2)} \leq t < Y_{(3)} \\ \ldots \end{cases}$$

$\hat{S}_n$ is step function with jumps at the event times

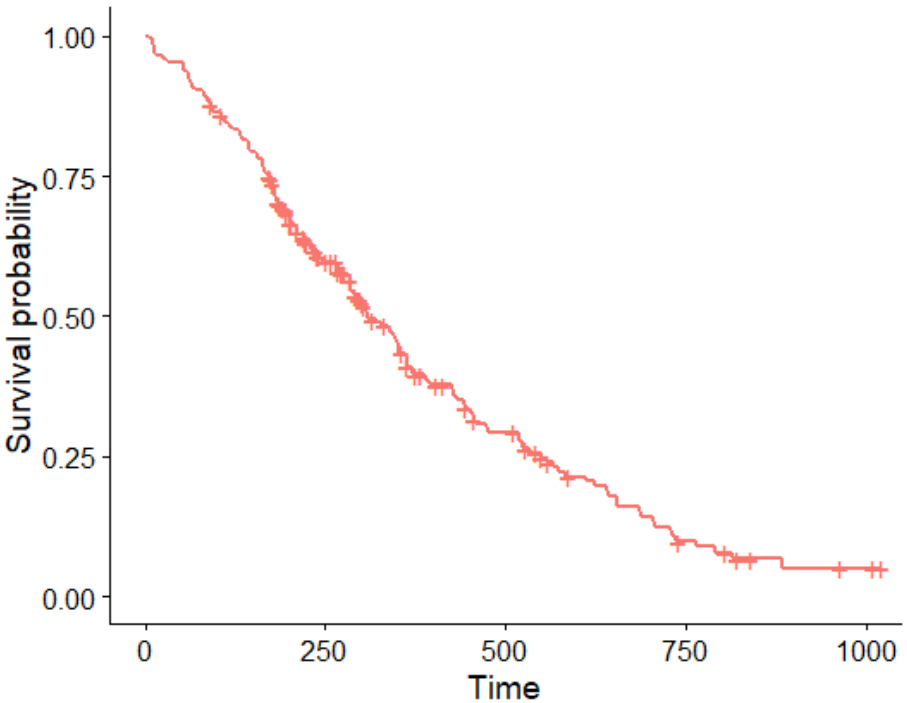
Example: NCCTG Lung Cancer Data

Survival in patients with advanced lung cancer from the North Central Cancer Treatment Group.

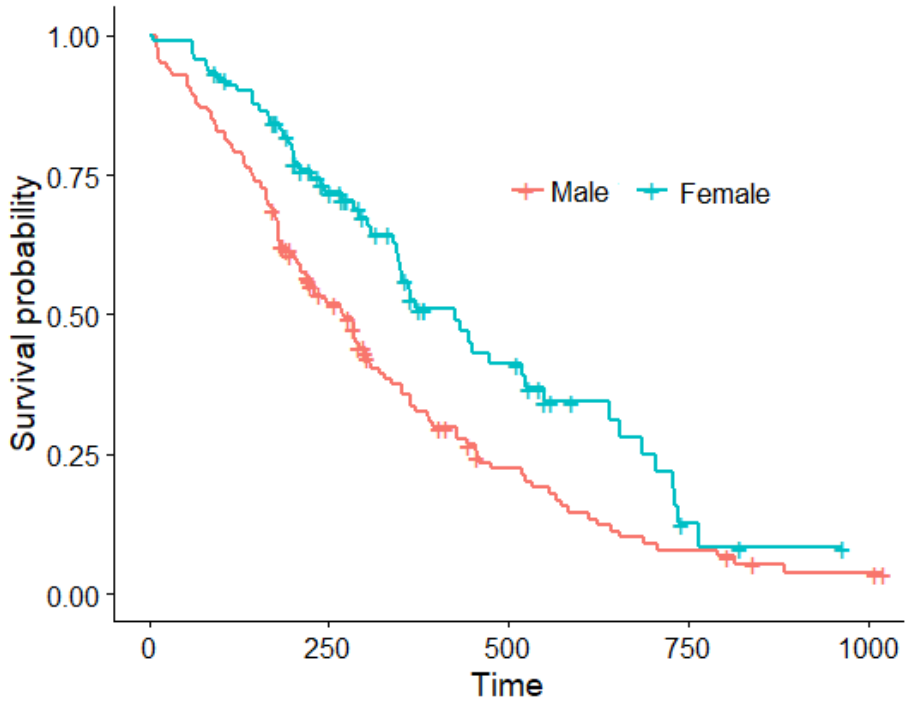
	inst	time	status	age	sex
1	3	306	1	74	1
2	3	455	1	68	1
3	3	1010	0	56	1
4	5	210	1	57	2
5	1	883	1	60	1

