

A Leonard Held
production

We need a Kaplan-Meier plot!
Can you do this?

Yes!

But, -- hmmm -- in your case,
..., I dont know, not yet!

We need it yesterday!

**Ok, let's call it Kaplan-Meier!
... and check details ... later**

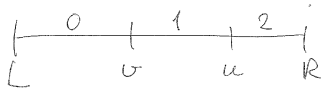
Interval censoring and the product limit method

*Daniel Commenges, Halina Frydman, Thomas Gerds,
Randi Grøn, Pierre Joly, Niels Keiding*

- ▶ Nonparametric estimation of the survival function
- ▶ A new one-step estimator
- ▶ Fillings in primary teeth
- ▶ Extension: competing risks

$$P_{02}(L, R) = \int_L^R e^{-\varepsilon(u) - \varphi(u)} \varepsilon(u) du$$
~~$$+ \int_L^R \left(\int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) dv \right) \cdot e^{-\alpha(u) - \alpha(v)} du$$~~

For $\alpha = \varepsilon$



$$\int_L^R \int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) e^{-\alpha(u) - \alpha(v)} \alpha(u) dv du$$

For $\alpha = \varepsilon$

~~$$\int_L^R \int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) e^{-\varepsilon(u) - \varepsilon(v)} \varepsilon(u) dv du$$~~

$$= \int_L^R e^{-\alpha(u) - \alpha(u)} \alpha(u) du$$

4.2 Likelihood for coarsened observation

Let us first compute the likelihood on the atoms of \mathcal{O}' (here we must not omit the proportionality term because we shall have to combine the likelihood on two events). Consider first $D_{1l} \cap D_{2l}$. From Lemma 2 or Theorem 2 in Commenges and Gégout-Petit (2006) we have on this event:

$$\mathcal{L}_{\mathcal{O}'} = \frac{1}{(e^{-v_{l-1}} - e^{-v_l})^2} \int_{v_{l-1}}^{v_l} f_1(s_1, s_2) ds_1 ds_2.$$

On $E_1 \cap D_{2l}$ the likelihood is:

$$\mathcal{L}_{\mathcal{O}'} = \frac{e^{v_l}}{(e^{-v_{l-1}} - e^{-v_l})^2} \int_{v_{l-1}}^{v_l} f_2(C, s_2) ds_2.$$

Here $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ represent the likelihoods for non-coarsened observation for possible values of T_1 and T_2 given by the first and second arguments respectively. We have for the first equation

$$f_1(s_1, s_2) = 1_{\{s_1 < s_2\}} \alpha_{01}(s_1 - b) \alpha_{12}(s_2) \exp\left[-\int_b^{s_1} \alpha_{01}(t - b) dt - \int_0^{s_1} \alpha_{02}(t) dt - \int_{s_1}^{s_2} \alpha_{12}(t) dt\right],$$

and in second equation:

$$f_2(C, s_2) = \alpha_{02}(s_2) \exp\left[-\int_b^{s_2} \alpha_{01}(t - b) dt - \int_0^{s_2} \alpha_{02}(t) dt\right].$$

Now on $[D_{1l} \cup E_1] \cap D_{2l}$ the likelihood $\mathcal{L}_{\mathcal{O}}$ is given by formula (1). It is easy to find that $P_{0l} = \frac{(e^{-v_{l-1}} - e^{-v_l})^2}{(e^{-v_{l-1}} - e^{-v_l})^2 + e^{-v_l}(e^{-v_{l-1}} - e^{-v_l})}$ so that the likelihood $\mathcal{L}_{\mathcal{O}}$ is proportional to:

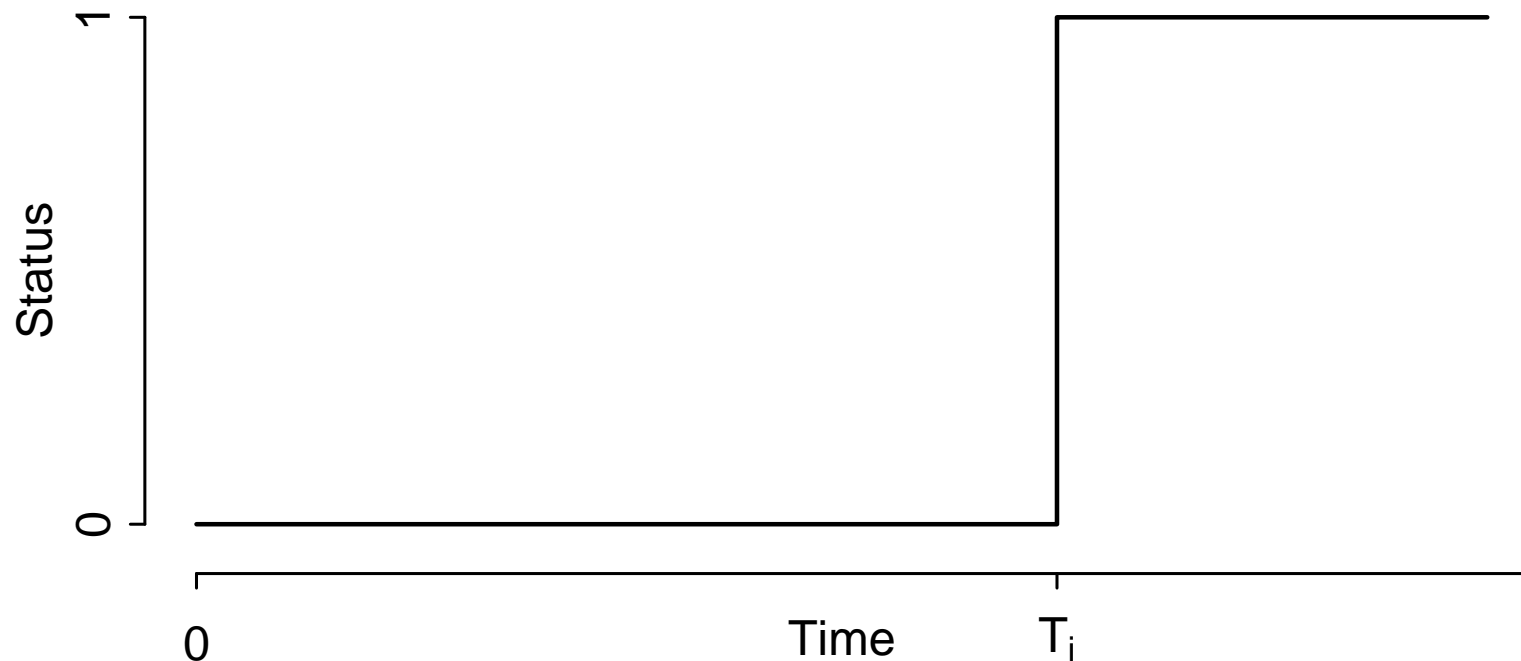
$$P_{\alpha}([D_{1l} \cup E_1] \cap D_{2l}) = \int_{v_{l-1}}^{v_l} f_1(s_1, s_2) ds_1 ds_2 + \int_{v_{l-1}}^{v_l} f_2(C, s_2) ds_2.$$

Note that in this observation scheme the smallest event observed has non-null probability: the observed sigma-field can be decomposed in atoms and the likelihood is proportional to the probability of the observed atom. We might have found a more direct way to compute this probability but it is not completely obvious anyway. It remains to compute the likelihood on the other atoms. One must also take into account the truncation condition: the sample is constituted of teeth which have not exfoliated before b so that we must divide the likelihood by $P_{\alpha}(T_2 \geq b) = \exp\left[-\int_0^b \alpha_{02}(t) dt\right]$

Uncensored observations

Given T_1, \dots, T_n – *iid* event times of n individuals – the aim is to estimate $S(t) = P(T_1 > t)$.

The process $N_i(t) = \mathcal{I}\{T_i \leq t\}$ represents T_i :

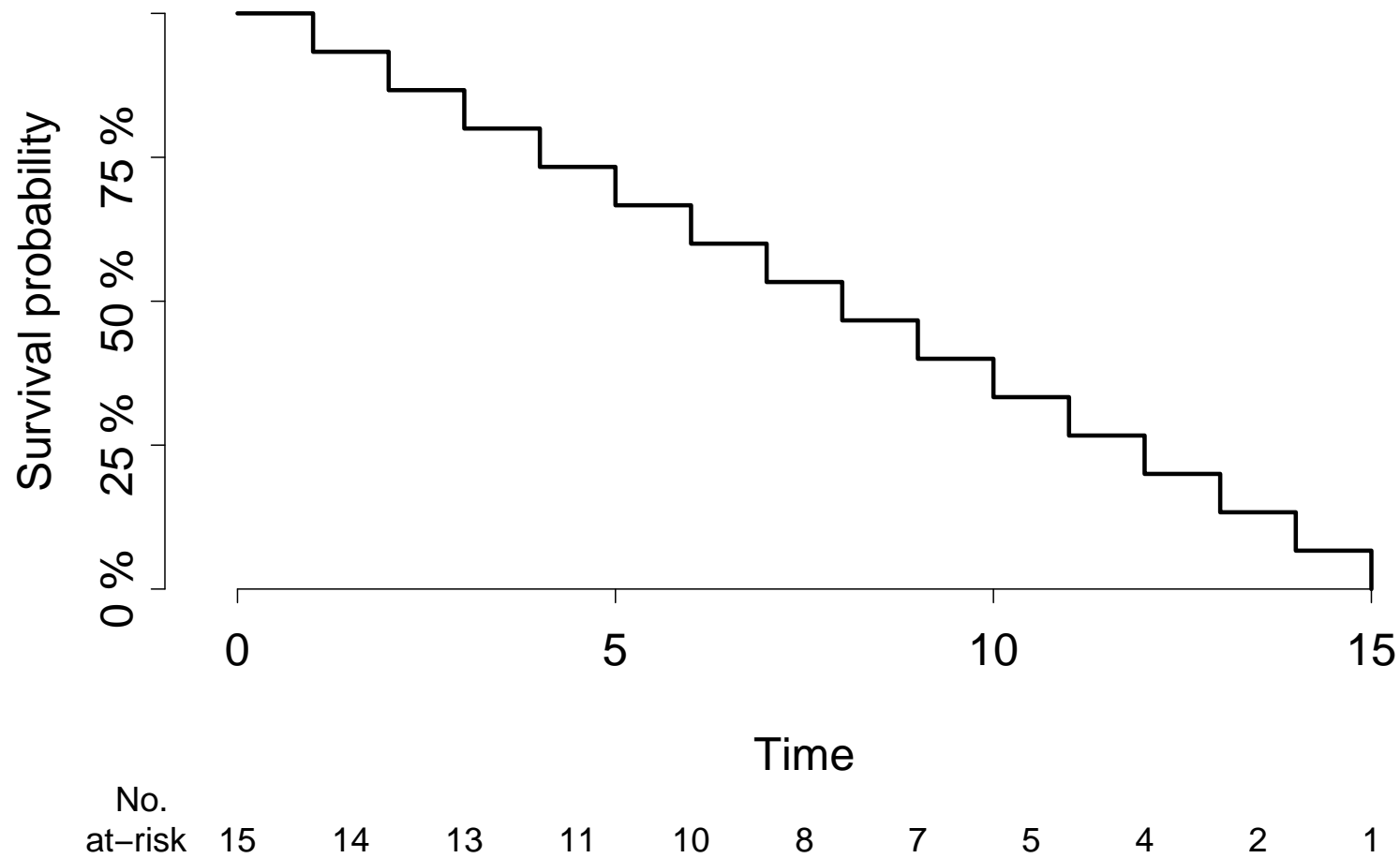


Uncensored observations

Each event time contributes a jump of

$$1/n$$

to the empirical survival distribution function:



