Analysis of co-time-to-event outcomes in randomized clinical trials

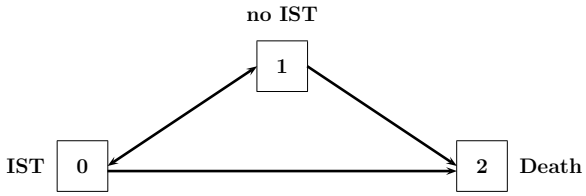
Jan Beyersmann

Joint work with Claudia Schmoor, Tobias Bluhmki, Harriet Sommer, Arthur Allignol...

Institute of Statistics, University of Ulm (JB, TB)

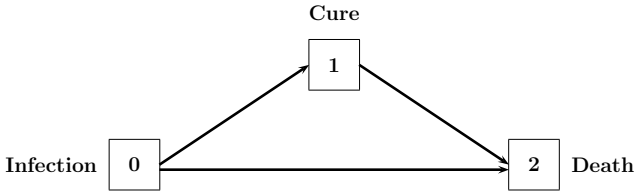
- Two data examples where neither Kaplan-Meier nor competing risks do the job.
- Some event history/multistate basics
- Comparing non-standard time-to-event outcomes using non-standard resampling and confidence bands

Example: immunosuppressive treatment (IST)



- Effect of graft-versus-host-disease (GvHD) prophylaxis on **probability to be alive without IST** in leukemia patients after allogeneic haematopoietic cell transplantation
- IST may be switched off and on a random number of times.
- Proportion of patients alive without IST goes up and down: neither Kaplan-Meier nor cumulative incidence functions apply.
- # arrows = # hazards
- Schmoor et al, Clinical Cancer Research 2013; Bluhmki et al, Biometrics 2018

Example: Antibiotic treatment of hospital infection



- Effect of Ceftobiprole to combat Gram-positive bacteria, outcome **probability to be cured and alive** in patients with hospital-acquired pneumonia
- Conflicting recommendations on outcome from EMA (cure) and FDA (28-day-mortality)
- The aim of the treatment is cure, but death shortly after cure happens.
- Sommer et al, Antimicrobial Agents and Chemotherapy 2018

Why hazards?

- Outcome: Jan's death
- Observation process: the audience is looking at me.
- I'm **at risk**: alive and under observation.
- Independent censoring: your presence
  - does not scare me too much (which might increase my hazard)
  - does not please me too much (which might decrease my hazard)
- If I die right now,
  - it'll happen with the same hazard as without you looking,
  - you'll observe it.
- So you can **estimate** my hazard (based on 100 Jans...)

This has little to do with me dying. . .

- Outcome: Jan breaks his right arm, too. (A possibly recurrent event in the presence of competing risk 'death'.)
- Observation process: the audience is looking at me.
- I'm at risk: alive **with right arm** and under observation.
- Independent censoring: your presence
  - does not scare me too much (which might **change** my arm-breaking hazard)
  - does not please me too much (which might **change** my arm-breaking hazard)
- If I break my right arm right now,
  - it'll happen with the same hazard as without you looking,
  - you'll observe it.
- So you can **estimate** it.

More formally

- $N(t)$ counts no. of observed events **of some type**
- Independent censoring **process** (light turned on/off/on/off/on. . .) such that

$$P(\mathrm{d}N(t) = 1 \,|\, \text{Past}) = Y(t)\alpha(t)\mathrm{d}t, \text{ where}$$

  - $\alpha(t)$ is as in the uncensored case, e.g.,

$$\lim_{\Delta t \searrow 0} P(\text{break arm in } [t, t+\Delta t) \,|\, \text{alive with arm at } t-)/\Delta t$$

  - $Y(t)$ the number of units that may experience the event at t and are under observation at $t-$
- OK, if censoring is entirely random, if one leaves the risk set because of a competing risk, or if. . .