- 228 patients
- time = survival time in days (event: death from lung cancer)
- status = censoring status
- age = age in years
- gender: 1=male, 2=female
- inst = institution code

Kaplan-Meier estimator



KM estimator for the whole population



KM estimators based on gender

Cox proportional hazards model (Cox, 1972)

- The most popular model that incorporates covariate information
- The conditional hazard rate is given by

$$\lambda(t|z) = \lambda_0(t)\, e^{\beta_0' z}.$$

where

  - the baseline hazard λ_0 describes the variation over time
  - the vector of regression coefficients β_0 describes the effect of the covariates

- Proportional hazards assumption

$$HR(z, \tilde{z}) = \frac{\lambda(t|z)}{\lambda(t|\tilde{z})} = \frac{\lambda_0(t)\, e^{\beta_0' z}}{\lambda_0(t)\, e^{\beta_0' \tilde{z}}} = e^{\beta_0'(z-\tilde{z})}$$

The hazard ratio remains constant over time.

Why is it so popular?

- Easy interpretation:

$$HR(z, \tilde{z}) = e^{\beta_{0,1}} \qquad \text{if} \qquad z = \tilde{z} + (1, 0, \ldots, 0)$$

- β_0 can be estimated independently of λ_0 by maximizing the partial likelihood function

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{e^{\beta' Z_i}}{\sum_{Y_j \geq Y_i} e^{\beta' Z_j}} \right]^{\Delta_i}$$

- $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Sigma)$ and it is easy to estimate the asymptotic variance-covariance matrix.

Breslow estimator (Breslow and Crowley, 1974)

$$\Lambda_n(t) = \sum_{i=1}^{n} \frac{\mathbb{1}_{\{T_i \le t\}}}{\sum_{j=1}^{n} e^{\hat{\beta}_n' Z_j} \mathbb{1}_{\{T_j \ge T_i\}}}.$$

It is the maximizer of the Cox log-likelihood

$$l(\beta, \Lambda) = \sum_{i=1}^{n} \Delta_i \log \lambda(Y_i) + \Delta_i \beta' Z_i - e^{\beta' Z_i} \Lambda(Y_i)$$

for fixed $\beta = \hat{\beta}_n$.