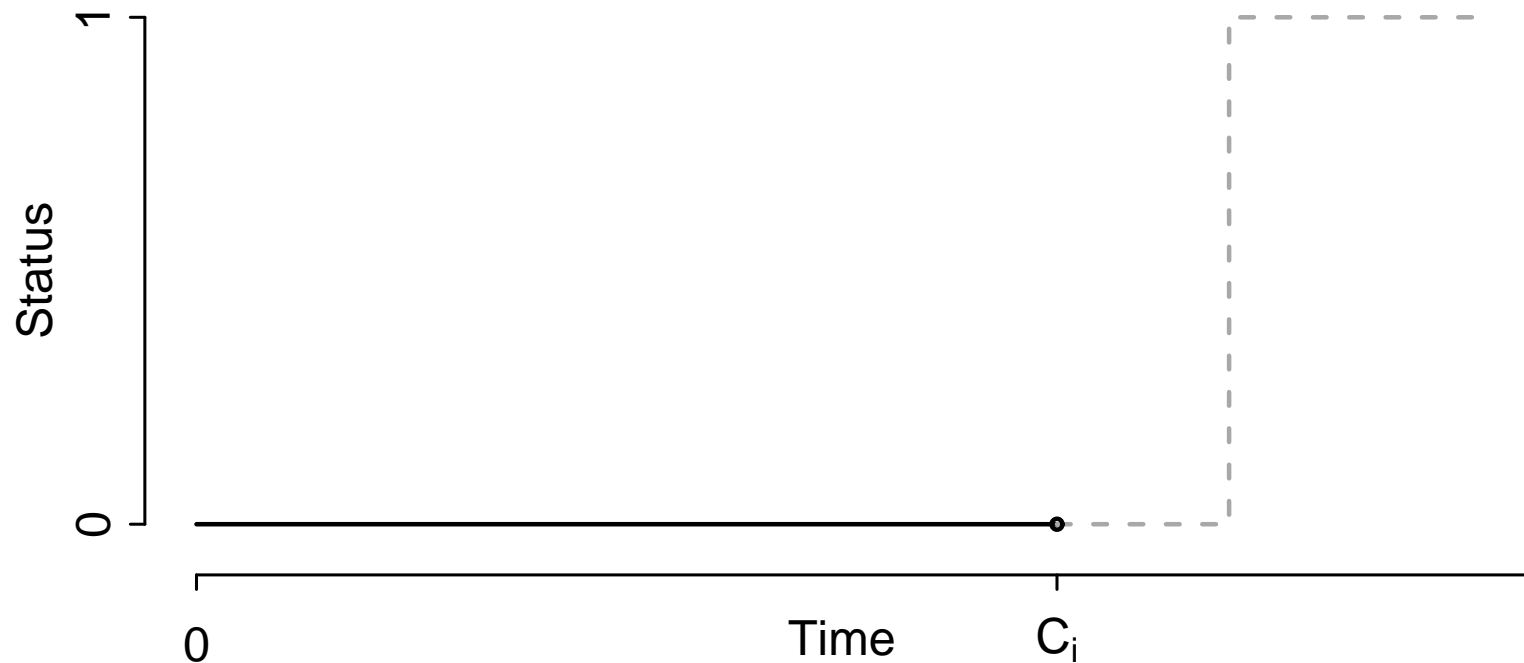
Right censored observations

With independent censoring times C_1, \dots, C_n the data are

$$\tilde{T}_i = \min(T_i, C_i)$$

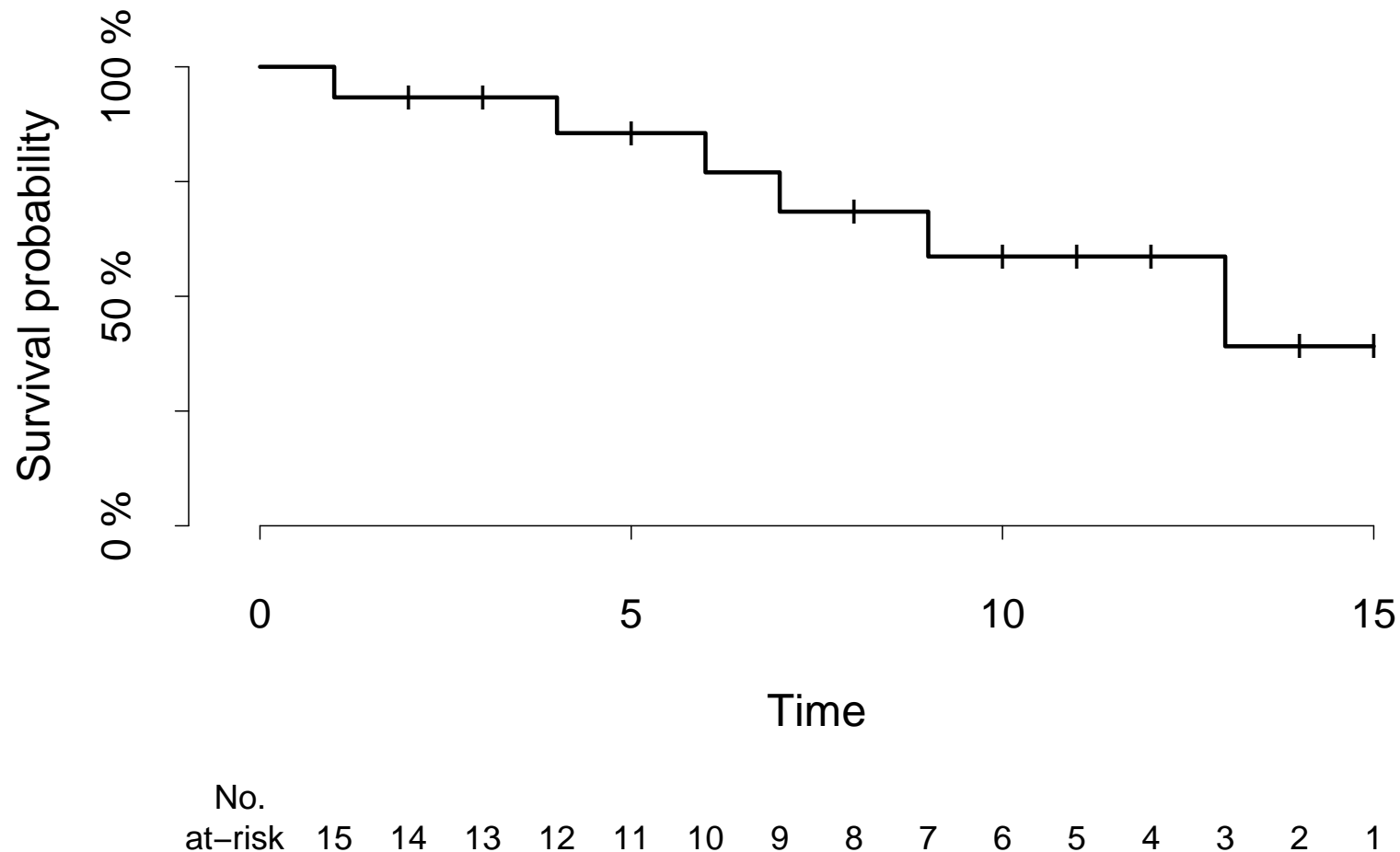
$$\Delta_i = \mathcal{I}\{T_i \leq C_i\}$$

An observation is right censored when $\Delta_i = 0$:



Right censored observations

The Kaplan-Meier estimate redistributes the mass $1/n$ of censoring times in equal pieces to the jumps at the **succeeding** observed event times



Right censored observations

With counting processes

$$N_i(dt) = \mathcal{I}\{\tilde{T}_i \in dt, \Delta_i = 1\}$$

$$Y_i(t) = \mathcal{I}\{\tilde{T}_i > t\}$$

$$N = \sum_i N_i$$

$$Y = \sum_i Y_i$$

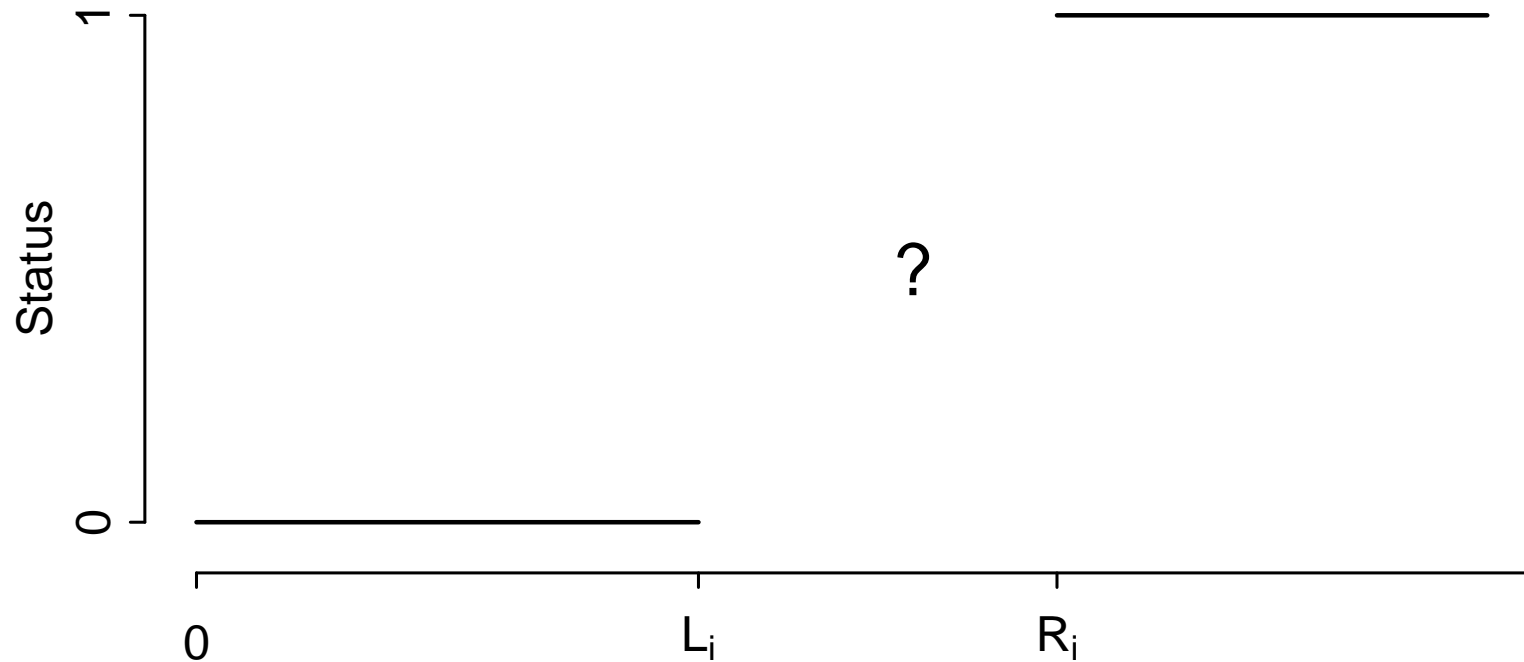
The Kaplan-Meier estimator is:

$$\hat{S}(t) = \prod_{s \leq t} \left(1 - \frac{N(ds)}{Y(s)} \right).$$

Interval censored observations

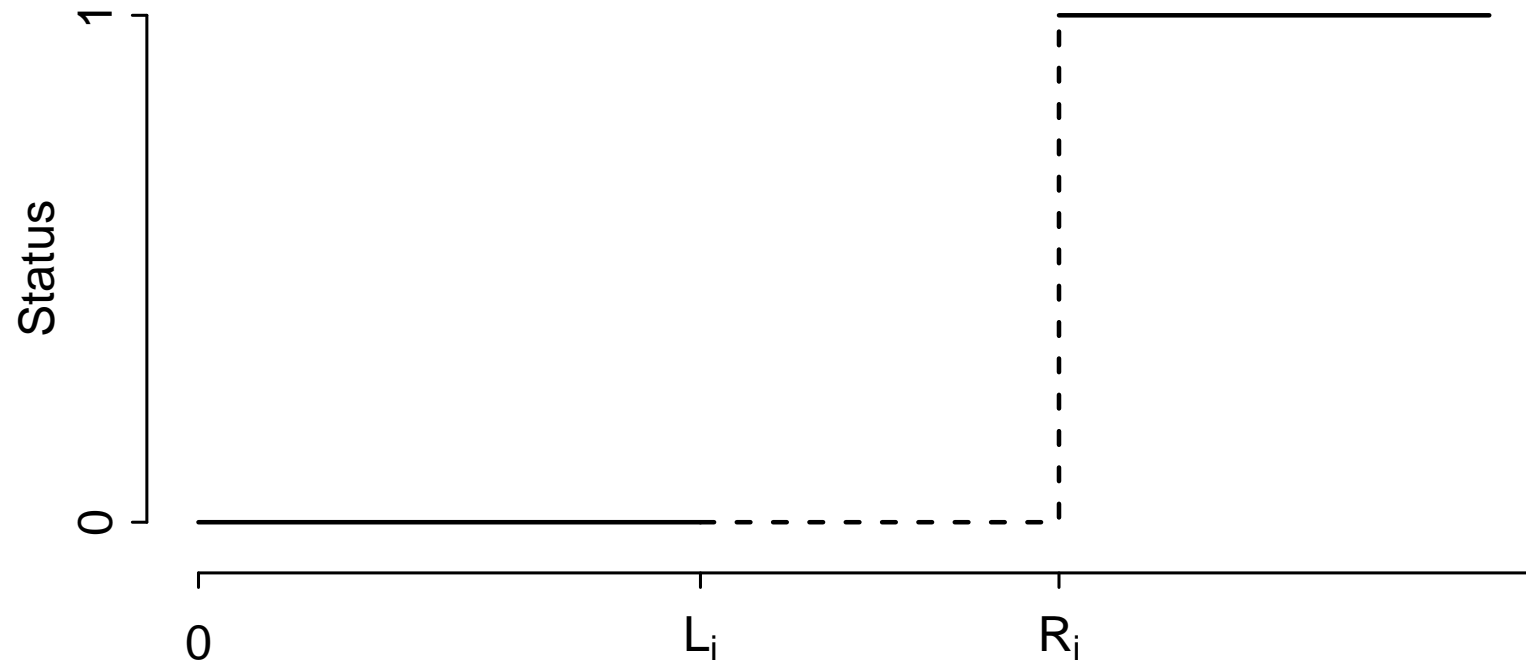
For random examination times L_i and R_i the data are $[L_i, R_i]$ such that

$$P(T_i \in [L_i, R_i]) = 1$$



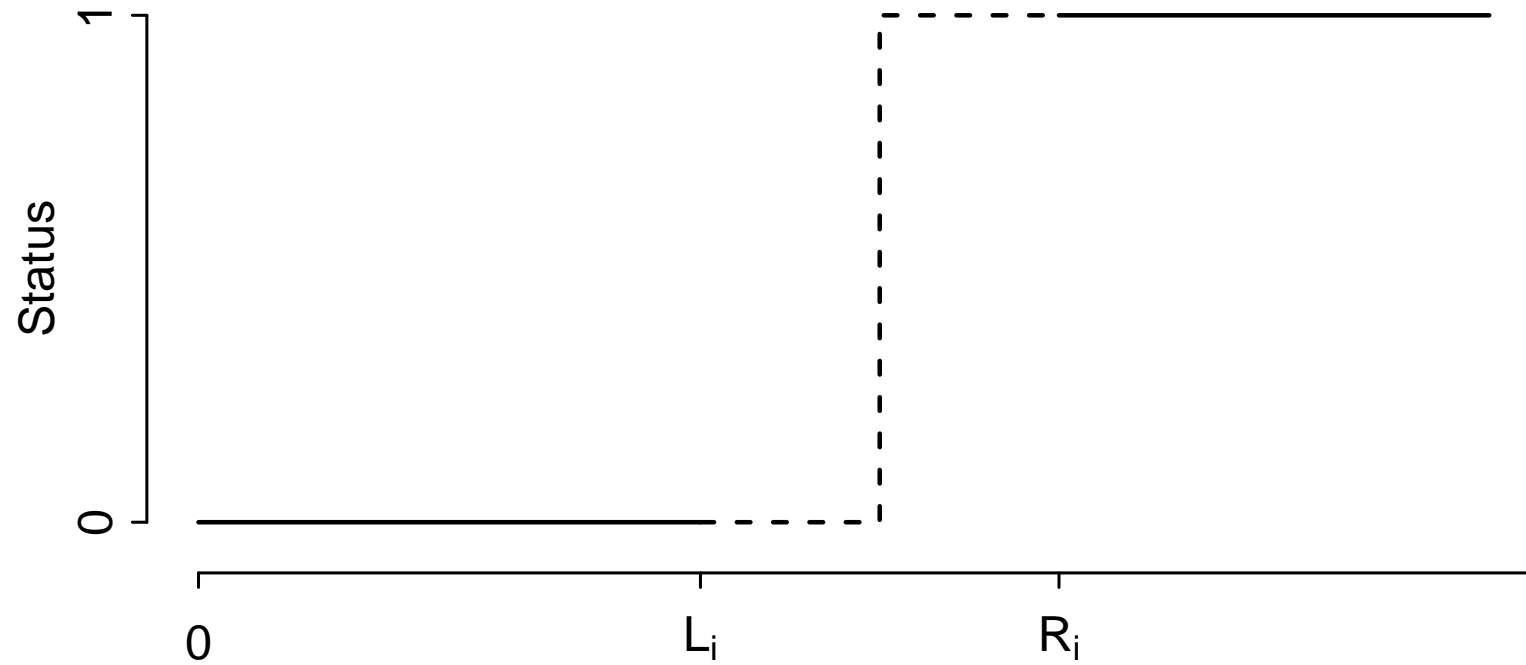
Interval censored observations

Ad-hoc approach 1



Interval censored observations

Ad-hoc approach 2



Interval censored observations

The nonparametric maximum likelihood estimator (NPMLE) is

$$S_n = \operatorname{argmax}_S \prod_{i=1}^n [S(L_i+) - S(R_i+)]$$

Constraint: S is a discrete subdistribution function which jumps only at the values of L_i and R_i for $i = 1, \dots, n$.

There is no explicit form for S_n but many algorithms:

- ▶ EM (Peto, 1973)
- ▶ Self-consistent (Turnbull, 1976)
- ▶ Iterative convex minorant (Groeneboom and Wellner, 1994)
- ▶ Hybrid (Wellner and Zhan, 1997; Dümbgen et al., 2006)

R-package (Icens) by Vandal & Gentleman

Cosmesis data

cosmesis

package:Icens

R Documentation

The time taken until cosmetic deterioration of breast cosmesis.

Description:

The 'cosmesis' data frame has 95 rows and 3 columns.

Format:

This data frame contains the following columns:

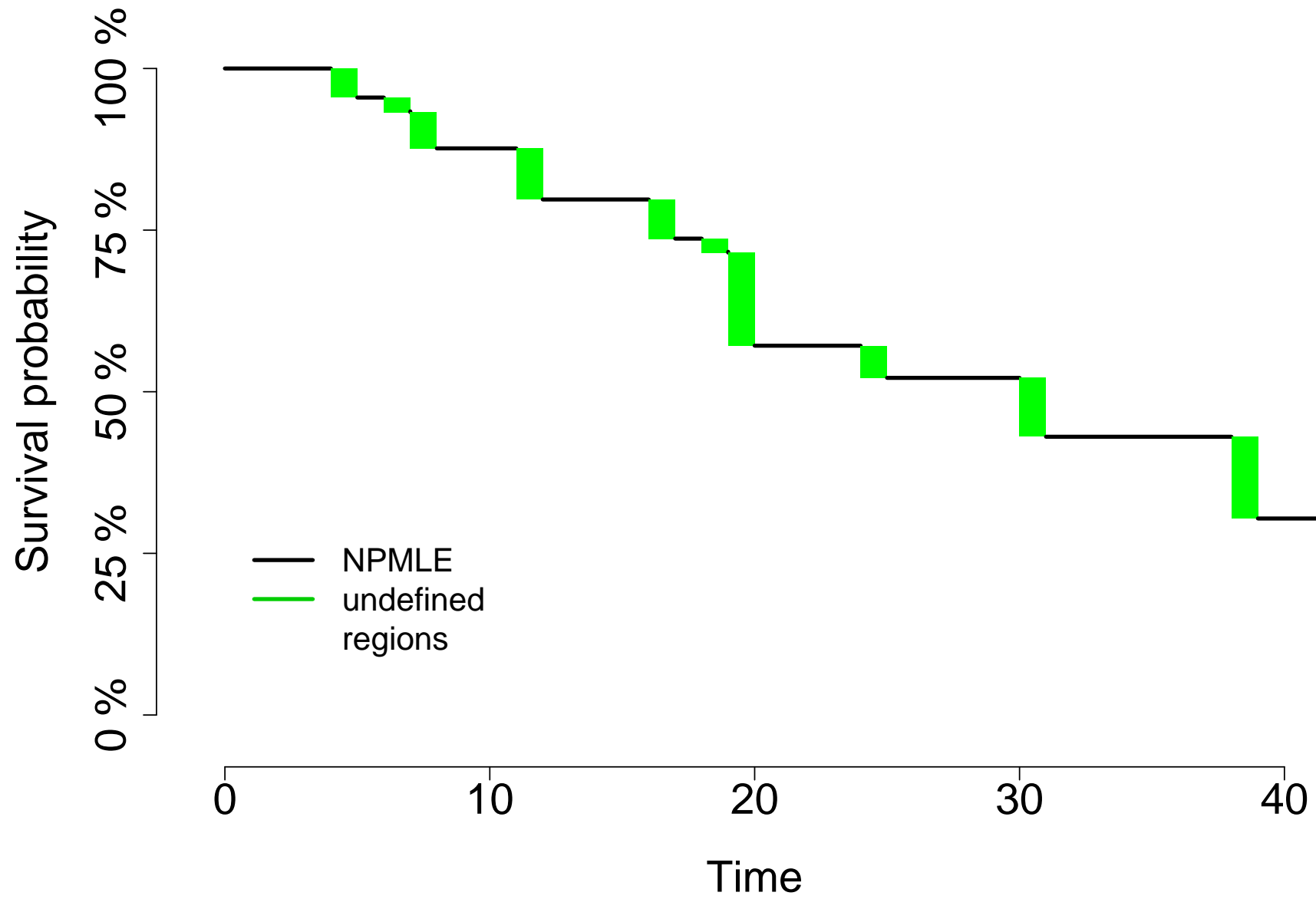
L The left end point of the cosmetic deterioration interval.

R The right end point of the cosmetic deterioration interval.

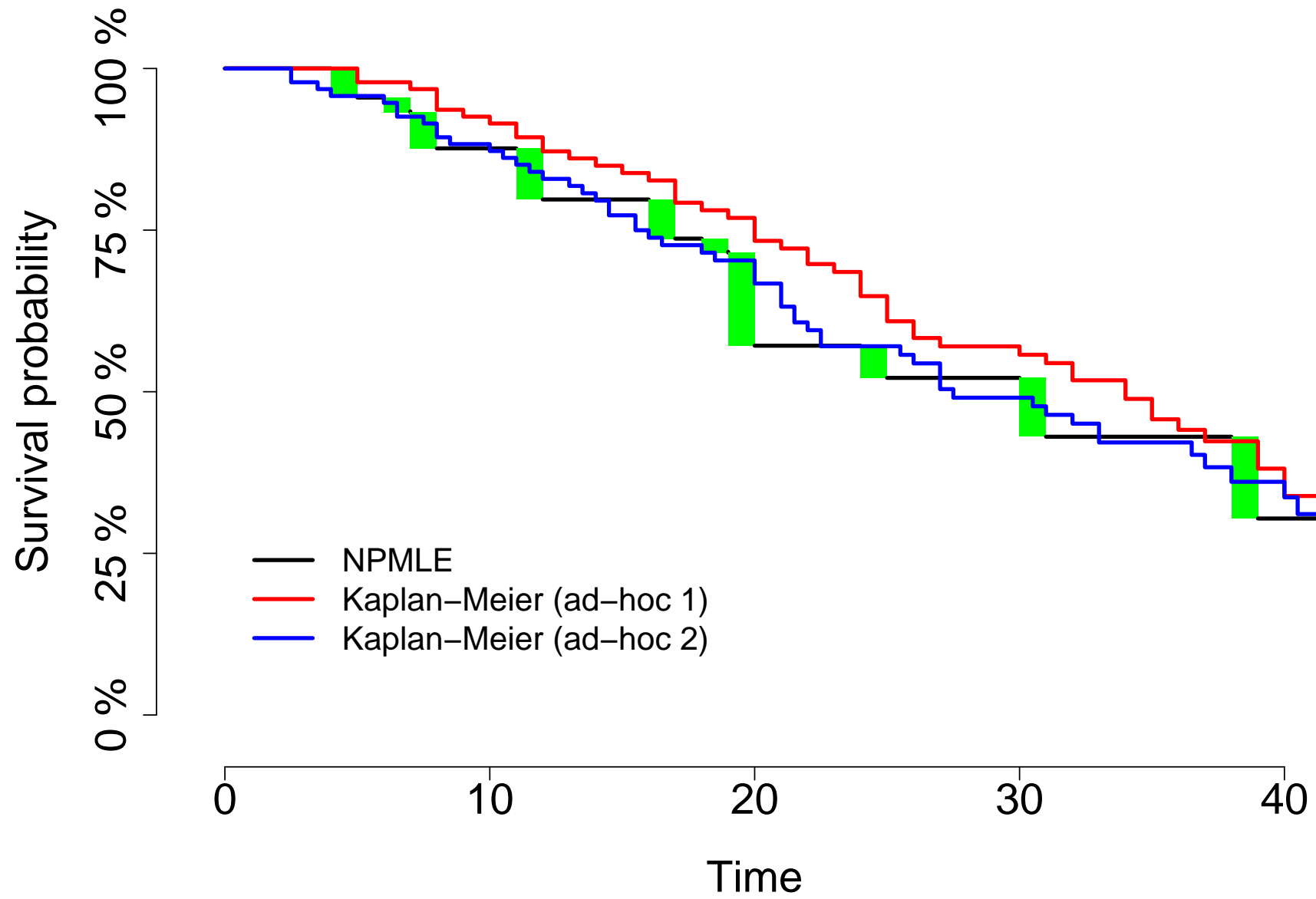
Source:

A semiparametric model for regression analysis of interval-censored failure time data, D. M. Finkelstein and R. A. Wolfe, 1985, Biometrics.

Cosmesis data



Cosmesis data

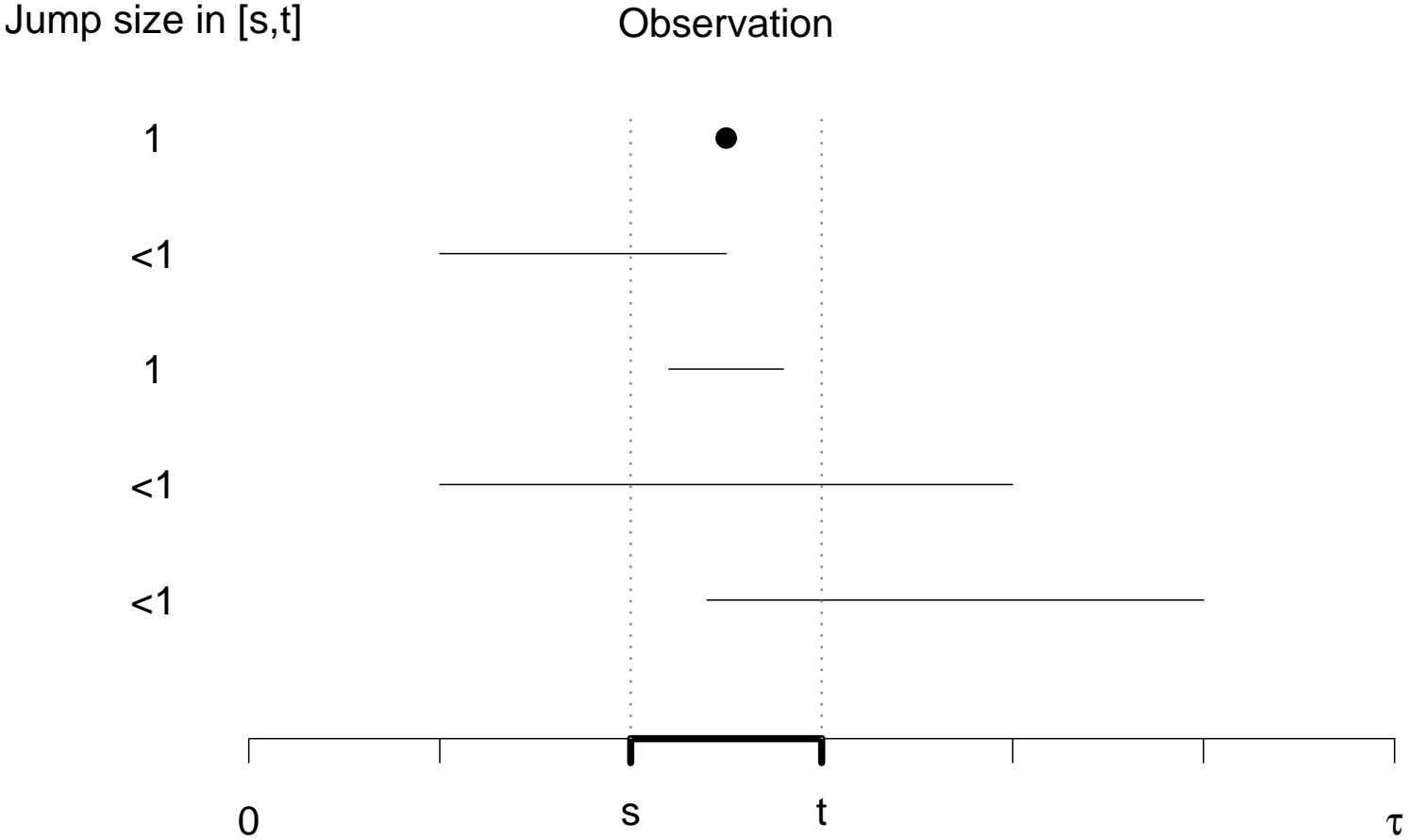


How do the interval censored observations contribute to NPMLE?

Under **coarsening at random** (Gill et al., 1995)

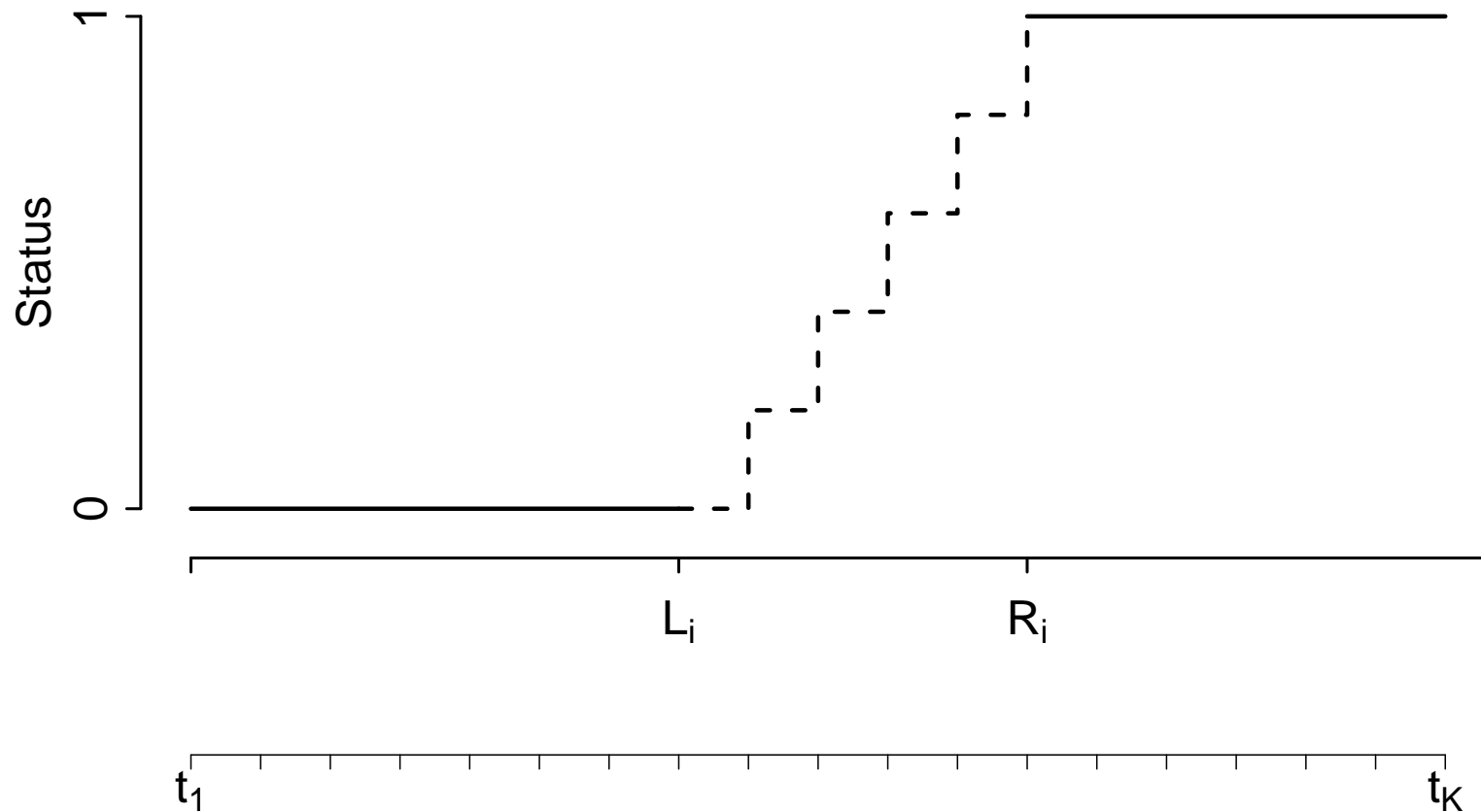
$$\begin{aligned} P(T_i \in dt \mid [L_i, R_i]) &\stackrel{!}{=} P(T_i \in dt \mid T_i \in [L_i, R_i]) \\ &= \frac{\mathcal{I}\{t \in [L_i, R_i]\}}{S(L_i) - S(R_i)} S(dt) \\ &\approx \frac{\mathcal{I}\{t \in [L_i, R_i]\}}{R_i - L_i} dt \\ &\equiv \text{Step one} \end{aligned}$$

Idea of the one-step estimate for mixed case interval censored observations



Interval censored observations

One step estimate: introduce a grid t_1, \dots, t_K and then redistribute the mass $1/n$ uniformly over $[L_i, R_i]$:



Interval censored observations

Let $Q_k = (t_k, t_{k+1})$ and

$$N_i^{(1)}(dQ_k) = \frac{\min(R_i, t_{k+1}) - \max(L_i, t_k)}{R_i - L_i}$$

$$Y_i^{(1)}(t_k) = \frac{R_i - \max(L_i, t_k)}{R_i - L_i}$$

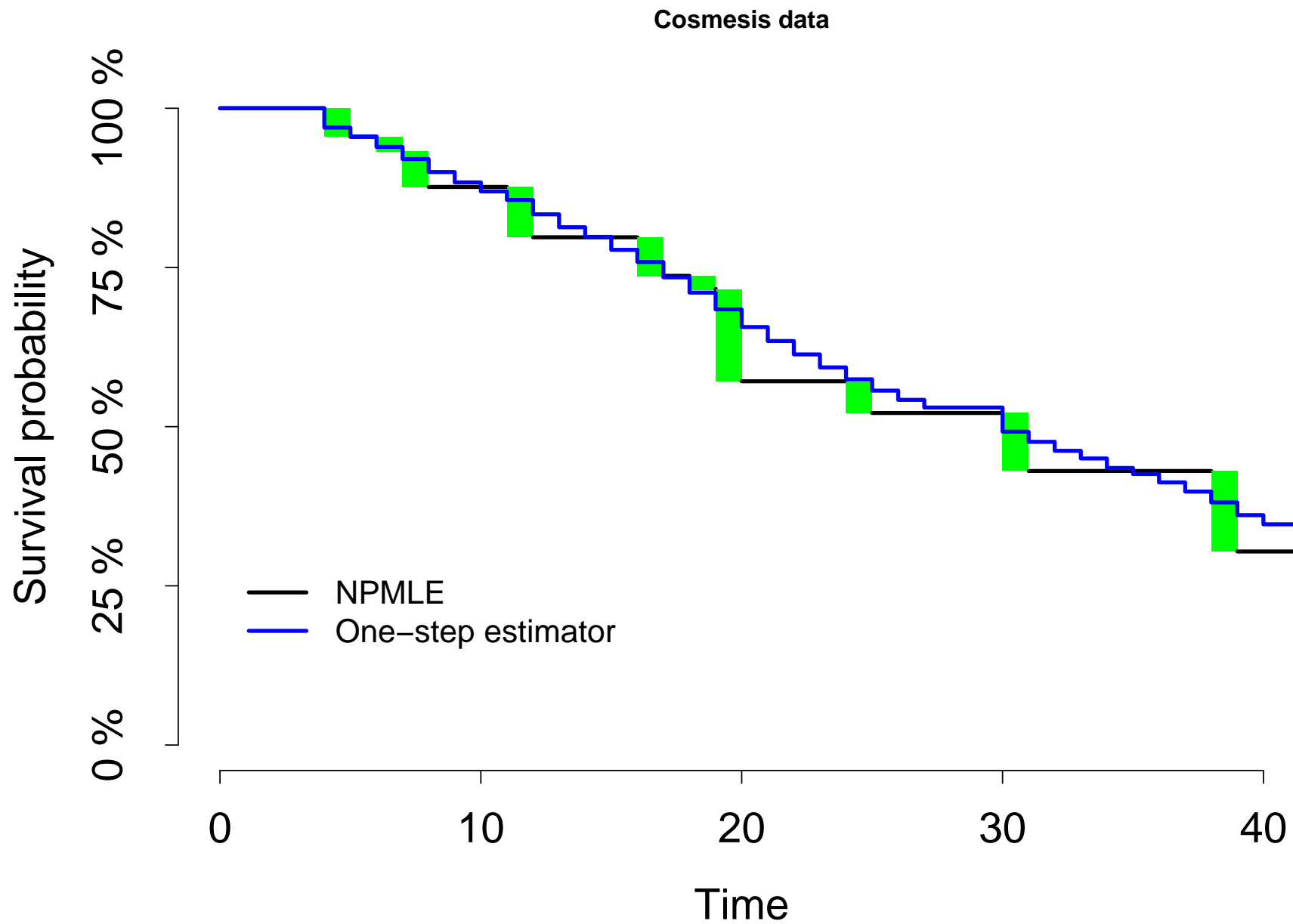
$$N = \sum_i N_i$$

$$Y = \sum_i Y_i$$

The one-step estimator:

$$\widehat{S}^{(1)}(t) = \prod_{k:t_k \leq t} \left(1 - \frac{N^{(1)}(dQ_k)}{Y^{(1)}(t_k)} \right).$$

Cosmesis data



Interval censored observations

Step $j+1$:

$$N_i^{(j+1)}(dQ_k) = \frac{\min(\widehat{S}^{(j)}(L_i), \widehat{S}^{(j)}(t_k)) - \max(\widehat{S}^{(j)}(R_i), \widehat{S}^{(j)}(t_{k+1}))}{\widehat{S}^{(j)}(L_i) - \widehat{S}^{(j)}(R_i)}$$

$$Y_i^{(j+1)}(t_k) = \frac{\min(\widehat{S}^{(j)}(L_i), \widehat{S}^{(j)}(t_k)) - \widehat{S}^{(j)}(R_i)}{\widehat{S}^{(j)}(L_i) - \widehat{S}^{(j)}(R_i)}.$$

The $j+1$ step estimator is:

$$\widehat{S}^{(j+1)}(t_r) = \prod_{k \leq r} \left[1 - \frac{dN^{(j)}(Q_k)}{Y^{(j)}(t_k)} \right].$$

With grid points $t_1, \dots, t_K = \text{sort}(\text{unique}(c(L_i, R_i)))$ Halina shows that this algorithm finds the NPMLE ...

Intermediate conclusions

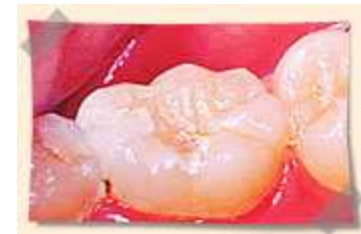
- ▶ Using counting processes for interval censored data leads to a new way of understanding NPMLE
- ▶ Right censored and exact observations enter into the product limit form in the usual way
- ▶ The adhoc midpoint approach has little bias (Lindsey, 1998) but usually underestimates the variability
- ▶ Maybe the asymptotic variance of the one-step estimator is the same as the asymptotic variance of NPMLE ...

Fillings in primary teeth

Between 1991 and 1992, 1000 fillings were placed by 14 dentists in type II cavities of 648 children (aged between 3 and 13 years). The aim of the study was to compare longevity and cariostatic effects of conventional glass ionomer Ketac-fil (GIC) and encapsulated amalgam (AM) fillings:

$$F(t) = Prob(\text{failure before } t \mid \text{material})$$

*Qvist et al. (2004)



Fillings in primary teeth

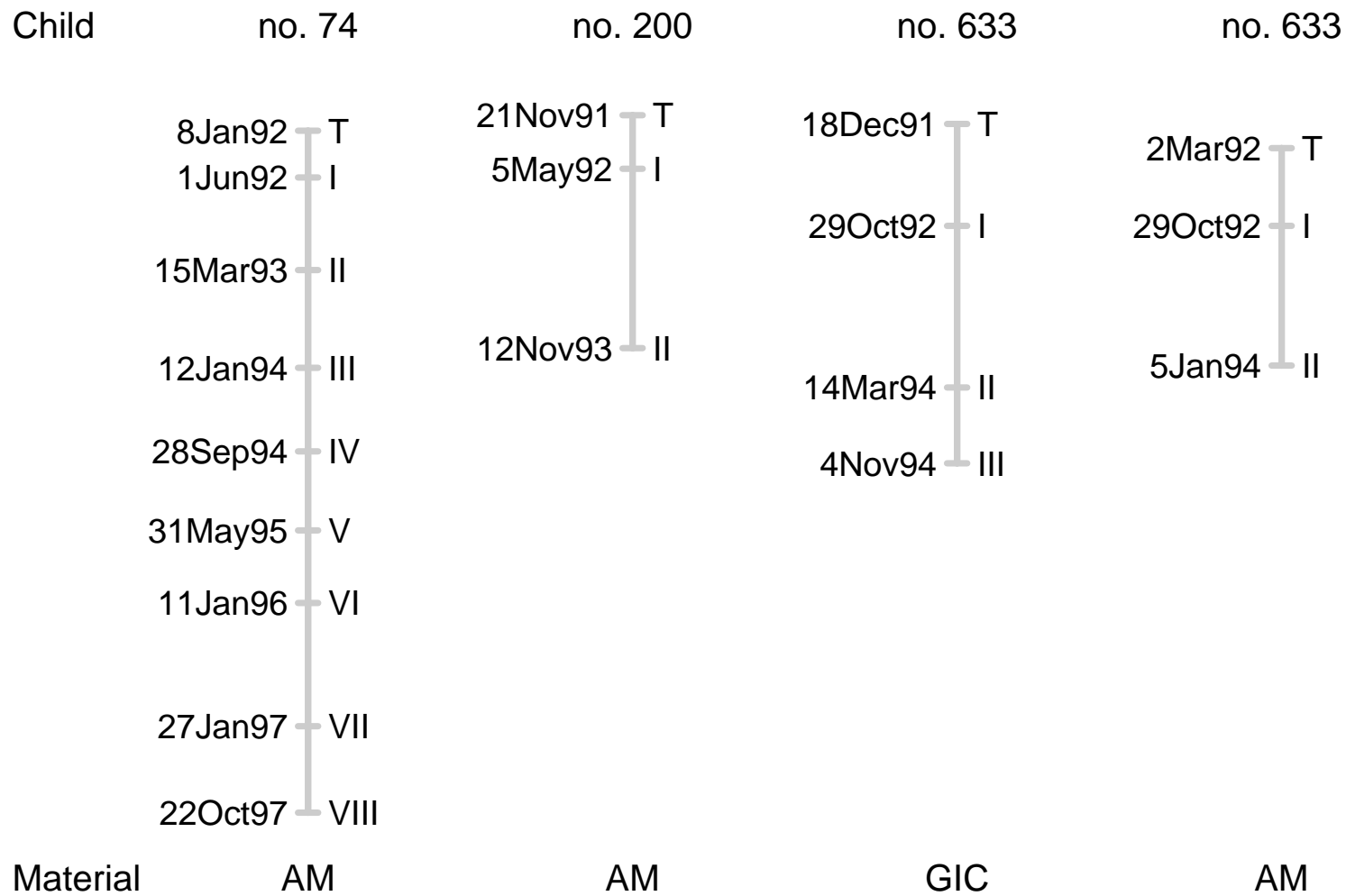
End of the study time:

Event	Overall	AM	GIC
Exfoliation*	561	334	227
Endodontic complications	72	38	34
Fracture of restoration	146	28	118
Other complications	96	43	53
Drop-out	125	86	39
Total	1000	529	471

*The replacement of a primary by a permanent teeth.