Event-driven trials

- Toy example: 2 patients put on trial at the same time, stop after 1 observed event.
- The data

$$T_1 \wedge T_2, \quad \mathbf{1}(T_1 \leq T_2)$$
$$T_1 \wedge T_2, \quad \mathbf{1}(T_2 \leq T_1)$$

not independent

- E.g., Efron's bootstrap requires independence, because it samples with replacement from the patients.
- General counting process/martingale/wild bootstrap machinery to follow does not.

A martingale estimating equation

- $\mathrm{d}N(t)$ no. observed events, $Y(t)$ no. at risk, target quantity $\alpha(t)$,

$$P(\mathrm{d}N(t) = 1 \,|\, \text{Past}) = Y(t)\alpha(t)\mathrm{d}t,$$

  or

$$\mathrm{d}N(t) - Y(t)\alpha(t)\mathrm{d}t = \mathrm{d}M(t),$$

  with $\mathrm{E}(\mathrm{d}M(t) \,|\, \text{Past}) = 0$ and M is a martingale/error process.
- Nelson-Aalen

$$\int_{(0,t]} \frac{1}{Y(u)} \mathrm{d}N(u) = \sum_{u \leq t} \frac{\Delta N(u)}{Y(u)} \to A(t) = \int_{(0,t]} \alpha(u)\mathrm{d}(u)$$

  with error process (martingale)

$$\int_{(0,t]} \frac{1}{Y(u)} \mathrm{d}M(u),$$

  mean zero, approximately normal.

Wild bootstrapping Nelson-Aalen (Dobler et al., LiDA, to appear)

- We do not know the precise value of

$$\mathrm{d}M(t) = \sum_{i=1}^{n} \left(\mathrm{d}N_i(t) - Y_i(t)\alpha(t)\mathrm{d}t\right),$$

but we approximate its distribution using

$$\mathrm{d}\hat{M}(t) = \sum_{i=1}^{n} \mathrm{d}N_i(t) \cdot \underbrace{G_i(t)}_{\sim N(0,1)},$$

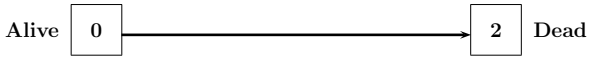
mean zero, normal, approximately right (co-) variance structure.

- Generate large number of $N(0,1)$-multipliers.
- We'll do this for every multistate transition hazard

$$\alpha_{lj}(t)\mathrm{d}t = P(\text{in state } j \text{ at time } t + \mathrm{d}t \,|\, \text{in state } l \text{ at time } t-).$$

- The trick now is switching to probabilities (where bootstrapping is really worthwhile)...

From hazards to probabilities



- A two-state model, a 2×2-matrix with unit matrix $\mathbf{I}$,

$$\mathbf{I} - \hat{\mathbf{A}}(t) = \begin{pmatrix} 1 - \frac{\Delta N(u)}{Y(u)} & \frac{\Delta N(u)}{Y(u)} \\ 0 & 1 \end{pmatrix}$$

where the sum over each row is one and
  - top left: known from Kaplan-Meier
  - bottom left: no rebirth
  - bottom right: once dead, stay dead
- Aalen-Johansen ($\hat{=}$ Kaplan-Meier here)

$$\prod_{u \le t} \left(\mathbf{I} - \hat{\mathbf{A}}(u)\right) = \begin{pmatrix} \text{KM} & 1 - \text{KM} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \hat{P}_{00}(0,t) & \hat{P}_{01}(0,t) \\ \hat{P}_{10}(0,t) & \hat{P}_{11}(0,t) \end{pmatrix}$$

