Tutorial on application of survival analysis to the assessment of multi-state models in clinical data analysis

Joachim Grevel

PAGE 2021 Tutorial, September 2021

Comparison of time-to-event analysis methods

Method	Most useful for
Cox PH model	Hazard ratio for relative risk (with conf. interval) for active treatment vs placebo (PFS, OS)
Parametric Cox PH model	Simulation of treatment effect, exposure-response (PFS, OS)
Fine-Gray PH model	Hazard ratio for absolute risk with dropout as competing risk (with conf. interval) for active treatment vs placebo (PFS, OS)
Semi-Markov multi-state model	Effects of active treatment vs placebo on transition risks and sojourn times

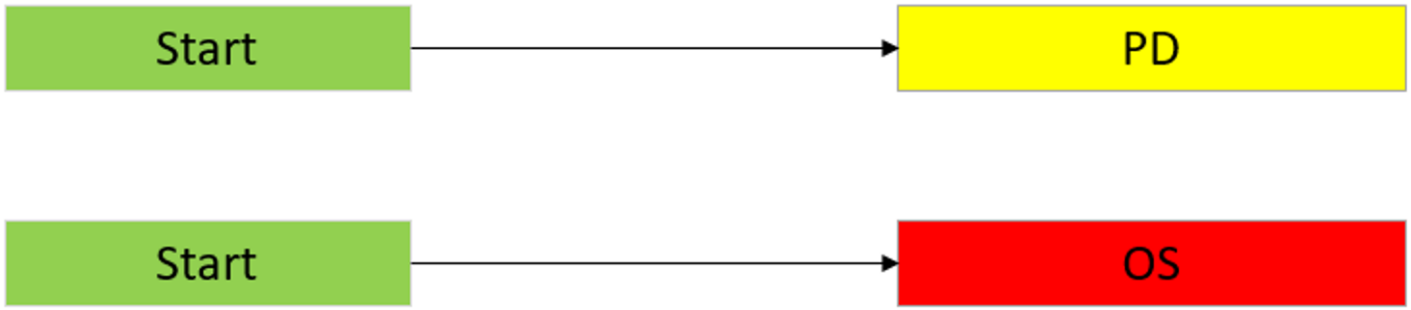
Comparison of time-to-event analysis methods (cont'd)

Method	Shortcomings
Cox PH model	Competing risks treated as right-censored
Parametric model	Competing risks treated as right-censored
Fine-Gray PH model	Competing risks weighted but unnatural risk set
Semi-Markov multi-state model	More complex models and larger data sets required for reliable inference

What events (states) are of interest in a phase 3 clinical study in oncology?

State	Type of state
Randomisation to treatment arm (start)	Initial state
Progressive disease (PD)	Transient state
Treatment discontinuation due to adverse effects (DISC)	Transient
Lost to follow-up (LOST)	Absorbing
Death	Absorbing
Administrative data cut-off	Not treated as own state

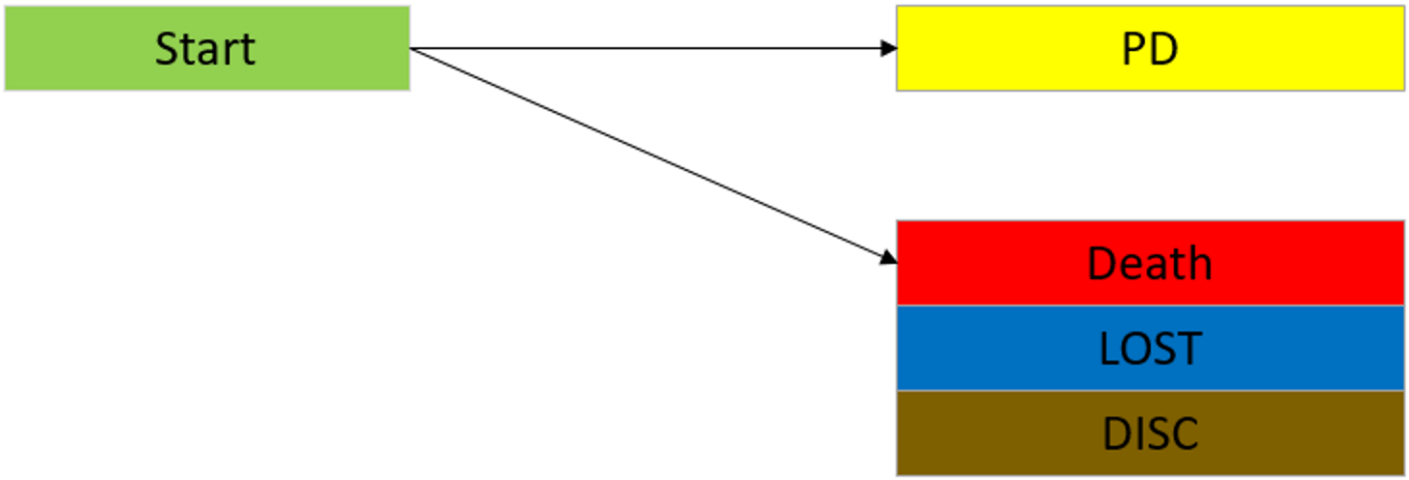
Cox PH model and parametric time-to-event model



Treated as right-censored:

- Administrative data cut-off
- DISC
- LOST
- Death
- PD

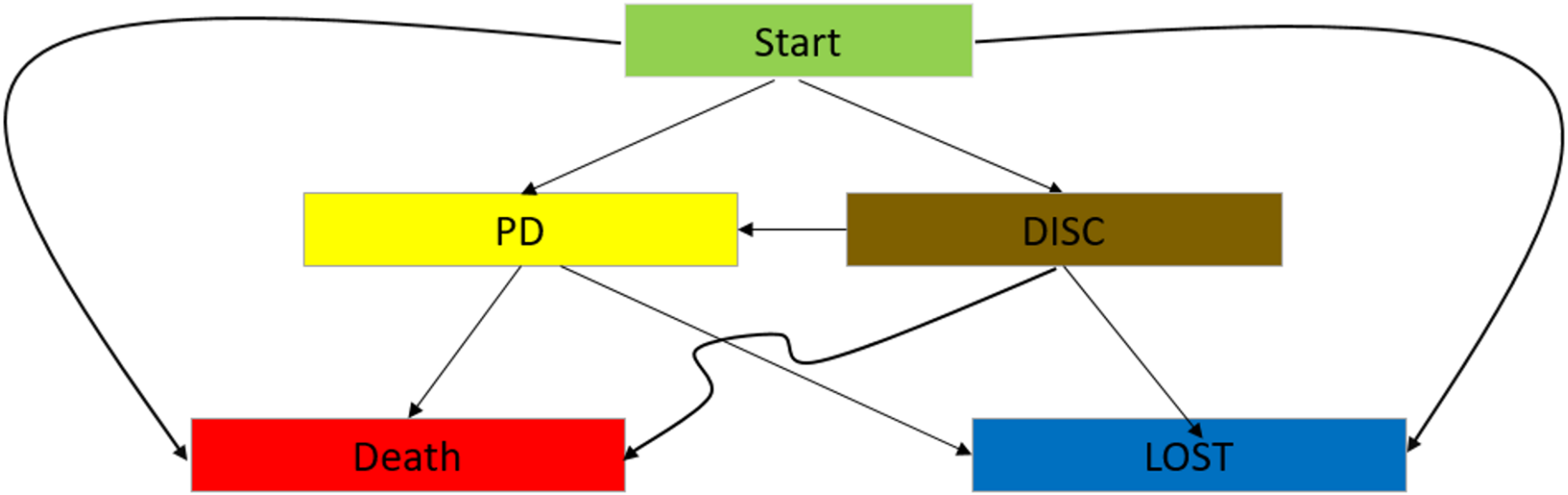
Fine-Gray PH model



Treated as right-censored:

- Administrative data cut-off

Semi-Markov multi-state model



Treated as right-censored:

- Administrative data cut-off

Simple multi-state model

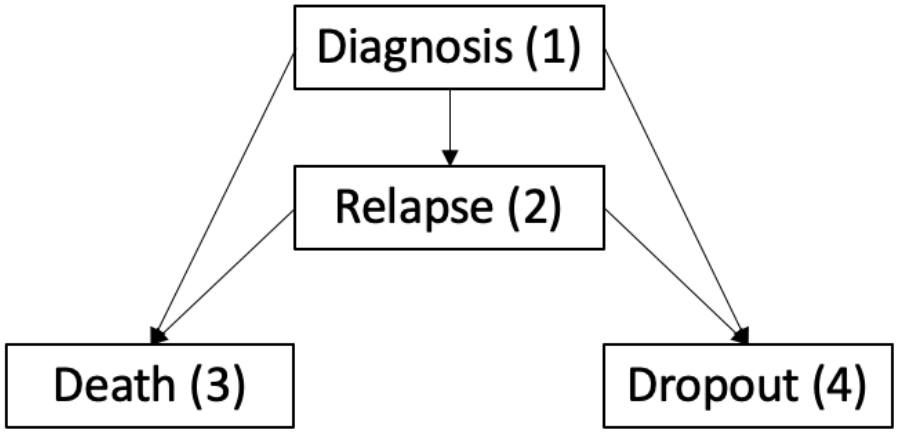


Figure: Diagrammatic representation of a simple multi-state model for an oncology drug trial