Fillings in primary teeth

Follow-up charts for fillings in primary teeth

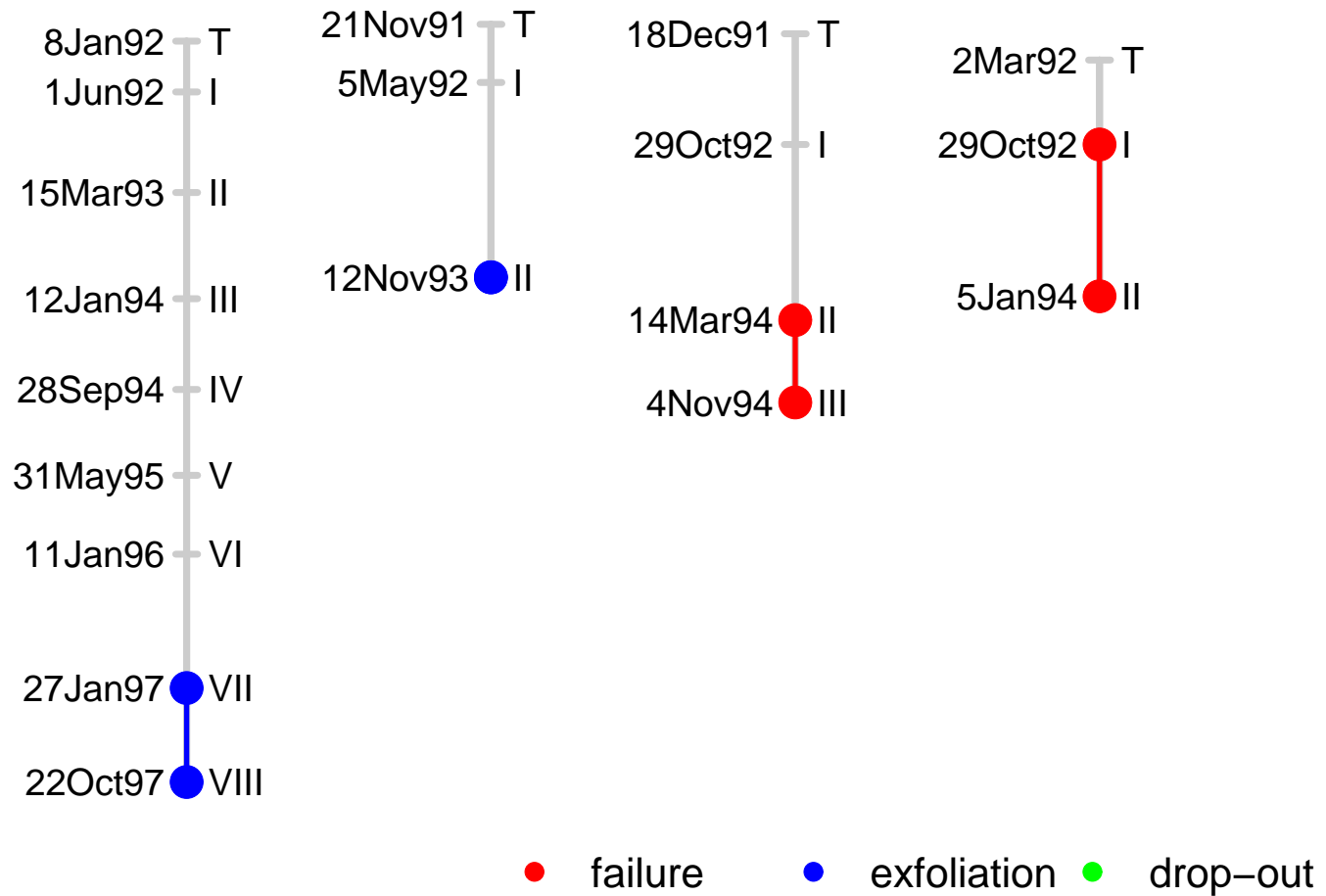


T : Treatment

I–XI : Visit 1–11

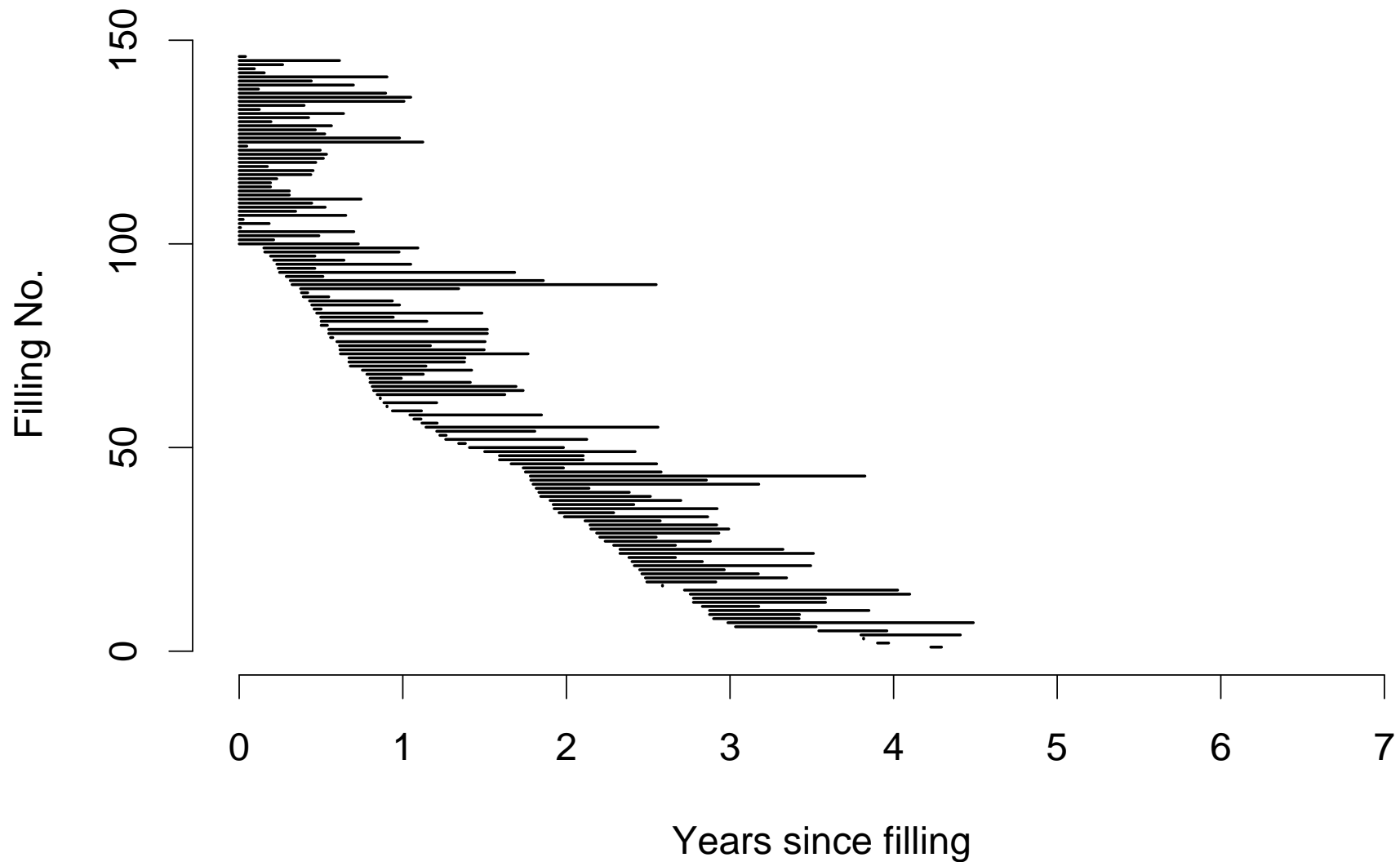
Fillings in primary teeth

Follow-up charts for fillings in primary teeth



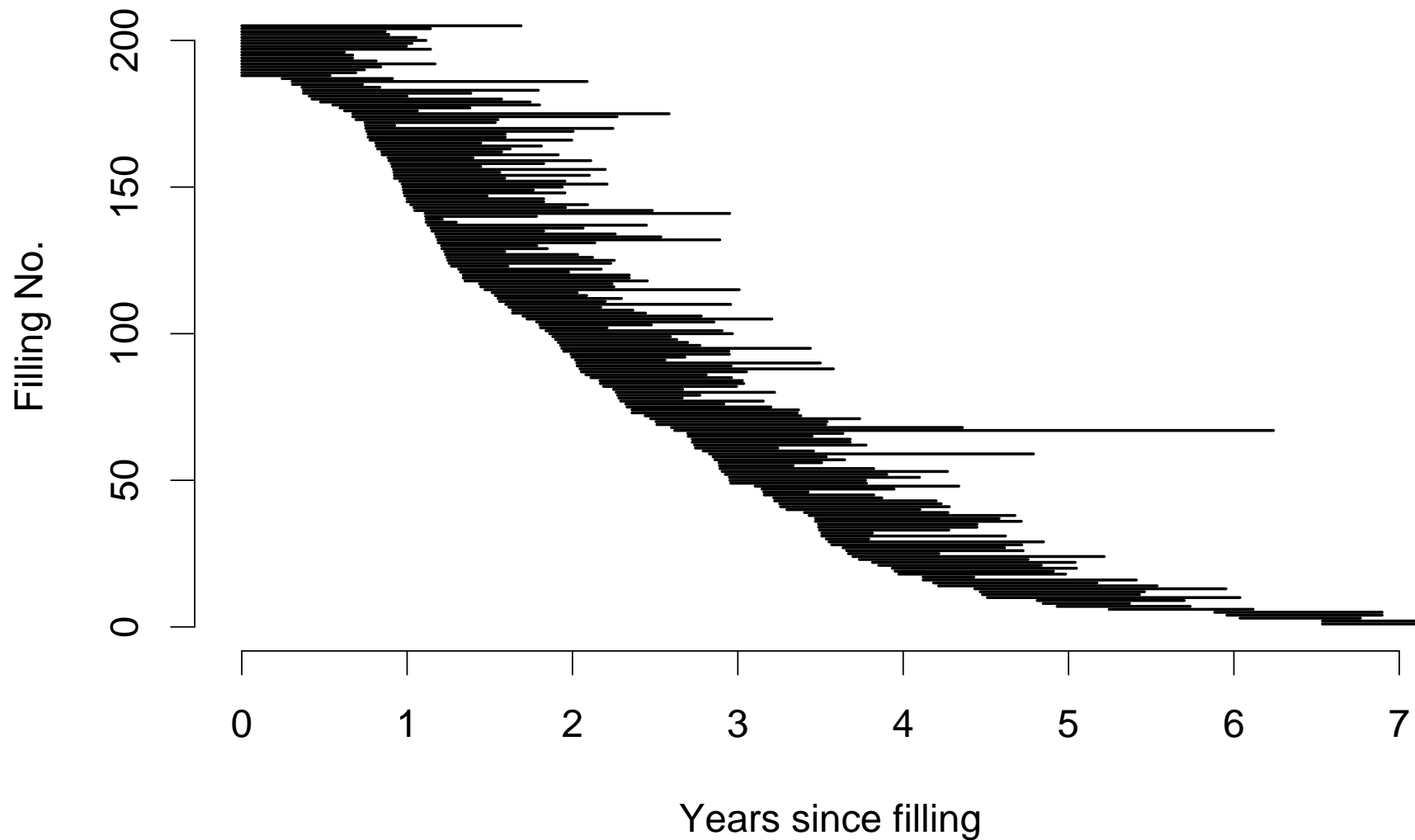
Fillings in primary teeth

Interval censored observations of fractured fillings (both materials)

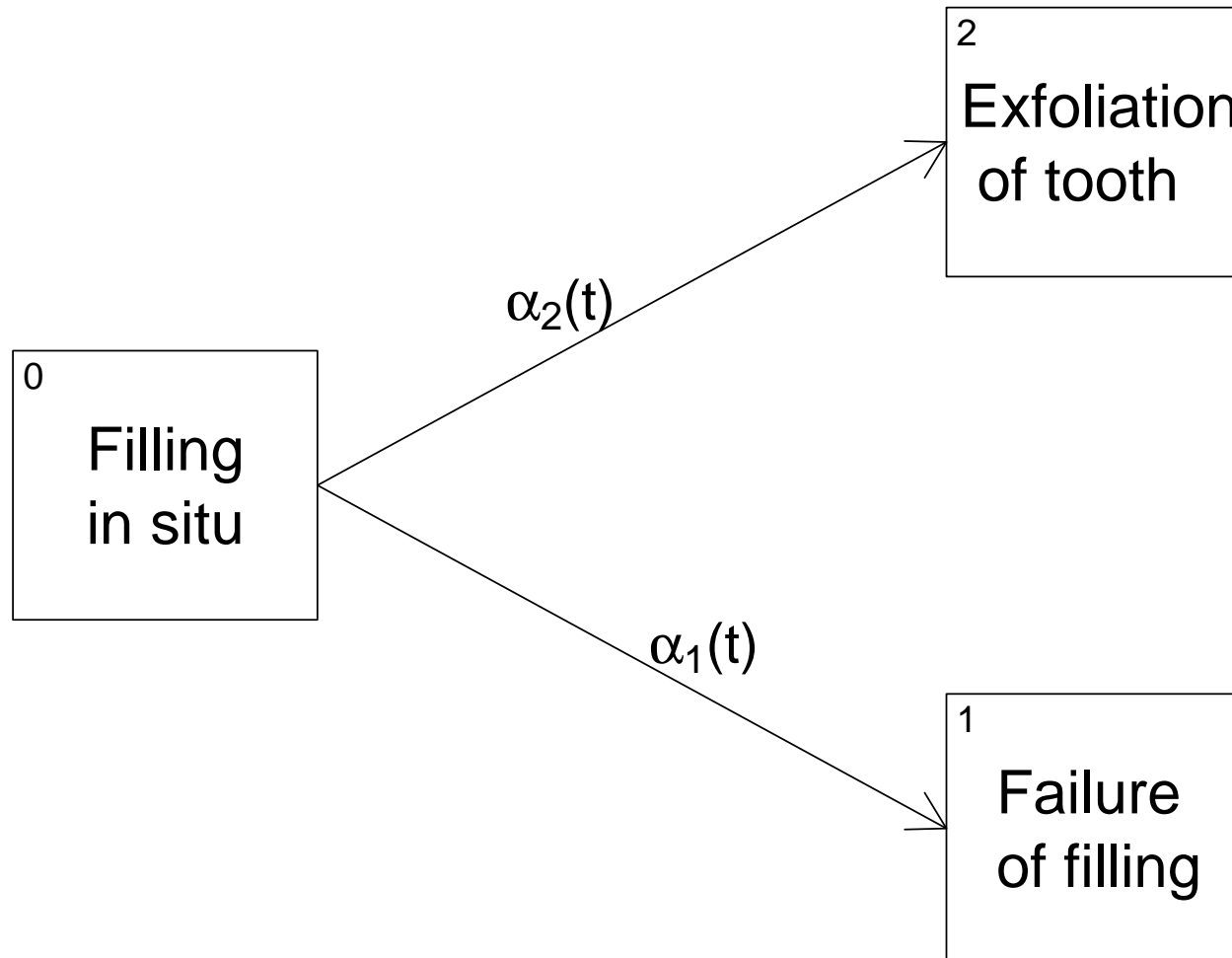


Fillings in primary teeth

Interval censored observations of exfoliated teeth (only GIC)



Competing risks model



Competing risks: long-term predictions

In the competing risks model the probability that a filling fails in the interval $[0, t]$, i.e. failure *before* exfoliation, is well described by the sub-distribution function F_1 .