Example: NCCTG Lung Cancer Data

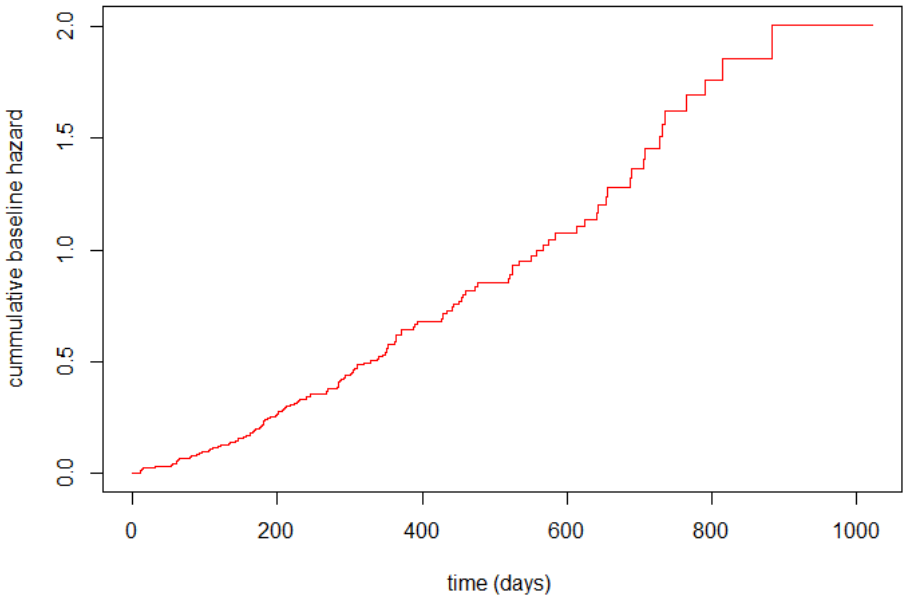
```
> coxph(Surv(time, status)~sex+age, data=Data)
Call:
coxph(formula = Surv(time, status) ~ sex + age, data = Data)

          coef exp(coef)  se(coef)      z Pr(>|z|)
sex -0.513219  0.598566  0.167458 -3.065  0.00218 **
age  0.017045  1.017191  0.009223  1.848  0.06459 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

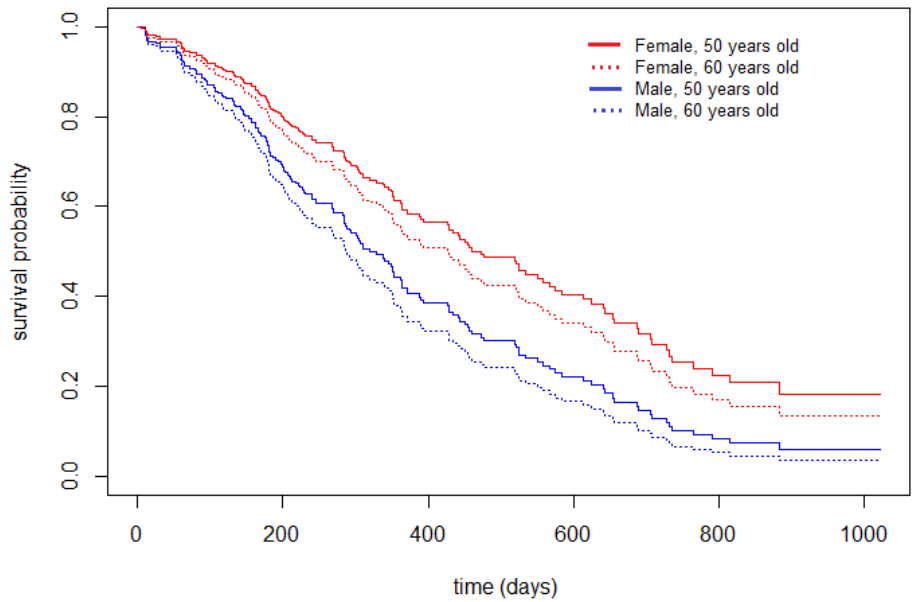
$$\exp(\text{coef}) = \text{HR} \begin{cases} = 1 & \text{no effect} \\ > 1 & \text{increase in hazard} \\ < 1 & \text{reduction in hazard} \end{cases}$$

Going from male to female results in nearly 40% reduction in hazard.

One year increase in age results in nearly 2% increase in hazard.

Example: NCCTG Lung Cancer Data





Estimator of survival probability for a given patient

$$\hat{S}_n(t|z) = \exp\left\{-\hat{\Lambda}_n(t)\exp\left(\hat{\beta}_1 * \text{Sex} + \hat{\beta}_2 * \text{Age}\right)\right\}$$

Part II

Beyond standard methods

What if the PH assumption is violated?

Some alternatives to the Cox model

- Cox model with time-varying effects

$$S(t|z) = \exp\left\{-\Lambda_0(t)\exp(\beta_t' z)\right\}$$

  The effect of certain variables might get stronger or weaker with time.

- accelerated failure time model

$$S(t|z) = \exp\left\{-\Lambda_0\left(t\exp(\beta' z)\right)\right\}$$

  If $\exp(\beta' z) > 1$ things happen faster in the life history of an individual.

Some other non-standard settings

- **Noparametric estimation under shape constraints**: the hazard rate after a successful medical treatment is expected to be decreasing.
- **Mismeasured covariates**: some variables such as the systolic blood pressure, tumor size, dietary intake cannot be measured precisely.
- **Dependent censoring**: patients may withdraw from the study because their condition is deteriorating or improving
- **Competing risks**: many mutually exclusive events are of interest, such as death from different causes, and the occurrence of one of these will prevent any other event from ever happening.

Part III

Cure rate models

The concept of 'cure'