Aalen-Johansen and Wild Bootstrap

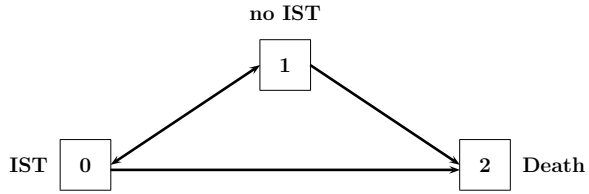
$$\hat{\mathbf{P}}(s,t) = \prod_{u \in (s,t]} \left(\mathbf{I} - \Delta\hat{\mathbf{A}}(u)\right)$$

with entries $\hat{P}_{lj}(s,t)$ and, for $l \neq j$,

$$\Delta\hat{\mathbf{A}}_{lj}(u) = \frac{\text{no. of observed } l \to j \text{ transitions at } u}{\text{at risk in state } l \text{ at } u-}$$

- Aalen-Johansen Hadamard-differentiable mapping (product integration) of multivariate Nelson-Aalen on cadlag function space $D[0,\tau]^{(K+1)\times K}$.
- Wild Bootstrap, if model is time-inhomogeneous Markov (Bluhmki et al. 2018): Transform Wild Bootstrap for Nelson-Aalen according to Hadamard-derivative (a functional delta method argument).
- Wild bootstrap converges in distribution in probability to the right limit, given the data, and allows to construct **simultaneous confidence bands** based on supremum statistics.

Example: immunosuppressive treatment (IST)



- Grafalon aka ATG-Fresenius aka ATLG $n = 103$, control $n = 98$.
- ATG-F decreased GvHD hazard, no effect on death w/o prior GvHD hazard (Finke et al Lancet Onco 2009, Schmoor et al Clin Canc Res 2013).
- ATG-F increased no-IST hazard (HR $_{1.41}\,2.02_{2.91}$) and decreased IST-hazard (HR $_{0.18}\,0.31_{0.55}$), no harmful effect on death. (Socié et al Blood 2011, Schmoor et al Clin Canc Res 2013).
- Probabilities?
- Following results based on 'empirical artificial' example (see later).

Aalen-Johansen estimators of $P_{0j}(0, t)$

(A) ATG-F: no harmful effect on survival.

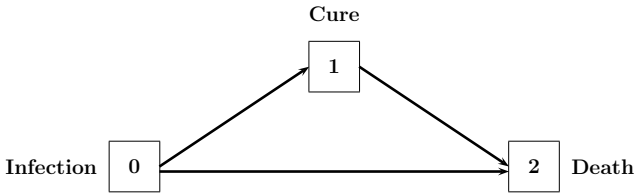
(B) ATG-F: smaller probability of 'alive with IST'.

(B) ATG-F: higher probability of 'alive without IST'.

Confidence bands for ATG-F minus control

- Left: alive with IST. Right: alive without IST.
- Null effect is the dashed **line**. (Bluhmki et al., Biometrics 2018)

Example: Antibiotic treatment of hospital infection

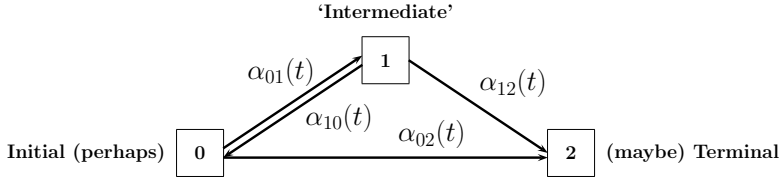


- Outcome: clinical cure at test of cure (ToC) for patients with hospital-acquired pneumonia
- Ceftobiprole was non-inferior compared to Ceftazidime/Linezolid at 15% margin, difference in proportions $_{-10.0\%}$ 49.9% – 52.8%$_{4.1\%}$, differential results if pneumonia was ventilator-associated (Awad et al., Clinical Infectious Diseases 2014).
- Aim now: account for vital status after cure and varying follow-up times, aim at stronger (time-simultaneous) non-inferiority statement

From Sommer et al., Antimicrobial Agents and Chemotherapy 2018

- Lower panel: difference of the P(alive & cured)'s with one-sided simultaneous confidence band

Simulations. But how to simulate?