F_1 is generally defined by

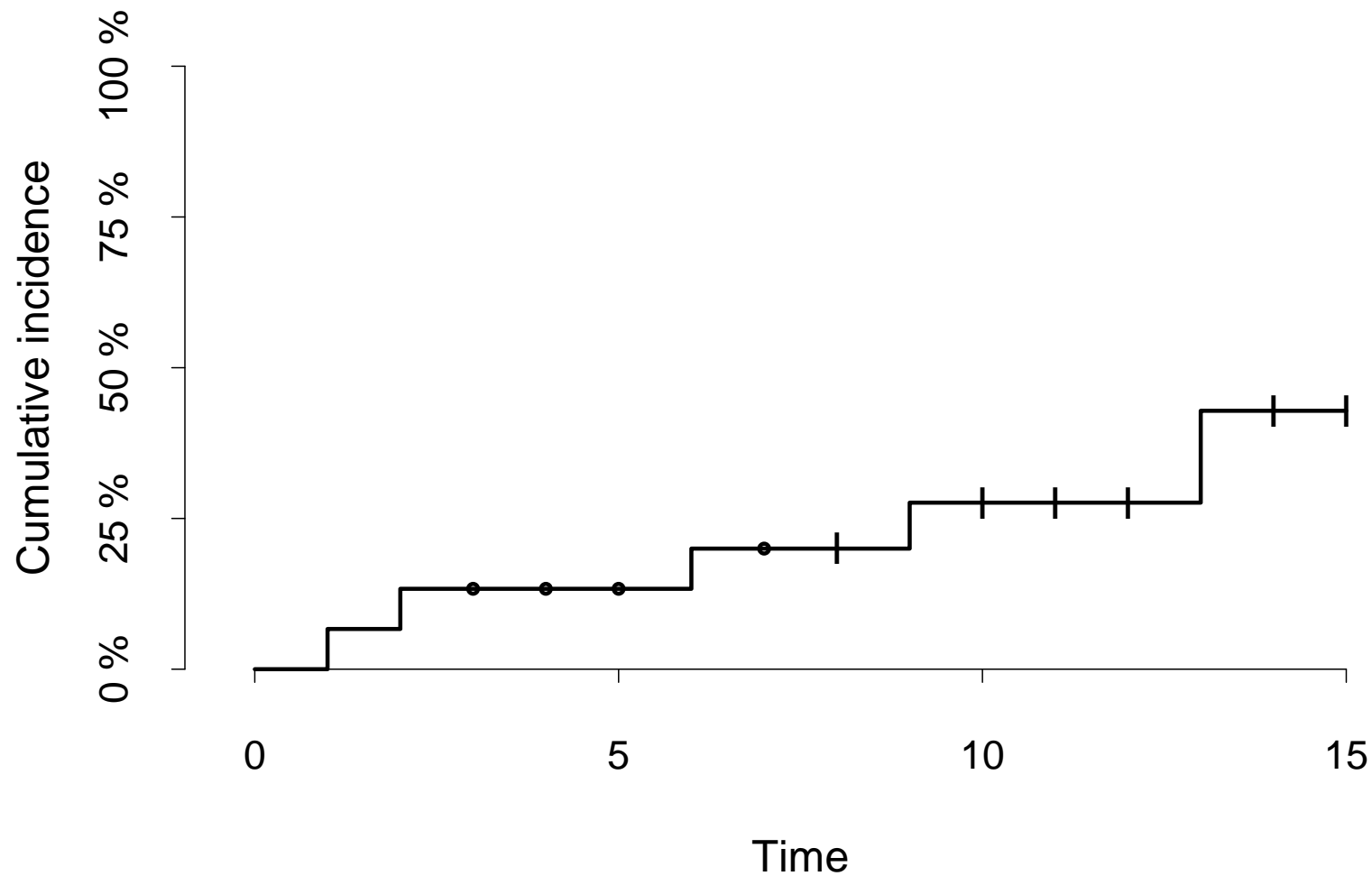
$$F_1(t) = P(T \leq t, \text{cause}=1) = \int_0^t S(s-) \alpha_1(u) du.$$

and is sometimes termed *cumulative incidence*.

Ad-hoc: apply the Aalen-Johansen estimator to the interval midpoints (or right endpoints).

Right censored observations

The Aalen-Johansen estimate also redistributes the mass $1/n$ of censoring times to the right and it absorbs the mass $1/n$ for events of competing causes



Consistent estimation of the cumulative incidence function

The nonparametric likelihood for F_1 and an algorithm for finding the NPMLE are provided by Hudgens et al. (2001)^a.

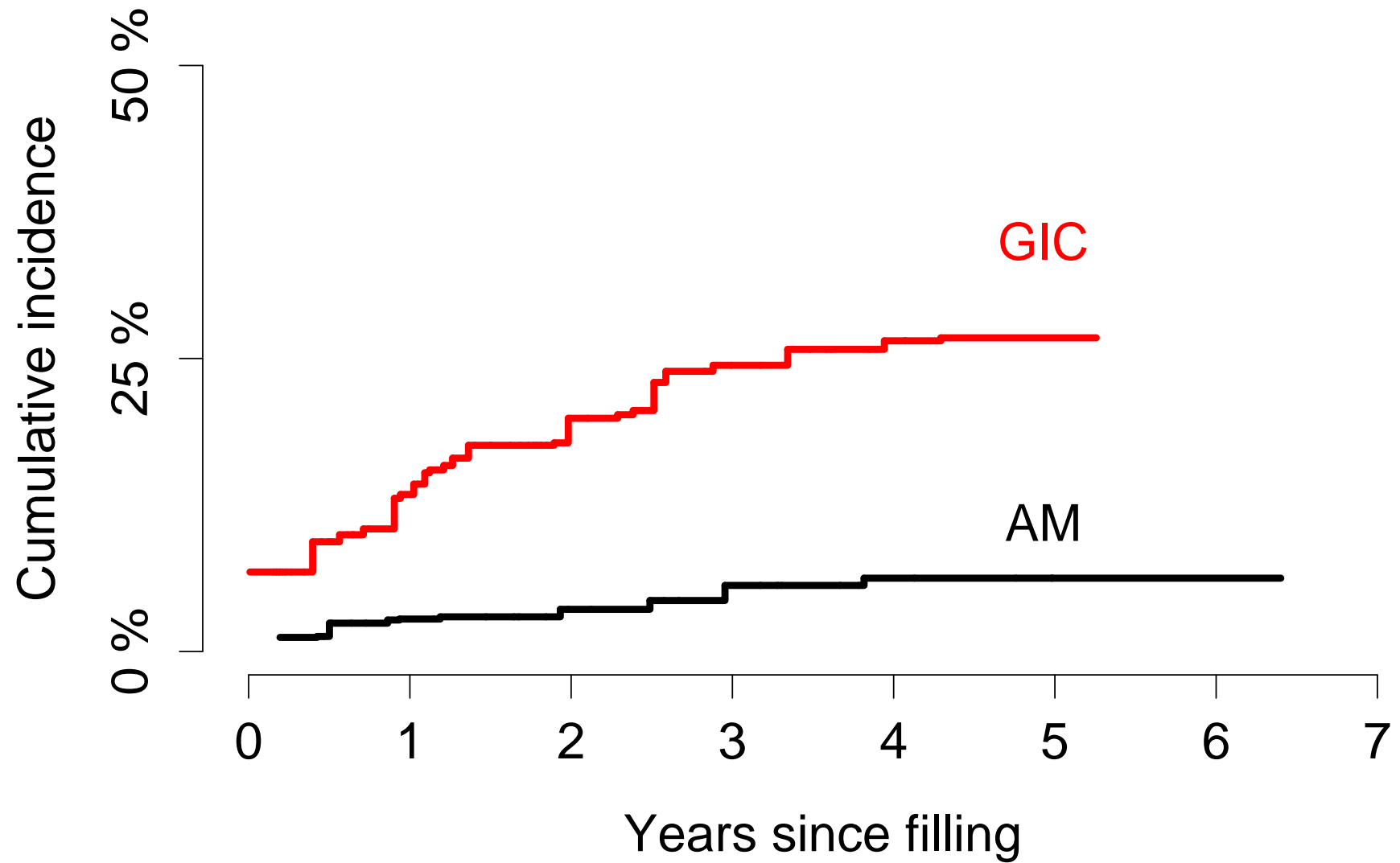
The one-step estimator and the algorithm using the product limit formula can be extended to competing risks (with Halina & Randi).

Standard errors via bootstrap?

^aSAS macro available: <http://www.bios.unc.edu/~mhudgens/>

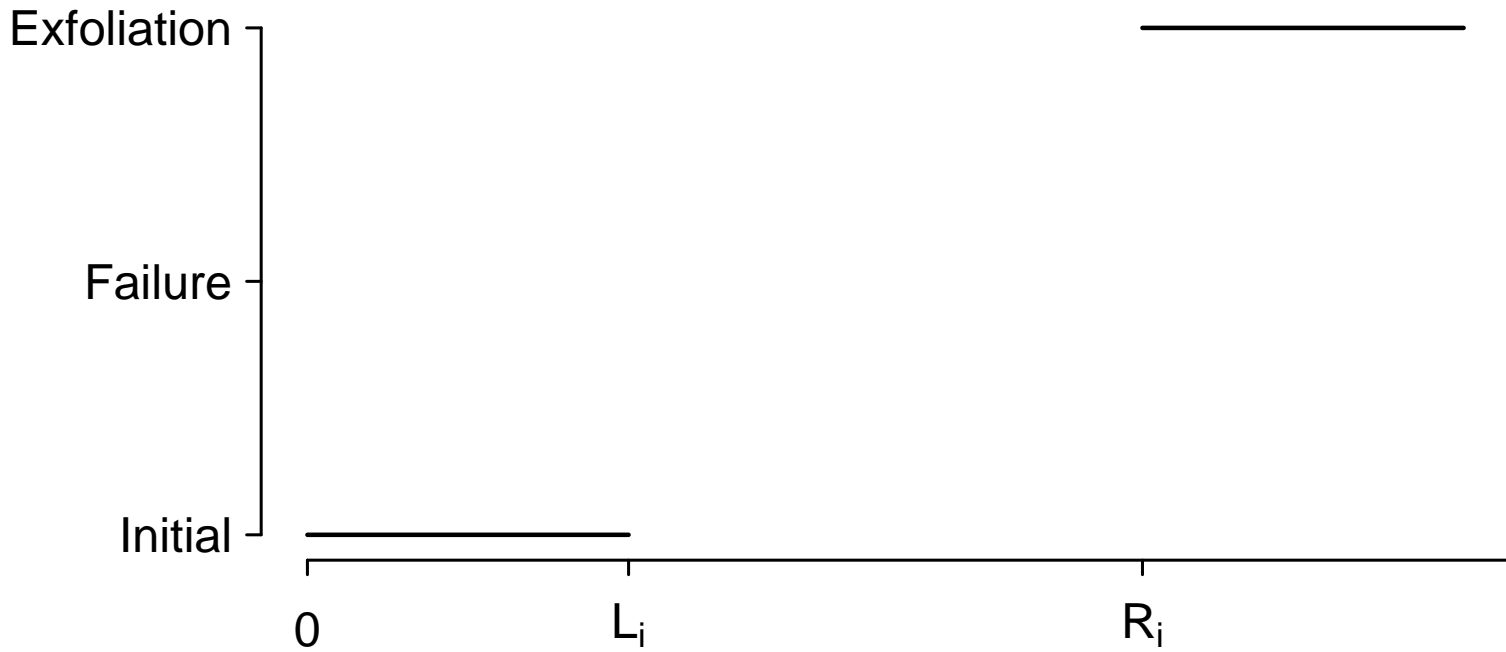
Consistent estimation of the cumulative incidence function