Comparing two possible approaches

Two possible approaches as per [Asanjarani et al., 2021]:

Approach I	Approach II
transition probabilities, sojourn time hazard functions	intensity transition functions (ITFs)
probabilities - direction hazard functions - rate	ITFs - both direction and rate
Incorporate covariates - yes	Incorporate covariates - yes
Able to specify parametric form	Able to specify parametric form
Able to simulate data	Able to simulate data
Larger computational expense	Moderate computational expense
Parametric and non-parametric estimates of quantities e.g. CIF	Parametric and non-parametric estimates of quantities e.g. CIF
95% confidence intervals of parametric estimates	95% confidence intervals of parametric estimates

Sojourn time hazard function and intensity transition function

As per [Asanjarani et al., 2021],

Sojourn time hazard function

$$\alpha_{ij}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < \tau_n \leq t + \Delta t | J_{n-1} = i, J_n = j, \tau_n > t) \quad (1)$$

Intensity transition function (ITF)

$$\tilde{\alpha}_{ij}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < \tau_n \leq t + \Delta t, J_n = j | J_{n-1} = i, \tau_n > t) \quad (2)$$

For a simple model with dichotomous covariate $Z \in \{0, 1\}$,

$$\alpha_{ij}(t|Z) = \alpha_{ij,0}(t) \exp\left(\beta_{ij} Z\right) \quad (3)$$

$$\tilde{\alpha}_{ij}(t|Z) = \tilde{\alpha}_{ij,0}(t) \exp\left(\tilde{\beta}_{ij} Z\right) \quad (4)$$

Transition of events among patients in simulated data

Weibull-like ITFs on each transition $i \to j$:

$$\tilde{\alpha}_{ij}(t|Z) = \frac{a_{ij}}{b_{ij}} \left(\frac{t}{b_{ij}} \right)^{a_{ij}-1} \exp(\beta_{ij} Z)$$

a_{ij} and b_{ij} are respectively similar to the shape and scale parameter associated with typical Weibull distributions (although ITFs are not associated with proper probability distributions!). β_{ij} is the covariate coefficient. Chosen values of parameters:

a_{12}	a_{13}	a_{14}	a_{23}	a_{24}
1.40	1.50	1.30	2.10	1.90
b_{12}	b_{13}	b_{14}	b_{23}	b_{24}
2.68	2.13	2.30	1.10	5.00
β_{12}	β_{13}	β_{14}	β_{23}	β_{24}
-0.90	-0.70	0.90	-0.10	-0.05

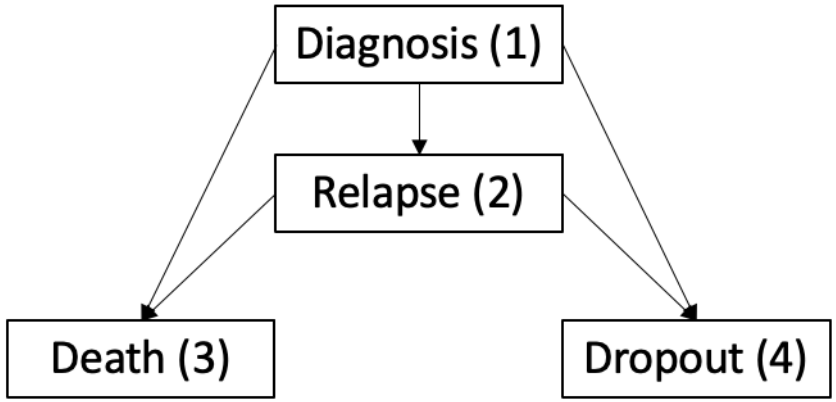
Simulation method

Let τ_n denote the sojourn time in previous state J_{n-1}.
Adapting algorithm in [Beyersmann et al., 2009]), for each patient:

1. Simulate an event time from the survival function of the holding time distribution at current state,
   $S_i(t) = P(\tau_n > t | J_{n-1} = i) = \exp\left(- \int_0^t \sum_{k \neq i} \tilde{\alpha}_{ik}(u) \mathrm{d}u \right)$
2. Carry out a multinomial experiment to decide the event based on
   $P(J_n = j | t < \tau_n \leq t + \Delta t, J_{n-1} = i, \tau_n > t) \approx \frac{\tilde{\alpha}_{ij}(t)}{\sum_{k \neq i} \tilde{\alpha}_{ik}(t)}$
3. Repeat until an absorbing state is reached

If censoring is desired, simulate a right-censoring time at step 1 and continue the algorithm until either (i) the sum of sojourn times in different states exceeds the censoring time, or (ii) an absorbing state is reached.
If the censoring time is exceeded as per (i), record the last sojourn time in a transient state as a right-censoring time.

Details of simulated study



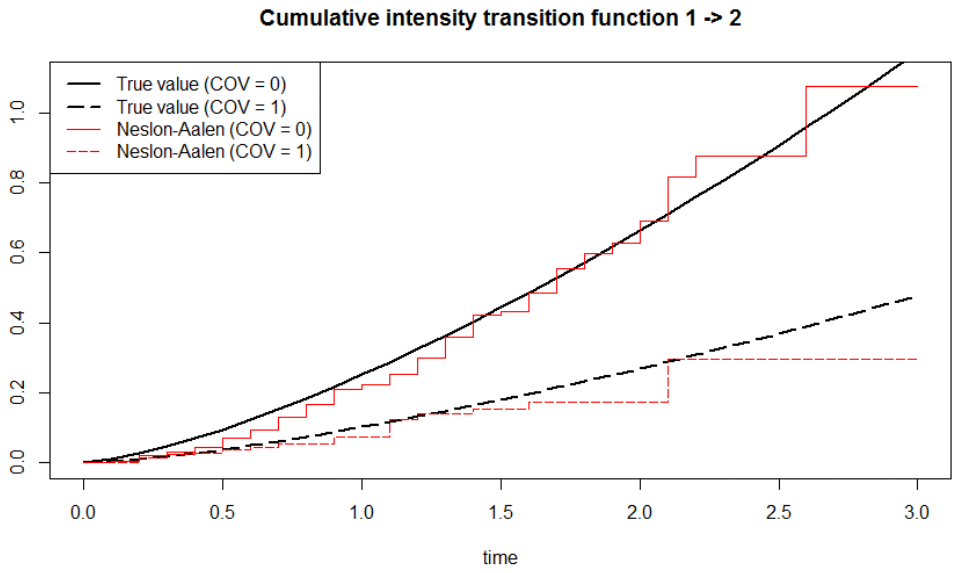
- $1 \rightarrow 2$: 159 patients
- $1 \rightarrow 3$: 220 patients
- $1 \rightarrow 4$: 500 patients