In classical survival analysis we assume that all the subjects will experience the event of interest

$$\lim_{t \to \infty} S(t) = 0.$$

In some situations, such assumption is not realistic. Instead we have

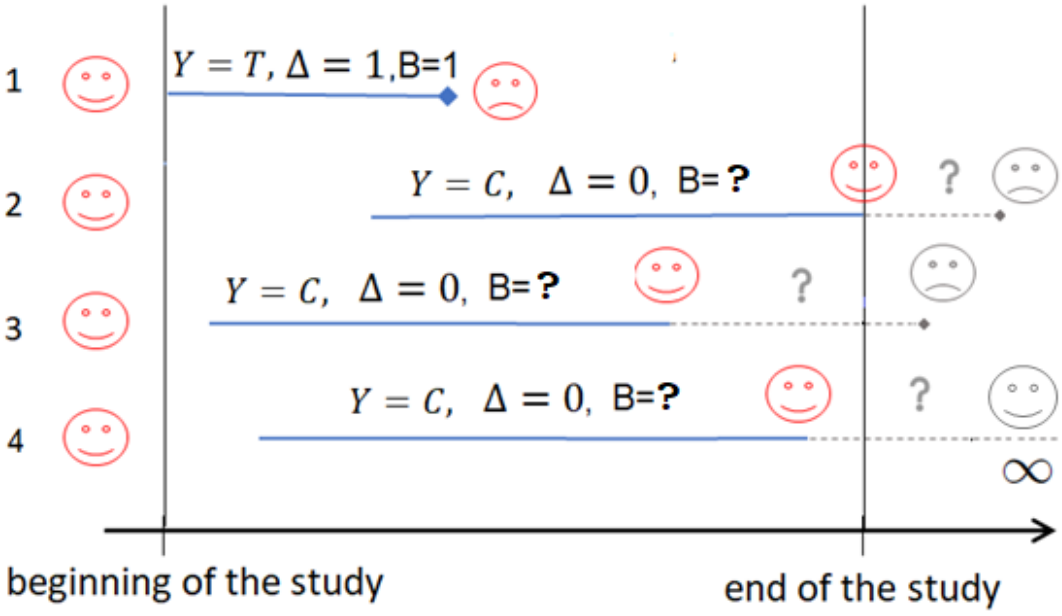
$$\lim_{t \to \infty} S(t) > 0.$$

- Curable diseases: some patients will never die from that disease
- Demography: time to a second child after a first one
- Finance: time until a bank or a business goes bankrupt
- Marketing: time until someone buys a new product

We refer to all the subjects that do not experience the event of interest as being cured and to the others as susceptibles.

Who is cured?

As a result of censoring, cured individuals cannot be distinguished from the uncured ones.



Cure status: $B \in \{0, 1\}$
Survival time:

$$T = BT^* + (1 - B)\infty$$

Censoring time: $C < \infty$

Observed data:

Follow-up time: $Y = \min(T, C)$
Censoring indicator: $\Delta = \mathbb{1}_{\{T \leq C\}}$
Covariate vector: $Z \in \mathbb{R}^q$

When can we use cure models?

- Medical evidence of cure
- Inspection of the Kaplan Meier survival curve (it levels up at some value larger than zero)
- Sufficient follow-up: long plateau with heavy censoring

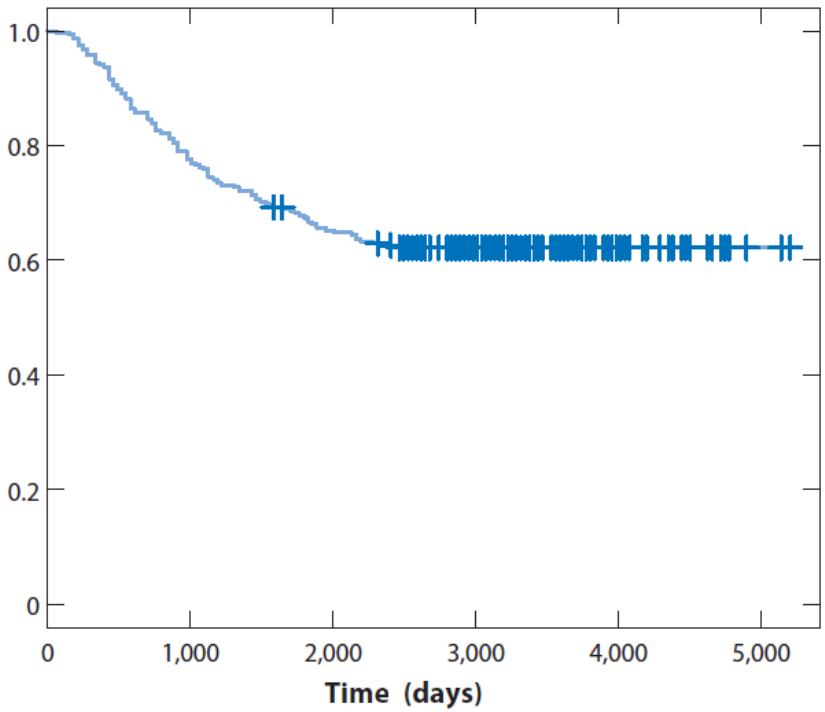


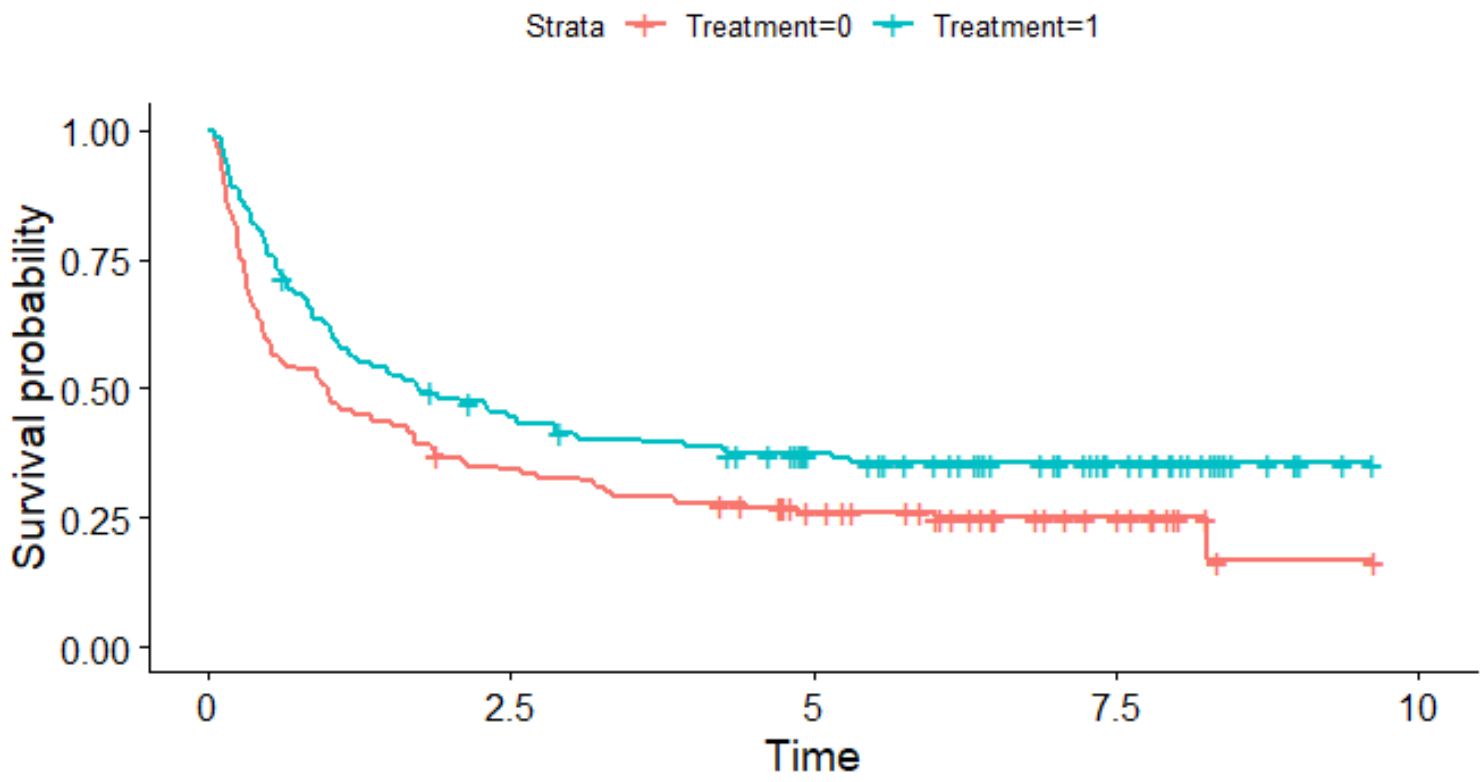
Figure from Amico & Van Keilegom (2018)

Kaplan-Meier estimator of the relapse free survival for breast cancer patients (Wang, 2005)

286 patients, 179 are censored out of which 88% are in the plateau

Eastern Cooperative Oncology Group (ECOG) Data

- phase III clinical trial to evaluate the high dose interferon alpha-2b (IFN) regimen against the placebo as the postoperative adjuvant therapy for melanoma patients
- 284 observations, 30% censored, 13 observations in the plateau
- The response variable is relapse free survival in years



The concept of cure in medical applications

- Is not always well defined, does not consider long-term adverse effects of cancer treatment
- It is reasonable to claim 'cure' for those who survive beyond a particular time point (long term survivors) when there is a negligible risk of the event of interest