- Start in 0, simulate waiting time from

$$t \mapsto 1 - \exp\left(-\int_0^t \alpha_{01}(u) + \alpha_{02}(u)\mathrm{d}u\right)$$

- Leave 0 at simulated time t_0 towards 1 with probability

$$\alpha_{01}(t_0)/\left(\alpha_{01}(t_0) + \alpha_{02}(t_0)\right)$$

- Say, trajectory moves into 1 at time t_0: Simulate waiting time in 1 from

$$t \mapsto 1 - \exp\left(-\int_{t_0}^t \alpha_{10}(u) + \alpha_{12}(u)\mathrm{d}u\right) \text{ etc..}$$

- Any latent 'time-to-progression' here? (Joint work w. M Meller, K Rufibach.)

How to simulate realistic data?

- Start in 0, simulate waiting time from

$$t \mapsto 1 - \exp\left(-\int_0^t \alpha_{01}(u) + \alpha_{02}(u)\mathrm{d}u\right)$$

- Leave 0 at simulated time t_0 towards 1 with probability

$$\alpha_{01}(t_0)/\left(\alpha_{01}(t_0) + \alpha_{02}(t_0)\right)$$

- Say, trajectory moves into 1 at time t_0: Simulate waiting time in 1 from

$$t \mapsto 1 - \exp\left(-\int_{t_0}^t \alpha_{10}(u) + \alpha_{12}(u)\mathrm{d}u\right) \text{ etc..}$$

- Simply replace $\alpha_{lj}(u)\mathrm{d}u$ by $\Delta\hat{A}_{lj}(u)$ above! (Bluhmki, Putter, et al, submitted.)
- May be based on published data only, e.g., for planning, and gives yet another bootstrap procedure — possibly without IPD!
- See Allignol et al. (BMC Med Res Meth 2011) and Ohneberg et al. (Stat Med 2017) for the competing risks case.

Simulations: wild bootstrap compared to 'empirical simulation'.

- OK, but slightly too low coverage for $n = 103$ as in first data example.
- Real data findings confirmed for level 0.025.

Non-Markov

- So far

$$\begin{aligned}\alpha_{lj}(t)\mathrm{d}t &= P(\text{in state } j \text{ at time } t + \mathrm{d}t \mid \text{in state } l \text{ at time } t-, \textbf{Past}) \\ &\overset{!}{=} P(\text{in state } j \text{ at time } t + \mathrm{d}t \mid \text{in state } l \text{ at time } t-)\end{aligned}$$

- What if $1 \to 2$ depends on t, state 1 **and** arrival time in 1?
- Recall Δ Nelson-Aalen

$$\frac{\text{no. observed } l \to j \text{ transitions at } t}{\text{no. under observation in } l \text{ at } t-}$$

estimates **partly conditional** transition rate (popular with recurrent events)

$$\begin{aligned}\tilde{\alpha}_{lj}(t)\mathrm{d}t &= P(\text{in state } j \text{ at time } t + \mathrm{d}t \mid \text{in state } l \text{ at time } t-) \\ &\neq P(\text{in state } j \text{ at time } t + \mathrm{d}t \mid \text{in state } l \text{ at time } t-, \textbf{Past}) \\ &\quad \text{in general,}\end{aligned}$$

and provided that **censoring** is **entirely random**. (Heuristics clear, proof less so.)

State occupation probabilities $P(X_t = j)$?

- No censoring:
  - Kaplan-Meier is survivor proportion.
  - Aalen-Johansen is state occupation proportion, even if non-Markov.
- Random censoring: Aalen-Johansen
  - always estimates $P(X_t = j)$, **if** Nelson-Aalen estimates cumulative partly conditional transition rates.
  - still estimates $P(X_t = j \mid X_s = l)$ using landmarking (Allignol et al., LiDA 2014, Putter & Spitoni SMMR 2016)
- Proof
  - in Datta and Satten (Stat Prob Lett 2001) for *non-Markov case under Markov assumption*
  - repaired and more general treatment in Müller et al. (submitted revision 2018)

From Müller et al., submitted revision 2018

- Aalen-Johansen estimator of $P(X(t) = 1)$ in non-Markov illness-death model without recovery. Simulations as before.
- E.g., Efron's bootstrap works.
- Event-driven trials?