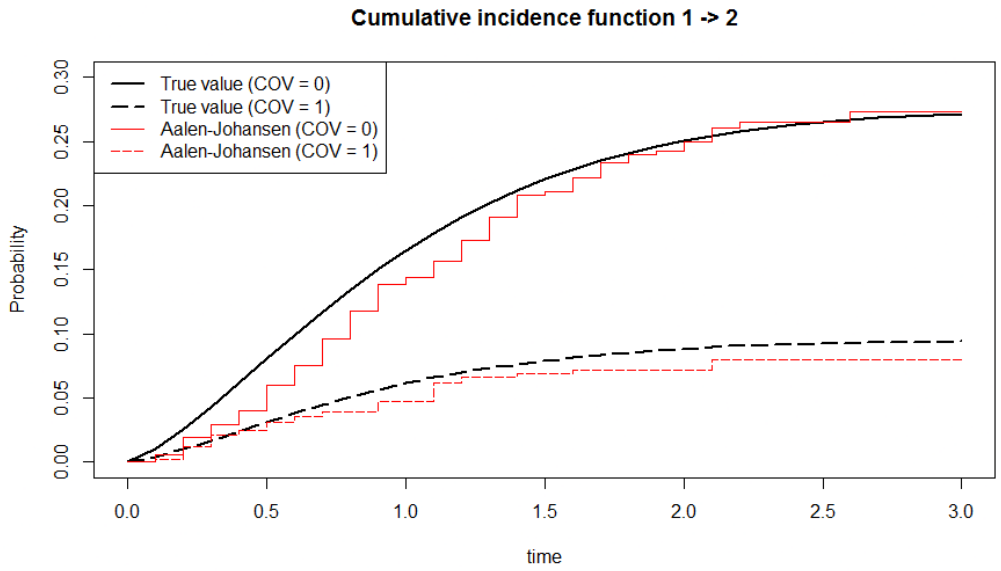
121 patients right-censored in state 1

- $2 \rightarrow 3$: 84 patients
- $2 \rightarrow 4$: 5 patients

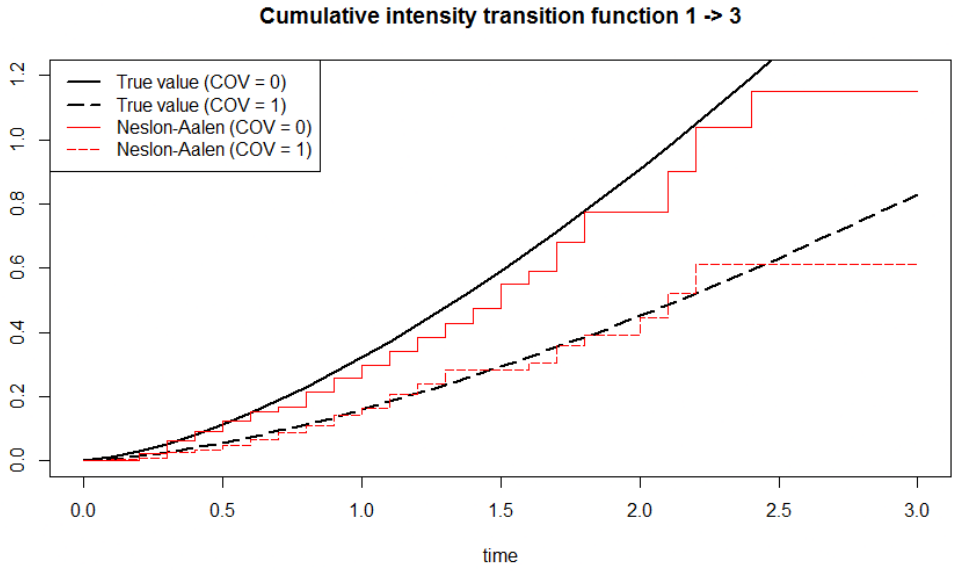
70 patients right-censored in state 2

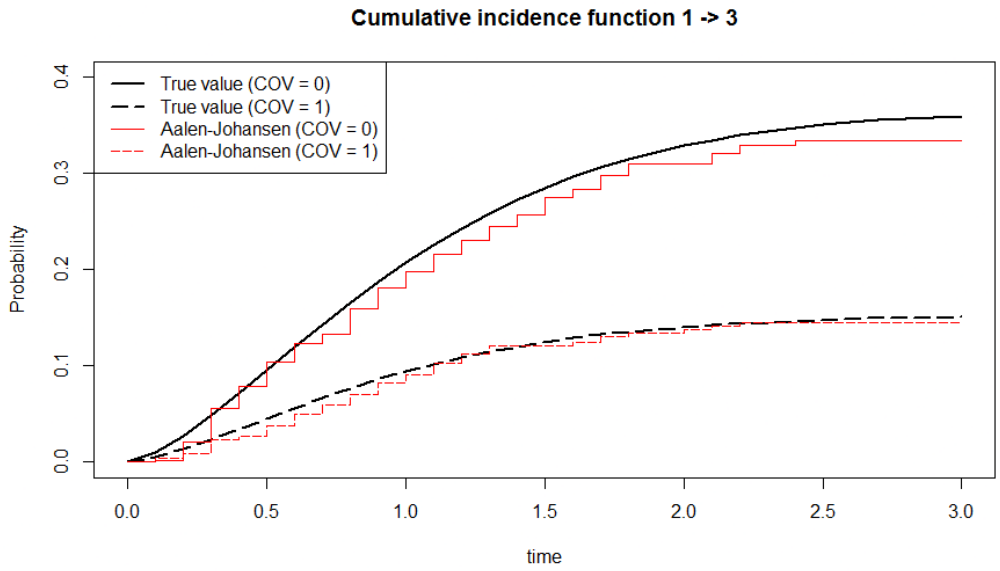
Visualising the transition Diagnosis (1) to Relapse (2)



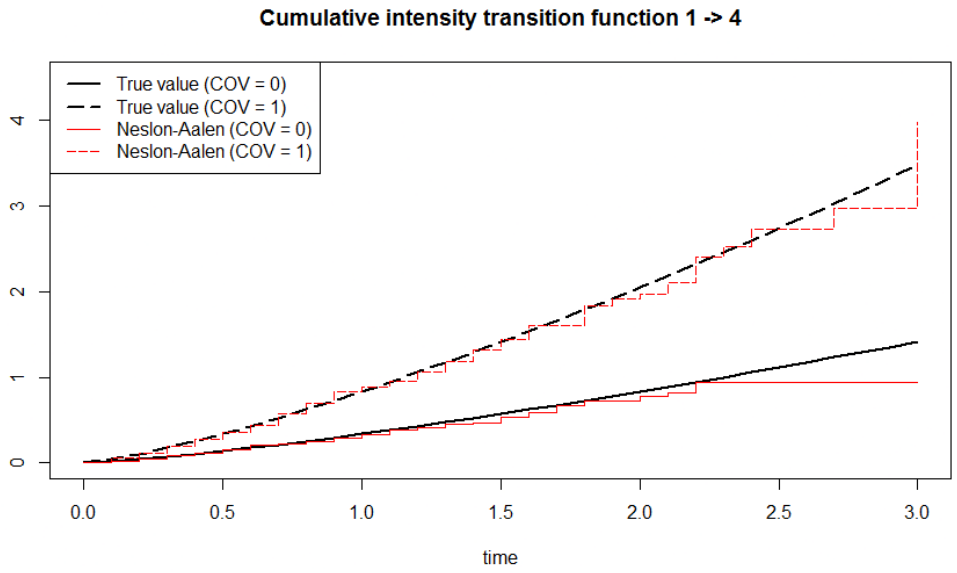


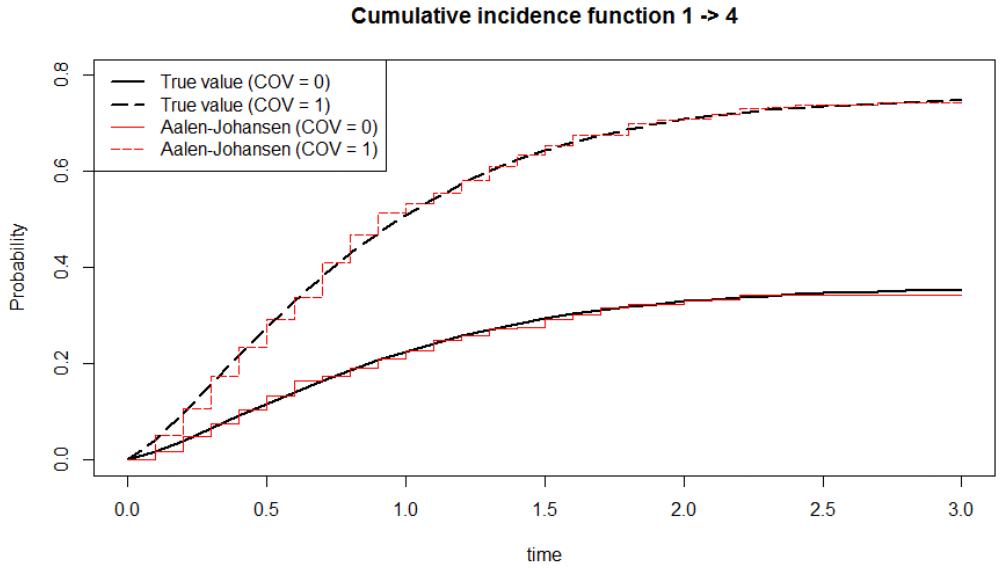
Visualising the transition Diagnosis (1) to Death (3)