- The patient survives until his mortality risk reaches the same level as the general population mortality risk (relative survival).

Why do we need cure models?

- The heterogeneity of the population leads to a violation of the PH assumption. Standard analysis might give misleading results.
- Cure models provide additional information: distinguish between a curative or a life-prolonging effect.
- Assessing the chances of being 'cured' is of interest for more informed decision making (for planning further treatment).

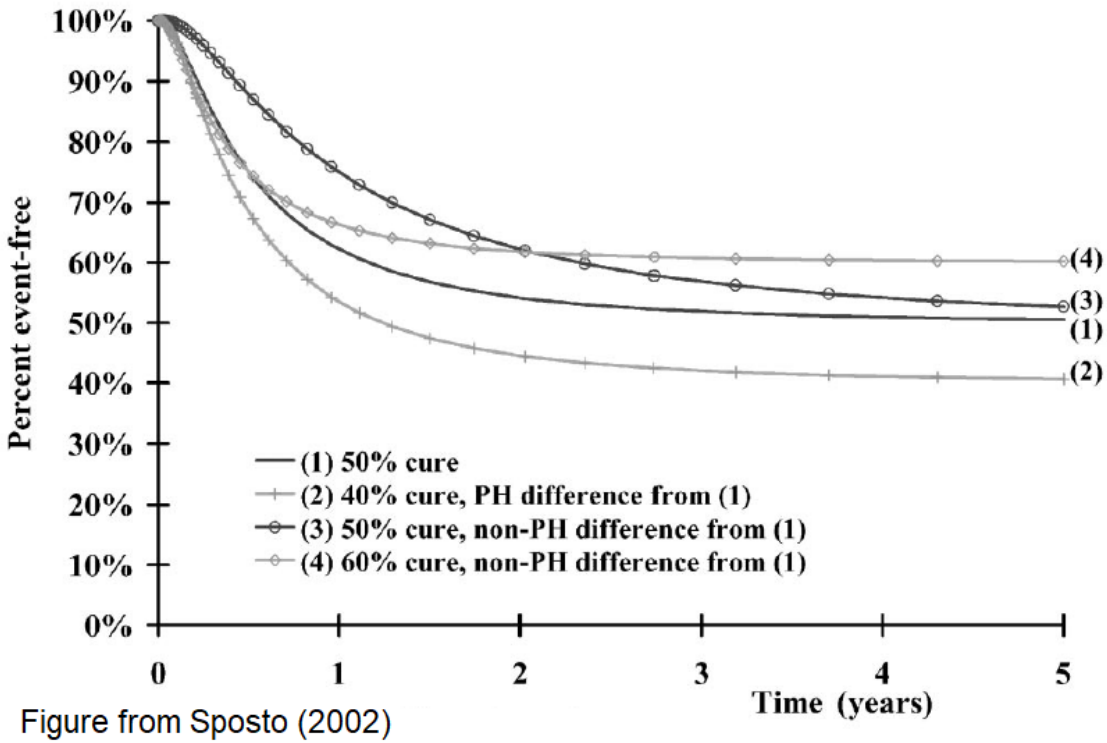


Figure from Sposto (2002)

Family I: promotion time cure models (Yakovlev et al. 1996)

Known also as bounded cumulative hazard model or PH cure model.

$$S(t|z) = \exp\{-\theta(z)\Lambda_0(t)\}$$

where

- The baseline cumulative hazard is bounded $\lim_{t\to\infty} \Lambda_0(t) = 1$
- Usually $\theta(z) = \exp(\beta' z)$ and the vector of covariates Z contains an intercept

The cure rate is $\exp(-\theta(x))$ and the PH assumption is still satisfied.

Biological interpretation: modelling cancer relapse

- After a first cancer, $N \geq 0$ carcinogenic cells can stay active in the organism
- It takes a certain time $\tilde{T}_k$, $k = 1, \ldots, N$ for each such cell to become an active tumor
- For individuals for whom $N \geq 1$ (for uncured observations), the survival time T is $\min\{\tilde{T}_1, \ldots, \tilde{T}_N\}$.
- For cured individuals $N = 0$ and hence $T = \infty$.
- If $N \sim \text{Poisson}(\theta)$ and $\tilde{T}_1, \ldots, \tilde{T}_n$ are i.i.d. with distribution F, then

$$\mathbb{P}(T > t) = \exp\{-\theta F(t)\}$$

Family II: mixture cure models (Boag, 1949)

The survival of a general individual of the population is given by

$$S(t|x,z) = \mathbb{P}(T > t|X = x, Z = z) = 1 - \pi(x) + \pi(x)S_u(t|z).$$

where

- $X \in \mathbb{R}^p$, $Z \in \mathbb{R}^q$ are two covariate vectors containing the variables affecting the probability of being cured and the survival of the susceptibles
- $\pi(x) = \mathbb{P}(B = 1|X = x)$ is the probability of being susceptible,
- $S_u(t|z) = \mathbb{P}(T > t|B = 1, Z = z)$ is the survival function of the susceptibles.

We have