Bootstrapping time-to-event data

- Drawing with replacement: relies on i.i.d. structure. But event history analysis can do much more, e.g.
  - event-driven trials,
  - nested case-control designs, e.g., if covariates are expensive.
- **Wild** bootstrap: you may bootstrap based on one unit only,

$$\mathrm{d}\hat{M}(t) = \sum_{i=1}^{n} \mathrm{d}N_i(t) \cdot \underbrace{G_i(t)}_{\sim N(0,1)}, \text{ even if } n = 1 \text{ (not recommended).}$$

- If you want to be **weird**, not **wild**: weird bootstrap (Andersen, Borgan, Gill, Keiding, *Statistical Models Based on Counting Processes*, 1993; Dobler et al, Biometrika 2017). Sample events at the observed event times using

$$B(Y(t), \Delta\hat{A}(t))$$

  which is weird, because you might die twice...
- Or just simulate analogously to using a published Kaplan-Meier curve.

Do we need all the stochastic process theory?

- Counting processes, martingales, Hadamard-differentiability on cadlag function spaces, ...? Yes!
- Standard textbook tale: survival data are censored, let's do Kaplan-Meier,

$$1 - \frac{\text{no. observed deaths at } t}{\text{no. of observed survivors at } t-},$$

  multiplying over all observed event times t.
- This conditions on the observed event times! (And on no. at risk, too.)
- So, just run a very expensive study, where you
  - follow-up every patient (no censoring)
  - with very frequent visits (no ties).

  This is very cheap! You can compute Kaplan-Meier without collecting the data! (Conditional on the actual event times...).

Discussion

- Multistate models for more compex time-to-event endpoints
  - account for ‘co-information’ like ‘no immunosuppressive treatment **and** alive’
  - Cox analyses as in Claudia’s talk (see the Andersen Gill 1982 Ann Stat paper)
  - probabilities may go up and down and are estimated by Aalen-Johansen
- Estimation of probabilities ‘an old story’, but direct comparisons a bit more challenging (perhaps):
  - this talk: confidence bands based on resampling
  - Andersen and colleagues: pseudo-values
- Aalen-Johansen ‘an old story’, but recent developments include resampling, non-Markov models, stabilization (Friedrich et al., Ann App Stat 2017), . . .

Some own references:

- Bluhmki, T., Schmoor, C., Dobler, D., Pauly, M., Finke, J., Schumacher, M., and Beyersmann, J. (2018). A wild bootstrap approach for the Aalen-Johansen estimator. *Biometrics* (early view)
- Sommer, H., Bluhmki, T., Beyersmann, J., Schumacher, M., et al. (2018). Assessing noninferiority in treatment trials for severe infectious diseases: an extension to the entire follow-up period using a cure-death multistate model. *Antimicrobial Agents and Chemotherapy*, 62(1):e01691–17.
- Allignol, A., Schumacher, M., Wanner, C., Drechsler, C., and Beyersmann, J. (2011). Understanding competing risks: a simulation point of view. *BMC Medical Research Methodology*, 11:86.
- Bluhmki, T., Putter, H., Allignol, A., and Beyersmann, J. (2018). Bootstrapping complex time-to-event data without patient individual data, with a view towards time-dependent exposures. *Submitted*.
- Dobler, D., Beyersmann, J., and Pauly, M. (2017). Non-strange weird resampling for complex survival data. *Biometrika*, 104(3):699–711.
- Müller, C., Allignol, A., and Beyersmann, J. (2018). Estimating state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring *Revised*.
- Beyersmann, J., Allignol, A., and Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. Springer, New York.