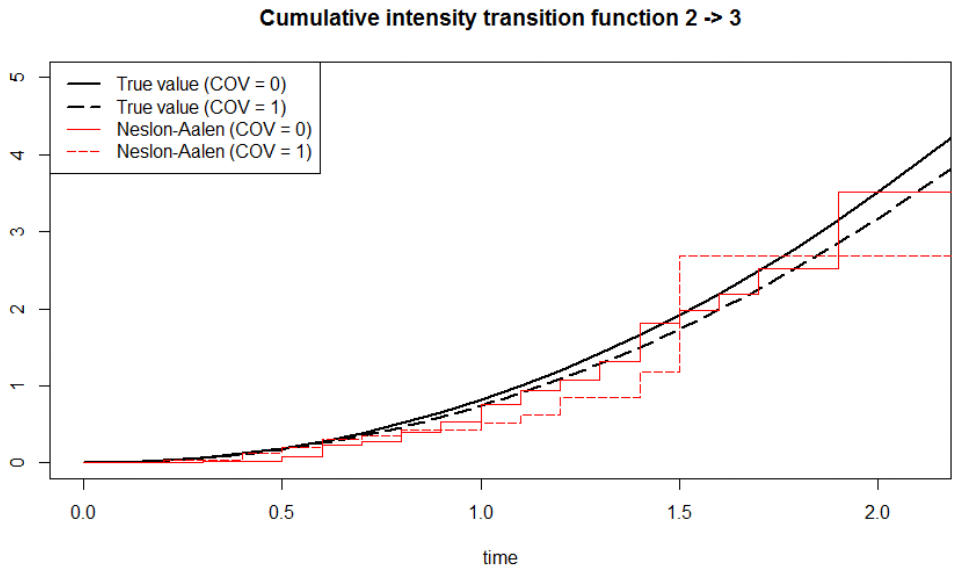


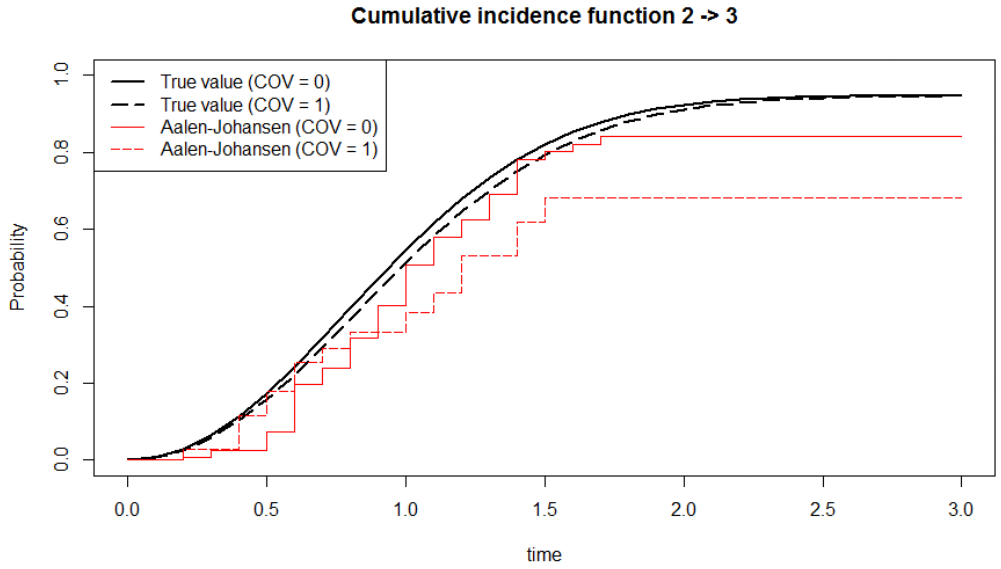
Visualising the transition Diagnosis (1) to Dropout (4)



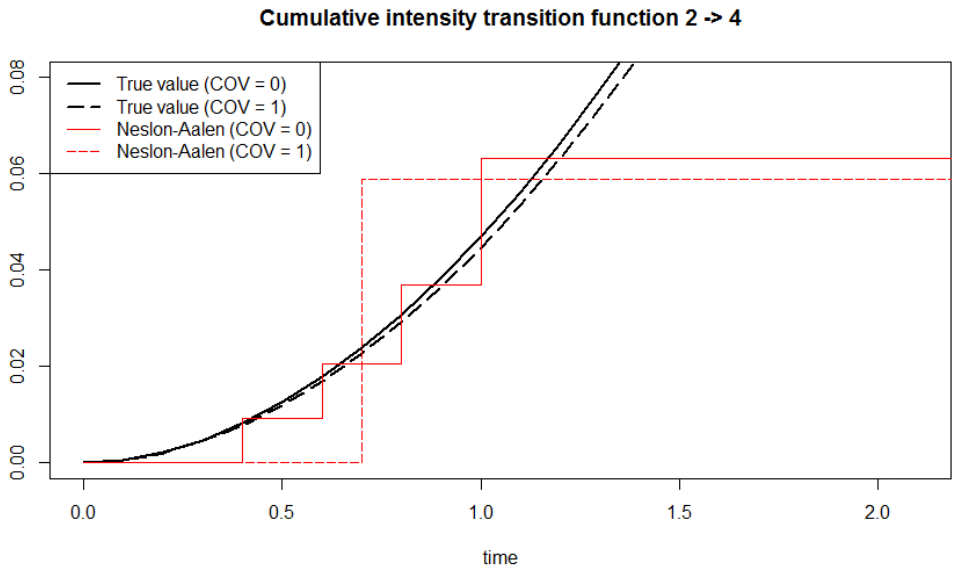


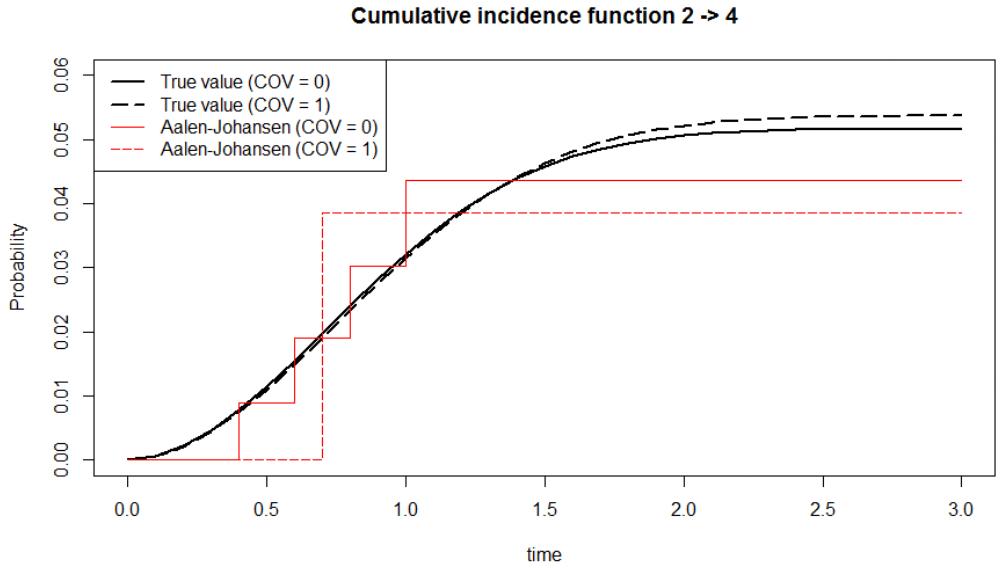
Visualising the transition Relapse (2) to Death (3)





Visualising the transition Relapse (2) to Dropout (4)





Chosen baseline model

Weibull-type intensity transition function:

$$\tilde{\alpha}_{ij}(t) = \frac{a_{ij}}{b_{ij}} \left(\frac{t}{b_{ij}} \right)^{a_{ij}-1}$$

Estimated parameter values after maximising likelihood (left), compared to true values (right). Details about the likelihood function are in the backup slides.

$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{23}$	$\hat{a}_{24}$
1.58	1.53	1.22	2.56	1.91
$\hat{b}_{12}$	$\hat{b}_{13}$	$\hat{b}_{14}$	$\hat{b}_{23}$	$\hat{b}_{24}$
3.06	2.55	1.58	1.15	5.19

a_{12}	a_{13}	a_{14}	a_{23}	a_{24}
1.40	1.50	1.30	2.10	1.90
b_{12}	b_{13}	b_{14}	b_{23}	b_{24}
2.68	2.13	2.30	1.10	5.00

Comparison of parametric models with and without covariate

Parameter estimates of model without covariates (left) and model with covariates (right):

$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{23}$	$\hat{a}_{24}$
1.58	1.53	1.22	2.56	1.91
$\hat{b}_{12}$	$\hat{b}_{13}$	$\hat{b}_{14}$	$\hat{b}_{23}$	$\hat{b}_{24}$
3.06	2.55	1.58	1.15	5.19

$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{23}$	$\hat{a}_{24}$
1.55	1.51	1.24	2.56	1.91
$\hat{b}_{12}$	$\hat{b}_{13}$	$\hat{b}_{14}$	$\hat{b}_{23}$	$\hat{b}_{24}$
2.55	2.23	2.42	1.15	5.16
$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{23}$	$\hat{\beta}_{24}$
-0.96	-0.60	0.97	-0.02	-0.06

$\left|\hat{\beta}_{ij}\right| \approx 0$ suggests insignificant effect on $\tilde{\alpha}_{ij}(t|Z) = \tilde{\alpha}_{ij,0}(t)\exp\left(\tilde{\beta}_{ij}Z\right)$.

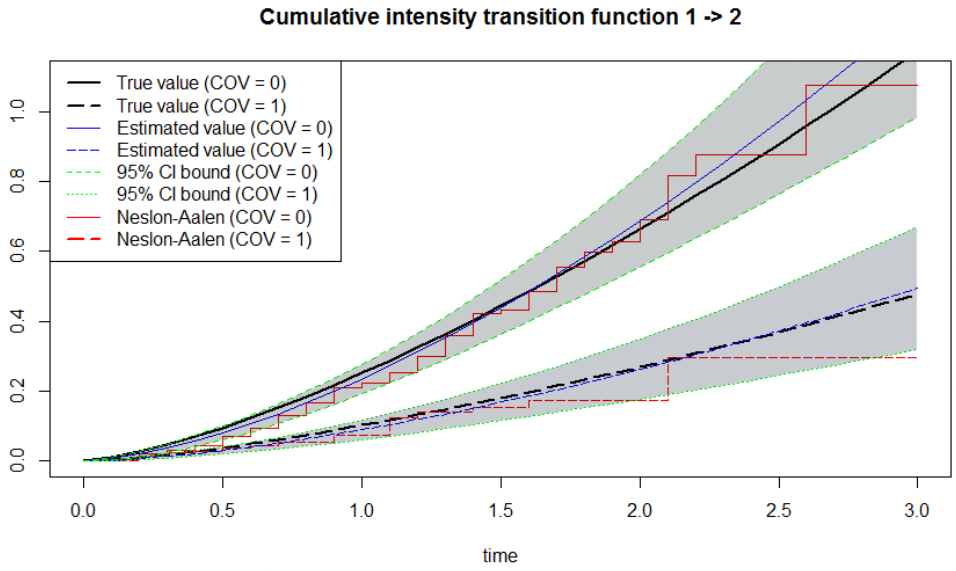
Positive (negative) value of $\hat{\beta}_{ij}$ suggests the risk increases (decreases) when $Z = 1$.