$$\lim_{t\to\infty} S_u(t|z) = 0 \quad \text{and} \quad \lim_{t\to\infty} S(t|x,z) = 1 - \pi(x) > 0$$

Most common mixture cure model

- Parametric logistic model for the incidence

$$\pi(x) = \mathbb{P}(B = 1|x) = \phi(\gamma_0, x) = \frac{e^{\gamma_0' x}}{1 + e^{\gamma_0' x}}$$

  where X contains an intercept.

- Semi-parametric Cox PH model for the latency

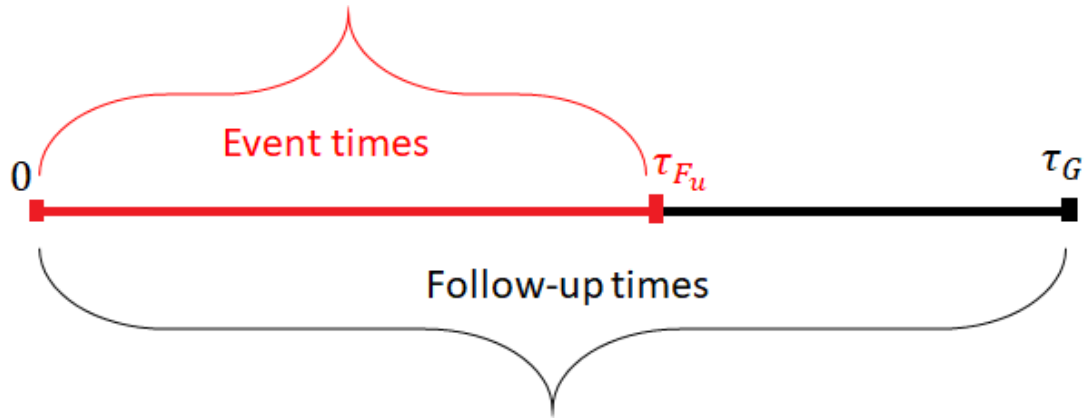
$$S_u(t|z) = \mathbb{P}(T > t|z, B = 1) = \exp\left\{-\Lambda_0(t)\exp(\beta_0' z)\right\}$$

At the level of the whole population the PH assumption is not satisfied!

Semi-parametric estimation

Identifiability condition:

$$\tau_{F_u} < \tau_G$$



- Partial likelihood method does not work
- The most common approach of maximizing the likelihood is via the EM algorithm (`smcure` package)
- Zero-tail constraint
- Standard errors of the estimators are computed using bootstrap procedures

Semi-parametric estimation

Observed likelihood

$$L(\gamma, \beta, \Lambda) = \prod_{i=1}^{n} \left\{ \phi(\gamma, X_i) \lambda(Y_i) \exp(\beta' Z_i) \exp\left[-\Lambda(Y_i) \exp(\beta' Z_i) \right] \right\}^{\Delta_i}$$
$$\left\{ 1 - \phi(\gamma, X_i) + \phi(\gamma, X_i) \exp\left[-\Lambda(Y_i) \exp(\beta' Z_i) \right] \right\}^{1-\Delta_i}$$

Full likelihood

$$L(\gamma, \beta, \Lambda) = \prod_{i=1}^{n} \left\{ 1 - \phi(\gamma, X_i) \right\}^{1-B_i} \phi(\gamma, X_i)^{B_i}$$
$$\times \left\{ \lambda(Y_i) \exp(\beta' Z_i) \right\}^{B_i \Delta_i} \left\{ \exp\left[-\Lambda(Y_i) \exp(\beta' Z_i) \right] \right\}^{B_i}$$

EM algorithm

Iterative procedure

$$\hat{\beta}_n^{(m+1)} = \underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} \left[\frac{\exp(\beta' Z_i)}{\sum_{Y_j \geq Y_i} W_j^m \exp(\beta' Z_j)} \right]^{\Delta_i}$$

$$\hat{\Lambda}_n^{(m+1)}(t) = \sum_{i=1}^{n} \frac{\mathbb{1}_{\{Y_i \leq t\}} \Delta_i}{\sum_{Y_j \geq Y_i} W_j^m \exp(\beta' Z_j)}$$

$$\hat{\gamma}_n^{(m+1)} = \underset{\gamma}{\operatorname{argmax}} \prod_{i=1}^{n} \phi(\gamma, X_i)^{W_i^m} \left[1 - \phi(\gamma, X_i)\right]^{1 - W_i^m}$$

where

$$W_i^m = \mathbb{E}\left[B_i \mid Y_i, \Delta_i, X_i, Z_i, \hat{\gamma}_n^m, \hat{\beta}_n^m, \hat{\Lambda}_n^m \right]$$