NPMLE

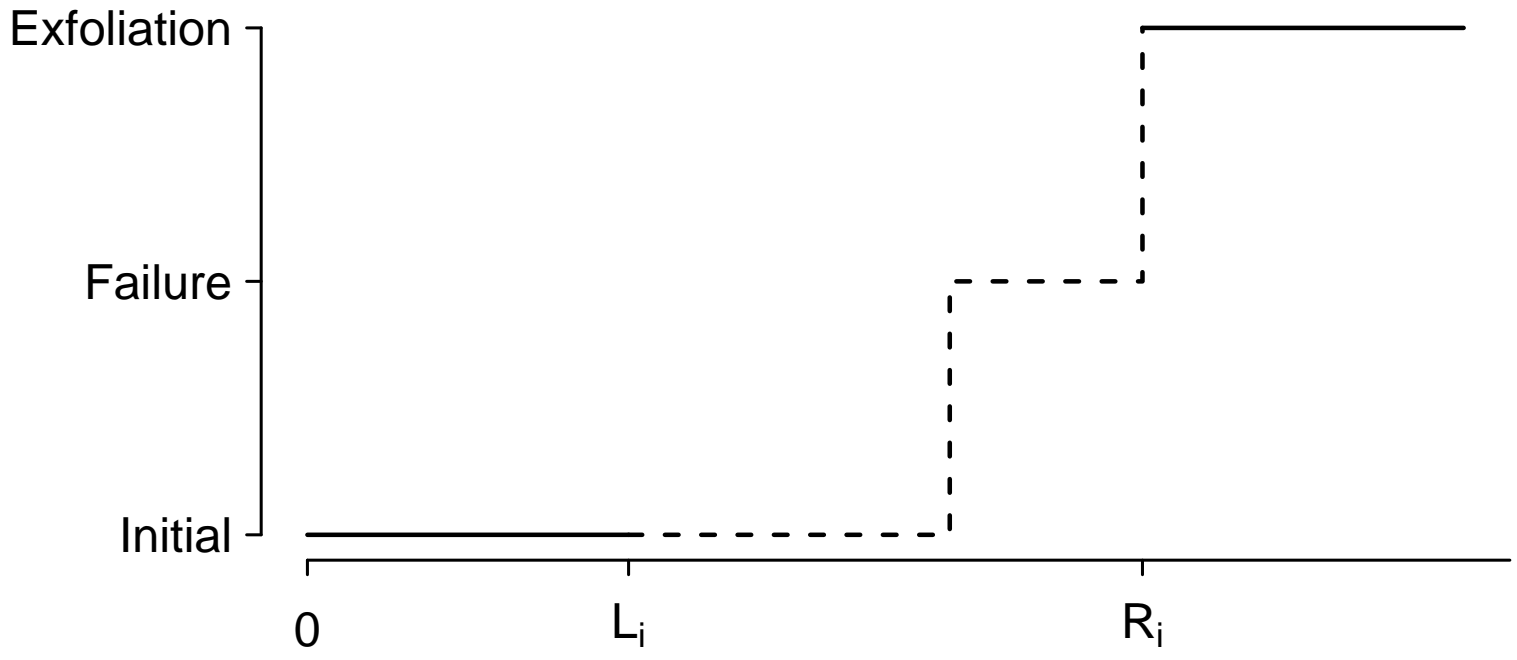


The end?

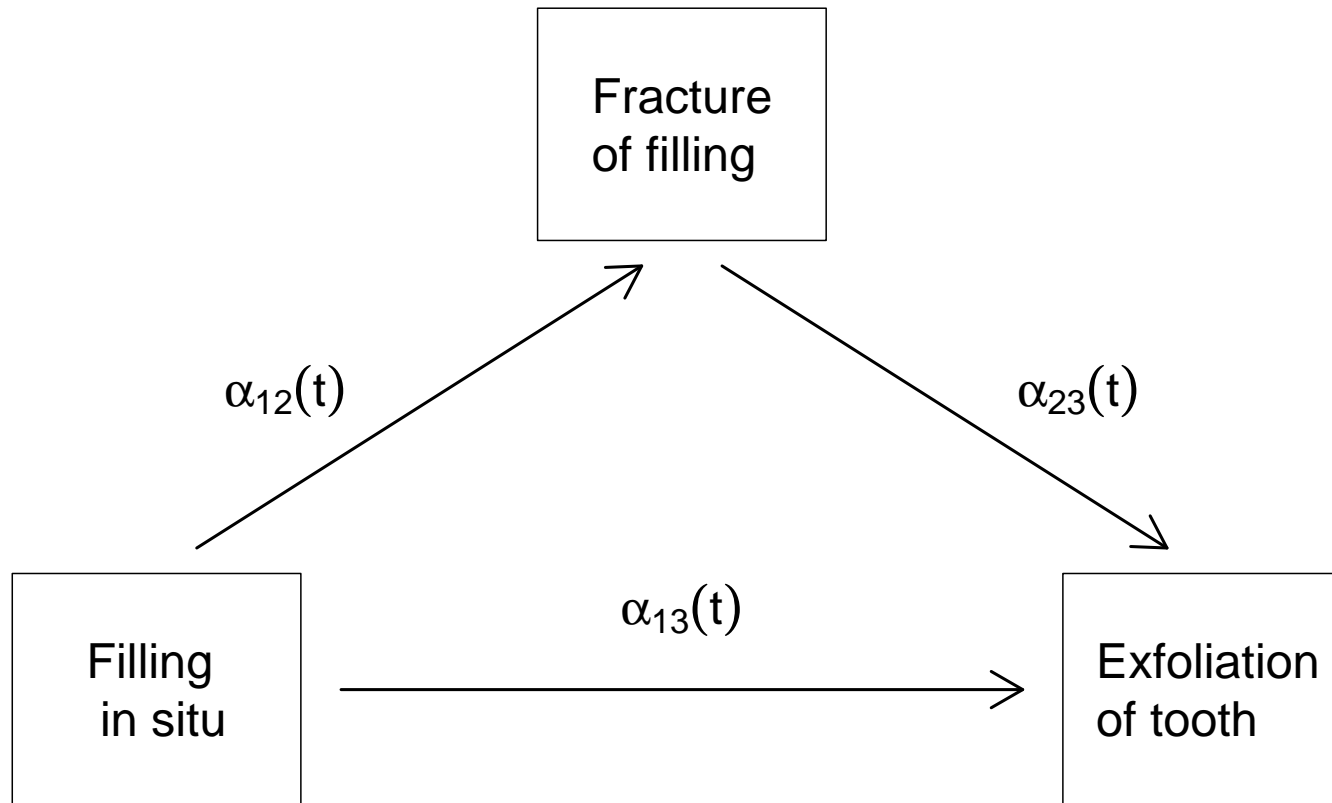
There could be unobserved failures ...



They would look like this ...



Because this really is an illness-death model:

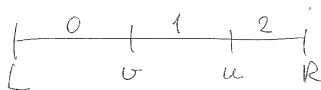


Modified likelihood

Niels K.

$$P_{02}(L, R) = \int_L^R e^{-\varepsilon(u) - \varphi(u)} \varepsilon(u) du$$
~~$$+ \int_L^R \left(\int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) dv \right) \cdot e^{-\varepsilon(u) - \alpha(u)} du$$~~

For $\alpha = \varepsilon$



$$\int_L^R \int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) e^{-\varepsilon(u) - \alpha(u)} \alpha(u) dv du$$

For $\alpha = \varepsilon$

~~$$\int_L^R \int_L^u e^{-\varepsilon(v) - \varphi(v)} \varphi(v) e^{-\varepsilon(u) - \alpha(u)} \alpha(u) dv du$$~~

$$= \int_L^R e^{-\varepsilon(u) - \alpha(u)} \alpha(u) du$$

Daniel C.

4.2 Likelihood for coarsened observation

Let us first compute the likelihood on the atoms of \mathcal{O}' (here we must not omit the proportionality term because we shall have to combine the likelihood on two events). Consider first $D_{1l} \cap D_{2l}$. From Lemma 2 or Theorem 2 in Commenges and Gégout-Petit (2006) we have on this event:

$$\mathcal{L}_{\mathcal{O}'} = \frac{1}{(e^{-v_{l-1}} - e^{-v_l})^2} \int_{v_{l-1}}^{v_l} f_1(s_1, s_2) ds_1 ds_2.$$

On $E_1 \cap D_{2l}$ the likelihood is:

$$\mathcal{L}_{\mathcal{O}'} = \frac{e^{v_l}}{(e^{-v_{l-1}} - e^{-v_l})} \int_{v_{l-1}}^{v_l} f_2(C, s_2) ds_2.$$

Here $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ represent the likelihoods for non-coarsened observation for possible values of T_1 and T_2 given by the first and second arguments respectively. We have for the first equation

$$f_1(s_1, s_2) = 1_{\{s_1 < s_2\}} \alpha_{01}(s_1 - b) \alpha_{12}(s_2) \exp\left[-\int_b^{s_1} \alpha_{01}(t - b) dt - \int_0^{s_1} \alpha_{02}(t) dt - \int_{s_1}^{s_2} \alpha_{12}(t) dt\right],$$

and in second equation:

$$f_2(C, s_2) = \alpha_{02}(s_2) \exp\left[-\int_b^{s_2} \alpha_{01}(t - b) dt - \int_0^{s_2} \alpha_{02}(t) dt\right].$$

Now on $[D_{1l} \cup E_1] \cap D_{2l}$ the likelihood $\mathcal{L}_{\mathcal{O}}$ is given by formula (1). It is easy to find that $P_{0l} = \frac{(e^{-v_{l-1}} - e^{-v_l})^2}{(e^{-v_{l-1}} - e^{-v_l})^2 + e^{-v_l}(e^{-v_{l-1}} - e^{-v_l})}$ so that the likelihood $\mathcal{L}_{\mathcal{O}}$ is proportional to:

$$P_{\alpha}([D_{1l} \cup E_1] \cap D_{2l}) = \int_{v_{l-1}}^{v_l} f_1(s_1, s_2) ds_1 ds_2 + \int_{v_{l-1}}^{v_l} f_2(C, s_2) ds_2.$$

Note that in this observation scheme the smallest event observed has non-null probability: the observed sigma-field can be decomposed in atoms and the likelihood is proportional to the probability of the observed atom. We might have found a more direct way to compute this probability but it is not completely obvious anyway. It remains to compute the likelihood on the other atoms. One must also take into account the truncation condition: the sample is constituted of teeth which have not exfoliated before b so that we must divide the likelihood by $P_{\alpha}(T_2 \geq b) = \exp\left[-\int_0^b \alpha_{02}(t) dt\right]$

Conclusions

- ▶ Nonparametric (Halina, Randi) and semiparametric methods (Daniel, Pierre) are soon available for this problem

Further research

- ▶ Appropriate confidence limits for cluster-correlated event times, e.g. with the bootstrap
- ▶ Regression models, e.g. via pseudo-values á la Andersen & Klein, or by treating L_i and R_i as current status data (van der Laan & Jewell).

References

- Dümbgen, L., S. Freitag, and G. Jongbloed (2006). Estimating a unimodal distribution from interval-censored data. *Journal of the American Statistical Association* 101, 1094.
- Gill, R. D., M. J. Van der Laan, and J. M. Robins (1995). Coarsening at random: Characterizations, conjectures and counter-examples. In D. Y. Lin and T. R. Fleming (Eds.), *Proceedings of the First Seattle Symposium in Biostatistics*, pp. 255–294. Springer Lecture Notes in Statistics.
- Groeneboom, P. and J. A. Wellner (1994). *Information Bounds and Non-parametric Maximum Likelihood Estimation*. Birkhäuser, Basel.
- Hudgens, M. G., G. A. Satten, and I. M. Longini (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* 57, 74–80.
- Lindsey, J. (1998). A study of interval censoring in parametric regression models. *Lifetime Data Anal.* 4(4), 329–354.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics* 22, 86–91.

Qvist, V., L. Laurberg, A. Poulsen, and P. T. Teglers (2004). Eight-year study on conventional glass ionomer and amalgam restorations in primary teeth. *Acta Odontol Scand* 62(1), 37–45.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. 38, 290–295.

Wellner, J. A. and Y. Zhan (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. 92, 945–959.