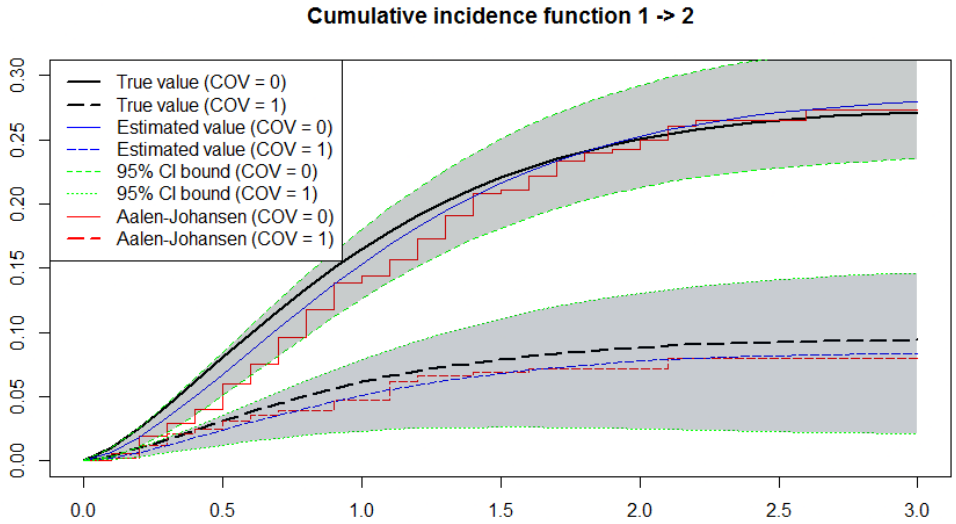
Comparison of parametric models with and without covariate (cont'd)

Comparing the two models overall,

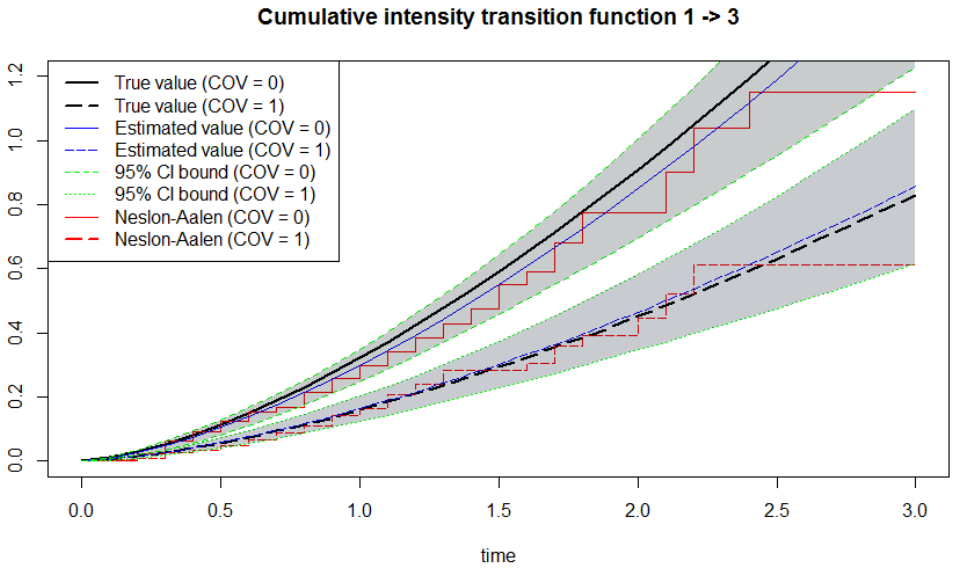
	Without covariate	With covariate
log likelihood	-1741.66	-1662.43
No. of parameters	10	15

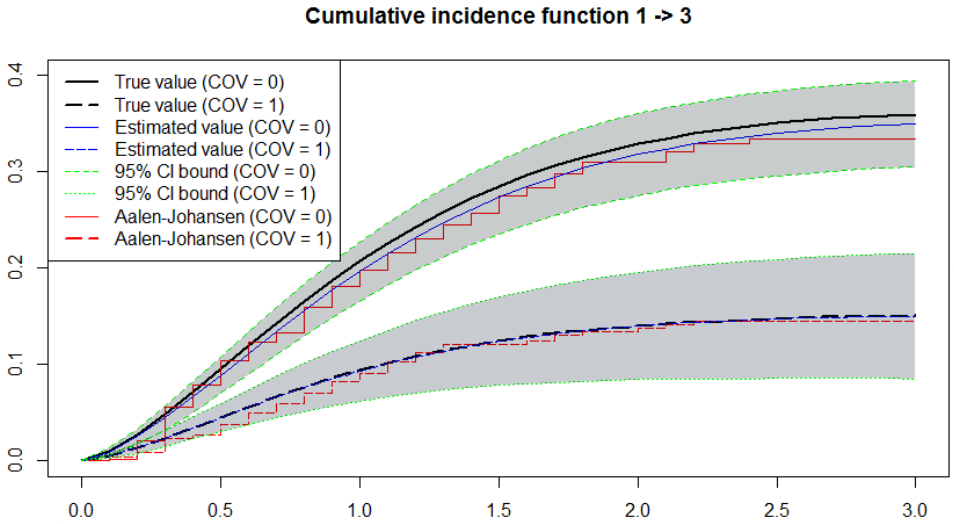
Assessing goodness of fit graphically for transition Diagnosis (1) to Relapse (2)



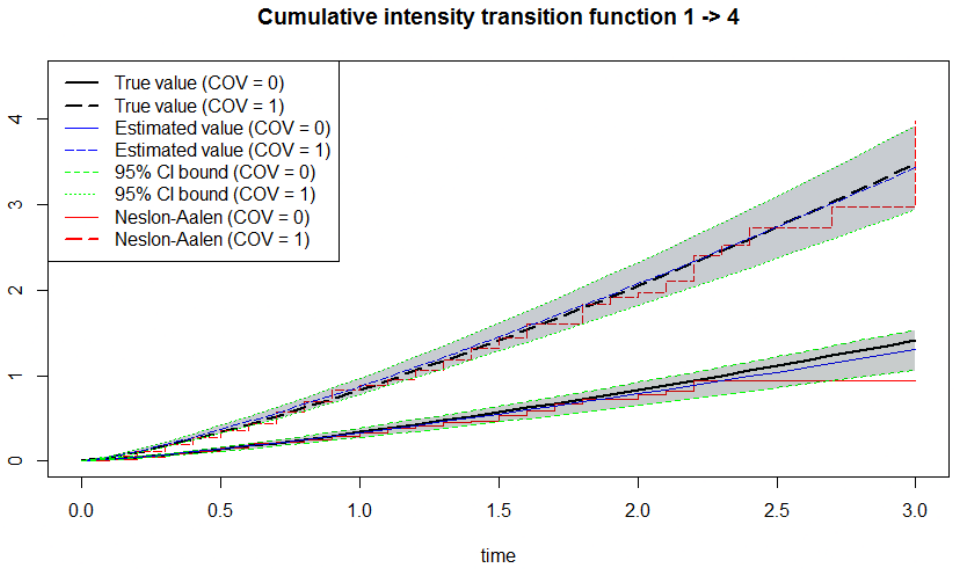


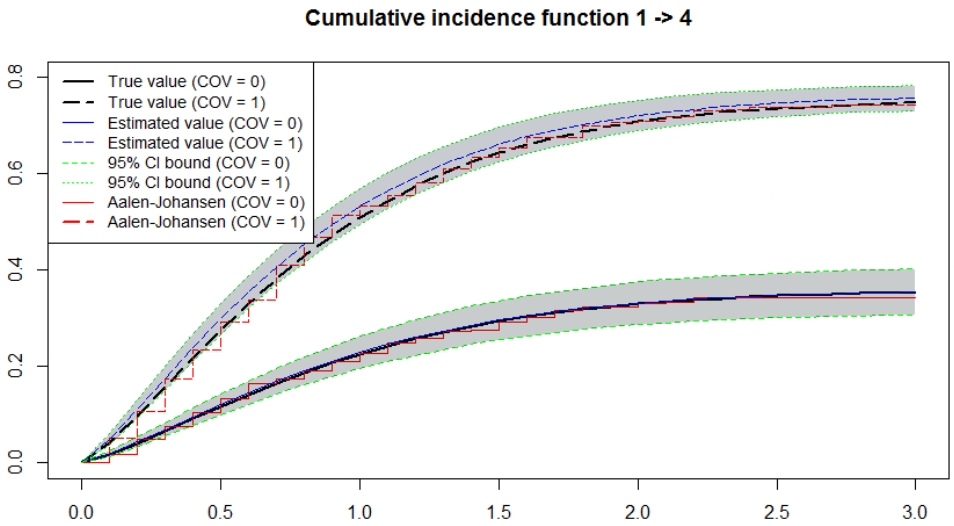
Assessing goodness of fit graphically for transition Diagnosis (1) to Death (3)