In particular $W_i^m = 1$ if $\Delta_i = 1$.

Estimation based on presmoothing

1. Estimate the cure probability $1 - \pi(x)$ non-parametrically by $\hat{S}_n(Y_{(r)}|x)$ where

$$1 - \hat{\pi}(x) = \hat{S}_n(t|x) = \prod_{t \in \mathbb{R}} \left(1 - \frac{\hat{H}_1(dt|x)}{\hat{H}([t, \infty)|x)} \right)$$

and

$$\hat{H}_k([t, \infty)|x) = \sum_{i=1}^{n} \frac{K_b(x - X_i)}{\sum_{j=1}^{n} K_b(x - X_j)} \mathbb{1}_{\{Y_i \leq t, \Delta_i = k\}}, \qquad k = 0, 1$$

2. 'Project' this nonparametric estimator to the class of logistic functions

$$\hat{\gamma}_n = \underset{\gamma}{\text{argmax}} \prod_{i=1}^{n} \phi(\gamma, X)^{\hat{\pi}(x)} \left[1 - \phi(\gamma, x)\right]^{1 - \hat{\pi}(x)}$$

3. Estimate β and Λ using the EM algorithm

Results breast cancer data

Incidence	Estimate	Standard error	$P(>\|Z\|)$
Intercept	1.1110	0.8850	0.2093
Age	−0.0382	0.0173	0.0270
ER+ versus ER−	0.1824	0.2704	0.4999
Tumor size	−0.0784	0.2054	0.7029
Menopausal (post- versus pre-)	0.7721	0.4445	0.0824
Latency	**Estimate**	**Standard error**	$P(>\|Z\|)$
Age	−0.0127	0.0179	0.4802
ER+ versus ER−	−1.0365	0.2317	<0.0001
Tumor size	0.5203	0.2184	0.0172
Menopausal (post- versus pre-)	0.0778	0.3970	0.8446

Table from Amico & Van Keilegom (2018)

Results ECOG melanoma data

		smcure package			Presmoothing		
		Estimates	SE	p-values	Estimates	SE	p-values
incidence	Intercept	1.3649	0.3457	$8 \cdot 10^{-5}$	1.6697	0.3415	10^{-6}
	Age	0.0203	0.0159	0.2029	0.0220	0.0104	0.0344
	Gender	−0.0869	0.3347	0.7949	−0.3039	0.3448	0.3493
	Treatment	−0.5884	0.3706	0.1123	−0.9345	0.3603	0.0095
latency	Age	−0.0077	0.0069	0.2663	−0.0079	0.0060	0.1861
	Gender	0.0994	0.1932	0.6067	0.1240	0.1653	0.4534
	Treatment	−0.1535	0.1715	0.3707	−0.0947	0.1692	0.5756

Covariates	Cox PH model Estimates	SE	p-value
Age	0.0049	0.0053	0.3554
Gender	-0.0180	0.1469	0.9023
Treatment	-0.3598	0.1437	0.0123

Results ECOG melanoma data

Cox model

- Treatment reduces the risk by 30% (HR= 0.7)

Mixture cure model (`smcure`)

- For patients that are not cured, treatment reduces the risk by 14% for the patients (HR= 0.86)