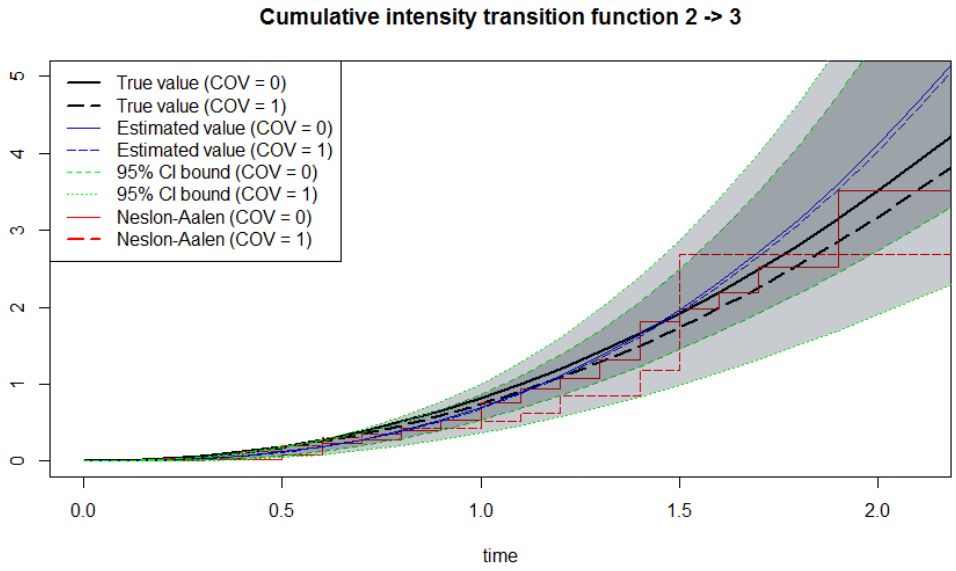


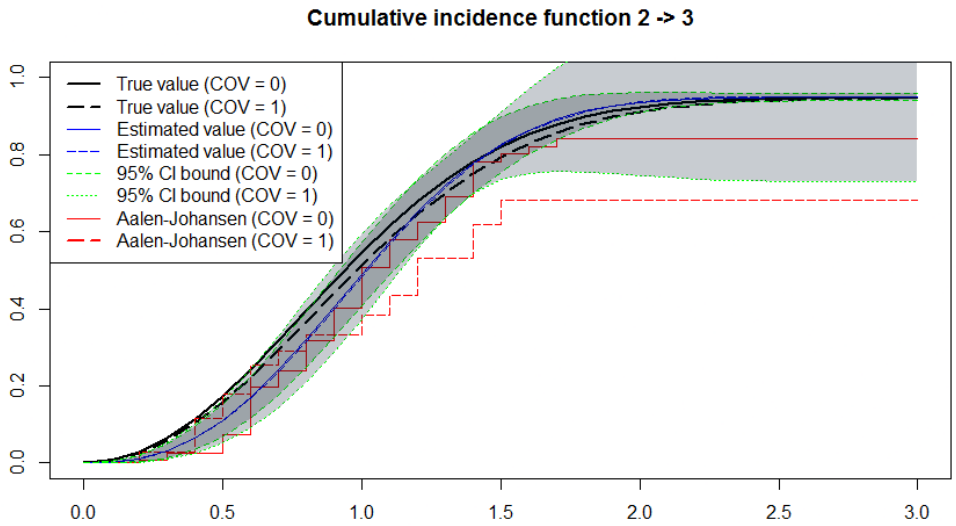
Assessing goodness of fit graphically for transition Diagnosis (1) to Dropout (4)



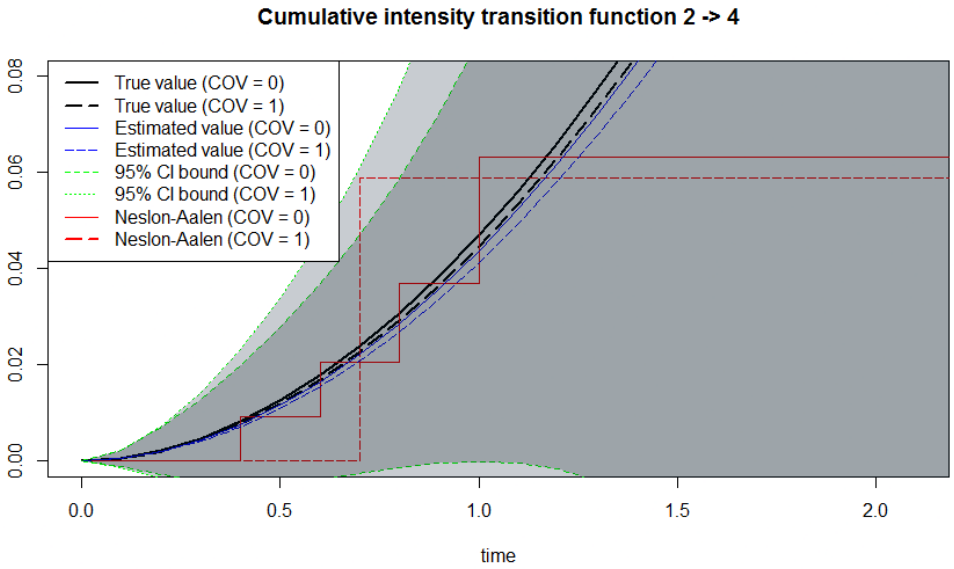


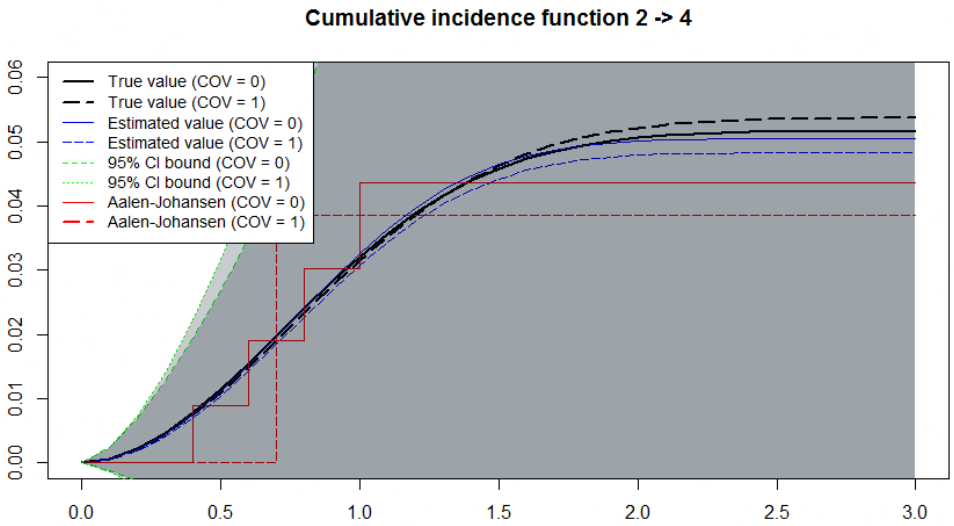
Assessing goodness of fit graphically for transition Relapse (2) to Death (3)





Assessing goodness of fit graphically for transition Relapse (2) to Dropout (4)





Hypothesis test to determine if patients on treatment have longer sojourn times before transitioning to undesirable states

Patients benefit from a drug if they have longer sojourn times before transitioning to undesirable states.

As per [Asanjarani et al., 2021], define the survival function of the holding time in state i, $S_i(t) = P(\tau_n > t | J_{n-1} = i)$. Writing in terms of ITFs,

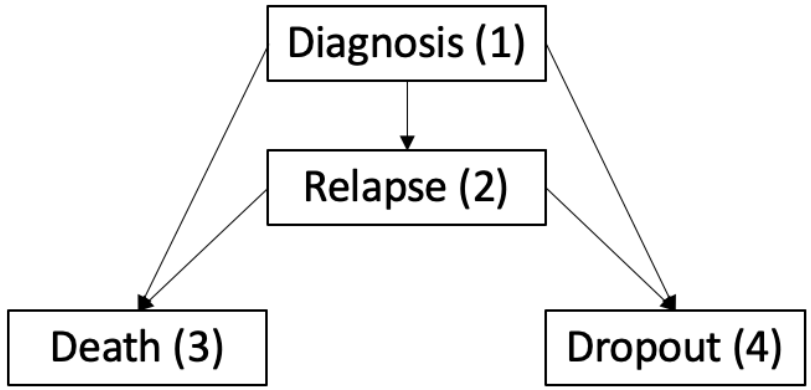
Survival function of holding time in state i

$$S_i(t) = \exp\Big(- \int_0^t \sum_{k \neq i} \tilde{\alpha}_{ik}(u) \mathrm{d}u \Big)$$

For those chosen model, in state 1,

$$S_1(t|Z) = \exp\left(- \sum_{k=2}^{4} \left(\frac{t}{b_{1k}} \right)^{a_{1k}} \exp(\beta_{1k} Z) \right)$$

Hypothesis test to determine if patients on treatment have longer sojourn times before transitioning to undesirable states (cont'd)



Suppose we wish to test the hypothesis

$$H_0 : S_1(t_0|Z = 1) = S_1(t_0|Z = 0) \text{ vs}$$
$$H_1 : S_1(t_0|Z = 1) < S_1(t_0|Z = 0)$$

for any fixed time point t_0. It can be shown that:

$$S_1(t_0|Z = 1) = S_1(t_0|Z = 0) \iff \sum_{k=2}^{4} \left(\frac{t_0}{b_{1k}}\right)^{a_{1k}} \Big(\exp(\beta_{1k}) - 1\Big) = 0$$

Hypothesis test to determine if patients on treatment have longer sojourn times before transitioning to undesirable states (cont'd)

Let $g(\theta) = \sum_{k=2}^{4} \left(\frac{t_0}{b_{1k}}\right)^{a_{1k}} \Big(\exp(\beta_{1k}) - 1 \Big)$. Here, θ is the vector of parameters in the full model.