- Treatment increases the probability of being cured for a male patient with age equal to the mean of the sample from 0.2 to 0.31.

Mixture cure model (presmoothing)

- For patients that are not cured, treatment reduces the risk by 9% for the patients (HR= 0.91)
- Treatment increases the probability of being cured for a male patient with age equal to the mean of the sample from 0.16 to 0.32.

References

- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. JASA 53, 457–481.
- Cox, D. R. (1972). Regression models and life-tables. JRSS B 34, 187–220
- Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. The Ann. of Stat., 437-453.
- Kalbfleisch, J. D., and Prentice, R. L. (2011). The statistical analysis of failure time data (Vol. 360). John Wiley & Sons.

- Boag, J. W. (1949), 'Maximum likelihood estimates of the proportion of patients cured by cancer therapy', JRSS 11, 15–53.
- Yakovlev, A. Y., Tsodikov, A. D. and Asselain, B. (1996), Stochastic Models of Tumor Latency and Their Biostatistical Applications, Vol. 1 of Mathematical Biology and Medicine, World Scientific, Singapore.
- Maller, R. A., and Zhou, X. (1996). Survival analysis with long-term survivors. John Wiley & Sons.

References

- Cai, C., Zou, Y., Peng, Y., and Zhang, J. (2012). smcure: An R-Package for estimating semiparametric mixture cure models. Computer methods and programs in biomedicine, 108(3), 1255-1260.
- Peng, Y. and Taylor, J.M.G. (2014) Cure models. Book chapter in Handbook of Survival Analysis, pages 113-134.
- Amico, M. and Van Keilegom, I. (2018) Cure models in survival analysis. Annual Review of Statistics and Its Application, 5(1).
- Legrand, C. and Bertrand, A. (2019) Cure models in cancer clinical trials. Book chapter in Textbook of Clinical Trials in Oncology.

Acknowledgments

Cure rate models:

Ingrid Van Keilegom (KU Leuven, Belgium)

Valentin Patilea (ENSAI, France)

Estimation of a monotone hazard in the Cox model

Hendrik P. Lopuhaä (TU Delft, Netherlands)

Cécile Durot (Université Paris Nanterre, France)

Thank you

for your attention

e.musta@uva.nl

UNIVERSITY OF AMSTERDAM

Your Expected Lifetime left : **31 YEARS***

* *Expectation can have an error of +/- 30 years*