We appeal to the *delta method* [Van der Vaart, 2000] (see backup slides for details). For chosen t_0 e.g. median observed sojourn time in state 1, the test statistic is

$$T = \left(\sum_{k=2}^{4} \left(\frac{t}{\hat{b}_{1k}}\right)^{\hat{a}_{1k}} \Big(\exp(\hat{\beta}_{1k}) - 1 \Big) \right) \Big/ \sqrt{\nabla g(\hat{\theta})^\top \hat{I}(\hat{\theta})^{-1} \nabla g(\hat{\theta})}$$
$$\sim N(0,1) \text{ under } H_0$$

where $\nabla g(\theta)$ denotes the vector of partial derivatives of $g(\theta)$. $I(\theta)$ is the Fisher information matrix, estimated by the negative of the Hessian matrix and evaluated at $\hat{\theta}$.

Hypothesis test to determine if patients on treatment have longer sojourn times before transitioning to undesirable states (cont'd)

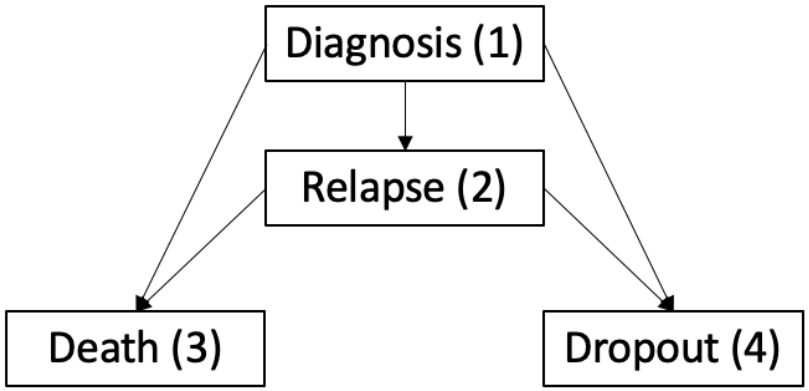
$$H_0 : S_1(t_0|Z = 1) = S_1(t_0|Z = 0) \text{ vs}$$
$$H_1 : S_1(t_0|Z = 1) < S_1(t_0|Z = 0)$$

H_0 is rejected if T is significantly greater than zero.

Based on the estimated parameters of the model, $T = 5.13$ and so H_0 is rejected in favour of H_1 at any reasonable level of significance (p-value 1.43×10^{-7}).

Hence, there is overwhelmingly strong evidence that patients on active treatment leave state 1 and transition to undesirable states significantly faster than patients not on active treatment. The patient is not benefiting!

Cox PH model vs fully parametric semi-Markov MSM



Fit Cox PH model on single transition (1) $\rightarrow$ (2):

$$h^C_{relapse}(t|Z) = h^C_{relapse,0}(t)exp(\beta Z)$$

We get $\hat{\beta} = -0.95$ with 95% conf. int. (-1.33 , -0.58) (p-value 4.85×10^{-7})
Estimated HR $= 0.38$ with 95% conf. int. (0.27 , 0.56)

Cox PH model vs fully parametric semi-Markov MSM (cont'd)

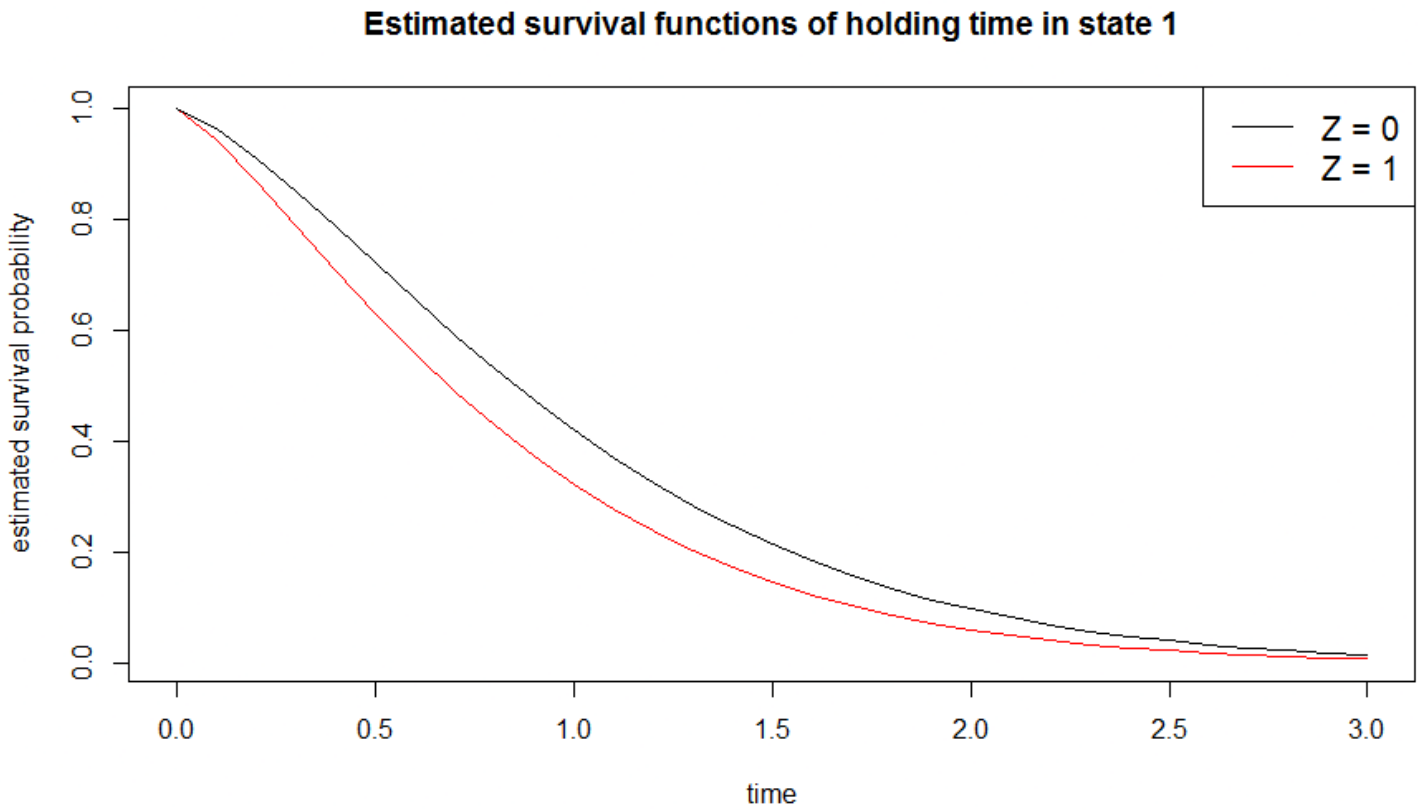
Fit parametric model on all transitions

$$\tilde{\alpha}_{ij}(t|Z) = \frac{a_{ij}}{b_{ij}} \left(\frac{t}{b_{ij}} \right)^{a_{ij}-1} \exp(\tilde{\beta}_{ij} Z)$$

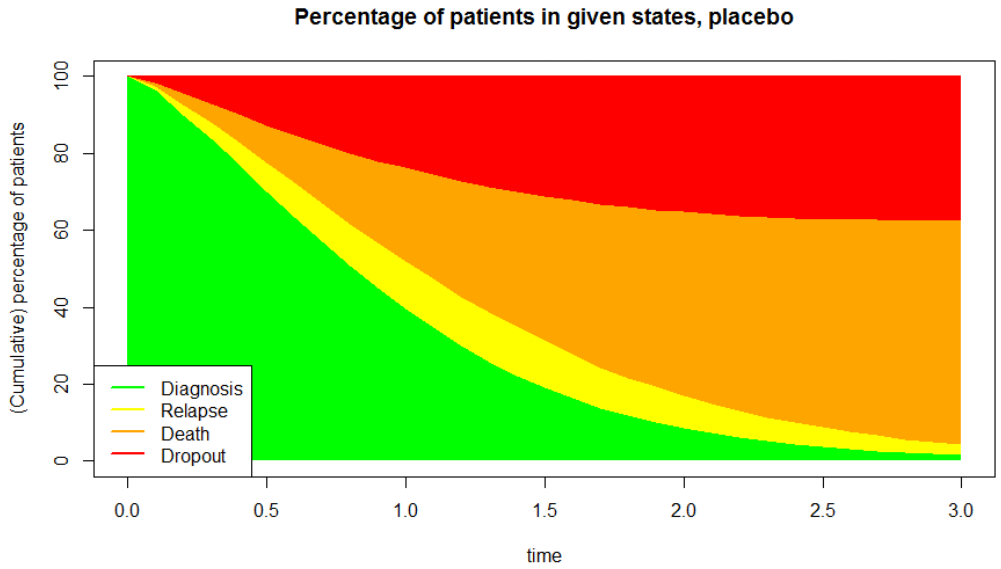
Comparing two groups' (Z=0 and Z=1) holding time in state 1 using estimated parameters:

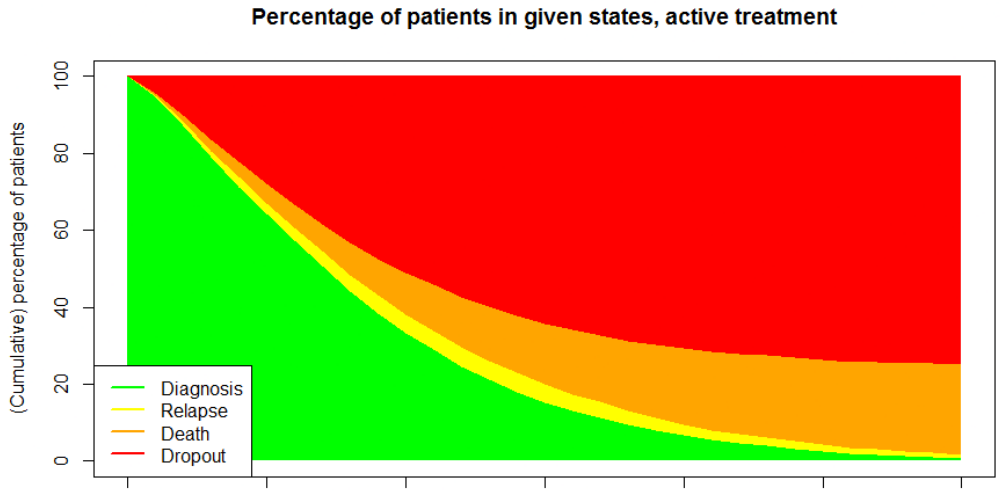
Cox PH model vs fully parametric semi-Markov MSM (cont'd)

$$\hat{S}_1(t|Z) = \exp\left(-\sum_{k=2}^{4} \left(\frac{t}{\hat{b}_{1k}} \right)^{\hat{a}_{1k}} \exp(\hat{\beta}_{1k} Z) \right)$$



Visualisation of potential patient benefit based on whether patient is on active treatment





Acknowledgements

This tutorial was made possible by the contributions of the following scientists:

Concept, idea and scope: Joachim Grevel

Scientific inquiry, translation into probability framework, execution of simulations and estimations, slide deck:
A. Haris Jameel (University of Nottingham)

Academic guidance, specifically towards semi-Markov processes, theory of hypothesis tests:
Christopher Brignell (University of Nottingham)
Christopher Fallaize (University of Nottingham)
Gilles Stupfler (L'École nationale de la statistique et de l'analyse de l'information (ENSAI))
Blesson Chacko (BAST Inc Ltd)

Nelson-Aalen estimator (backup)

Nelson-Aalen estimator [Nelson, 1969]

$$\hat{A}_{ij}(t) = \sum_{r:t_r \leq t} \frac{d_{ijr}}{n_{ijr}} \tag{5}$$

where d_{ijr} is number of $i \to j$ transitions at time t_r, n_{ijr} is number at risk of $i \to j$ transitions at time t_r.

Aalen-Johansen estimator (backup)

Aalen-Johansen estimator [Aalen and Johansen, 1978]

$$\hat{CIF}_{ij}(t) = \sum_{r:t_r \leq t} \frac{d_{ijr}}{n_{ijr}} \hat{S}_i(t_r) \tag{6}$$

$\hat{S}_i(t) = \prod_{q:t_q \leq t} \left(1 - \frac{d_q}{n_q}\right)$ is the Kaplan-Meier estimator [Kaplan and Meier, 1958] for the survival function of the holding time at state i. d_q is number of transitions out of state i at time t and n_q is number at risk of any transition out of out state i at time t.

Estimating semi-Markov MSM by optimising likelihood (backup)

As per [Asanjarani et al., 2021], the likelihood each individual h can be written as

$$\begin{aligned}\mathcal{L}^{(h)} &= \left\{ \prod_{k=1}^{N^{(h)}} \tilde{\alpha}_{J_{k-1}^{(h)} J_k^{(h)}}(\tau_k) S_{J_{k-1}^{(h)}}(\tau_k) \right\} \left\{ S_{J_{N^{(h)}}^{(h)}}(U^{(h)}) \right\}^{1-\delta^{(h)}} \\ &= \left\{ \prod_{k=1}^{N^{(h)}} \tilde{\alpha}_{J_{k-1}^{(h)} J_k^{(h)}}(\tau_k) \exp\Big(- \int_0^{\tau_{J_k^{(h)}}} \tilde{\alpha}_{J_{k-1}^{(h)}}(u) \mathrm{d}u \Big) \right\} \times \\ &\quad \left\{ \exp\Big(- \int_0^{U^{(h)}} \tilde{\alpha}_{J_{N^{(h)}}^{(h)}}(u) \mathrm{d}u \Big) \right\}^{1-\delta^{(h)}}\end{aligned}$$

where $\tilde{\alpha}_i(t) = \sum_{k \neq i} \tilde{\alpha}_{ik}(t)$.

Individuals are assumed independent so the full likelihood for m patients is $\mathcal{L} = \prod_{h=1}^{m} \mathcal{L}^{(h)}$.

Underlying assumptions about asymptotic distribution of maximum likelihood estimator (backup)

It has been assumed the vector of maximum likelihood estimators $\hat{\theta}$ of parameter vector θ is such that $(\hat{\theta} - \theta) \sim N_p(0, I(\theta)^{-1})$ asymptotically, where $I(\theta)$ is the Fisher information matrix associated with the entire sample. The reason for this assumption is that we have (conditionally) independent and non-identically distributed data and so the usual desirable asymptotic properties of maximum likelihood estimators may not apply.

Justifying or proving the distributional assumption (and hence the validity of confidence intervals and hypothesis test as per the next two slides) is a work in progress, but initial results based on simulations suggest evidence of normality when the number of each transition is sufficiently large.

Details about confidence intervals and hypothesis test for difference in sojourn times before transitioning to undesirable states (backup)

Assuming $(\hat{\theta} - \theta) \sim N_p(0, I(\theta)^{-1})$ asymptotically as detailed in the previous slide, we can use the *delta method* to deduce that for a function $h(\theta)$ with (non-zero) gradient vector $\nabla h(\theta)$,
$(h(\hat{\theta}) - h(\theta)) \sim N(0, \nabla h(\theta)^\top I(\theta)^{-1} \nabla h(\theta))$ asymptotically.

Since the cumulative intensity transition functions and cumulative incidence functions can be written as functions of θ, it is possible to compute (asymptotic) confidence intervals at fixed time points.

Details about confidence intervals and hypothesis test for difference in sojourn times before transitioning to undesirable states (cont'd) (backup)

The general form of the test statistic used for the hypothesis test $H_0 : S_i(t_0|Z=1) = S_i(t_0|Z=0)$ vs $H_1 : S_i(t_0|Z=1) < S_i(t_0|Z=0)$ is

$$T = \frac{g(\hat{\theta})}{\sqrt{\nabla g(\hat{\theta})^\top \hat{I}(\hat{\theta})^{-1} \nabla g(\hat{\theta})}} \sim N(0,1) \text{ asymptotically under } H_0$$

where $g(\theta) = \int_0^{t_0} \sum_{k \neq i} \tilde{\alpha}_{ik,0}(u) \Big(\exp(\beta_{ik}) - 1 \Big) \mathrm{d}u$. The Fisher information matrix is estimated by the observed Fisher information matrix evaluated at $\hat{\theta}$.

H_0 is rejected if T is significantly greater than zero, unless the direction of the inequality in the alternative hypothesis is reversed. In such a case, H_0 is rejected if T is significantly less than zero.


Aalen, O. O. and Johansen, S. (1978).
An empirical transition matrix for non-homogeneous markov chains based on censored observations.
Scandinavian Journal of Statistics, pages 141–150.


Asanjarani, A., Liquet, B., and Nazarathy, Y. (2021).
Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches.
The International Journal of Biostatistics.


Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009).
Simulating competing risks data in survival analysis.
Statistics in medicine, 28(6):956–971.


Kaplan, E. L. and Meier, P. (1958).
Nonparametric estimation from incomplete observations.
Journal of the American statistical association, 53(282):457–481.


Nelson, W. (1969).
Hazard plotting for incomplete failure data.

Journal of Quality Technology, 1(1):27–52.

Van der Vaart, A. W. (2000).
Asymptotic statistics, volume 3.
Cambridge university press.