Adaptive designs for multiple time-to-event outcomes in Markovian multi-state models-

Moritz Fabian Danzer[1] Andreas Faldum[1] Rene Schmidt[1]

[1]Institute of Biostatistics and Clinical Research, University of Münster, Germany

20/04/2023

Project

Results of the DFG-funded project "Design and analysis of adaptive-sequential clinical trials with multiple, correlated time-to-event endpoints"

- Project leader: Rene Schmidt
- Location: Institute of Biostatistics and Clinical Research, University of Münster
- Period: 04/2019 - 03/2022 (extended until 09/2025)
- Grant number: 413730122

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Aims and scope

Several correlated time-to-event endpoints of interest given

- Simultaneous assessment of efficacy and toxicity

Enable adaptive design changes which can be based on all of these endpoints

$\Rightarrow$ Extending one-sample methods from Danzer et al. (2022) to a randomized, multi-arm setting

H_0: Joint distribution of selected time-to-event endpoints in different groups is the same

What's new?

Several time-to-event endpoints may be used to determine interim design changes!

General concern (cf. Bauer and Posch (2004)):

Consideration of surrogate (non-primary) endpoint in interim design changes may inflate type I error!

Existing solutions (patient-wise separation) lead to

- discarding of information or
- inevitable worst-case adjustment

Related methods

H_0: (Marginal) distribution of all selected time-to-event endpoints in different groups is the same

Tests for multiple time-to-event endpoints in Wei and Lachin (1984):

- Simultaneous log-rank tests for all variables
- Covariance matrix via multivariate CLT

Group-sequential extension in Lin (1991):

- Extended use of multivariate CLT for multiple time points
- component-wise "independent increments structure"

Problem: increments are not independent across components

Compromising the type I error level

Scenario similar to example in Bauer and Posch (2004) to exploit method from Lin (1991) in adaptive design:

- Two time-to-event endpoints
- Second endpoint occurs exactly t years after occurence of first endpoint
- Exact prediction of second component t years in advance possible
- Adaptation at interim analysis:
  - Stop recruitment if forecasted component will exceed decision bound
  - Else increase sample size

Challenge: Account for "shared information"

Notational conventions

Patients $i \in \{1, \ldots, n\}$

Treatment indicators $Z_i \in \{0, 1\}$

States resp. (component) events $j \in \{1, \ldots, k\}$

Composite events $E \subset \{1, \ldots, k\}$

Event resp. hitting times T_i^E

Counting processes N_i^E

Occupation indicators $Y_i^j \in \{0, 1\}$

Occupation counts $Y^j = \sum_i Y_i^j$ and $Y^{Z=1,j} = \sum_i Z_i Y_i^j$

Two notions of time:

- calendar time t
- time in trial s

From one- to two-sample methods

Patient-wise martingale

$$M_i^j(s) = N_i^j(s) - \int_0^s \lambda^j(u|\mathcal{F}_i(u))du$$

as basis for one-sample test

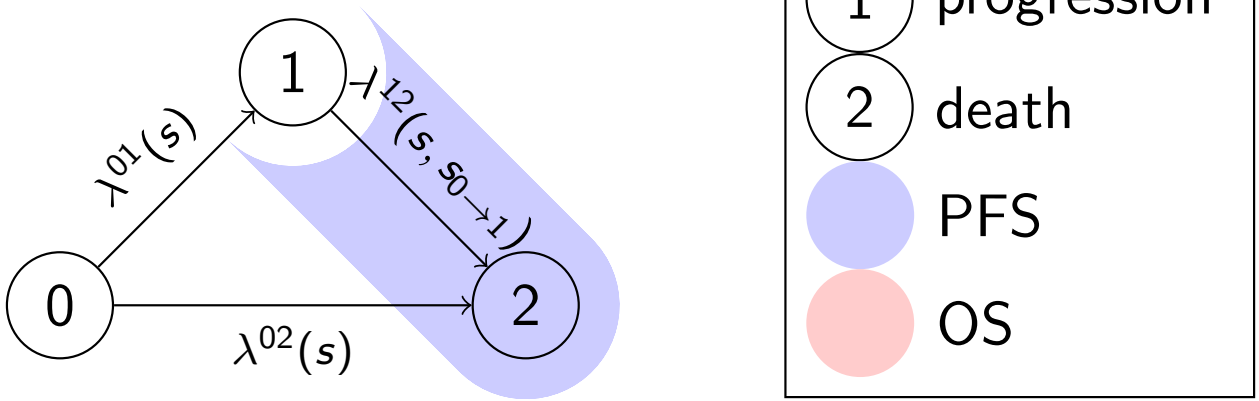
Advantages of one-sample method:

- Multiple dependence mechanisms applicable, e.g.
  - (Semi-)Markovianity of underlying multi-state model
  - Copula or frailty model
- Joint reference distribution specifies neccessary correction (i.e. $\lambda^j(u|\mathcal{F}_i(u))$)

Impose similar assumptions on the data generating process in two-sample case!

Illness-death models

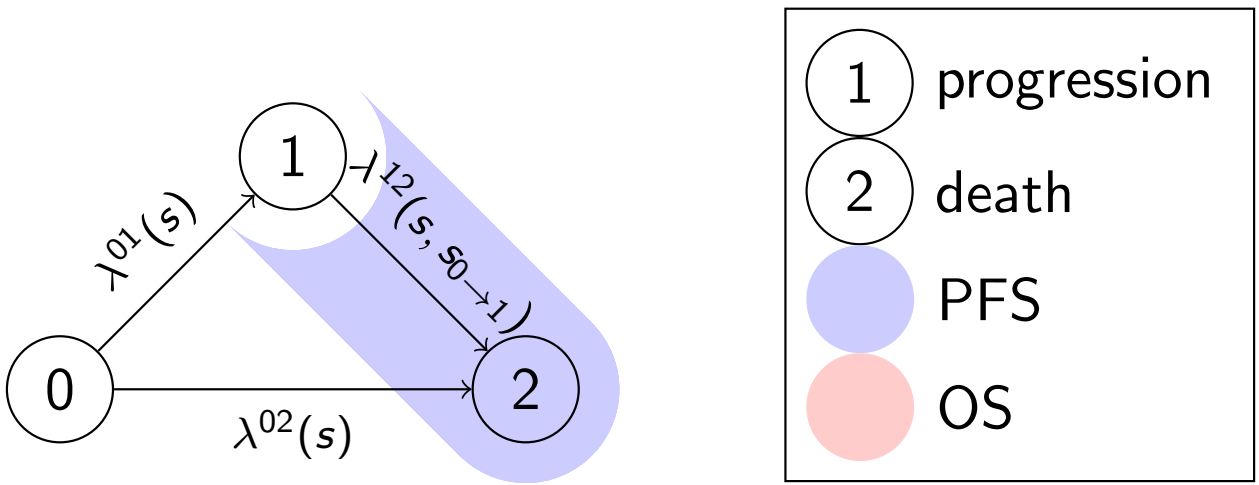
A multi-state model enables a more granular description of the course of disease



Events of clinical interest are given as "hitting times" of a set of nodes

Illness-death models

A multi-state model enables a more granular description of the course of disease

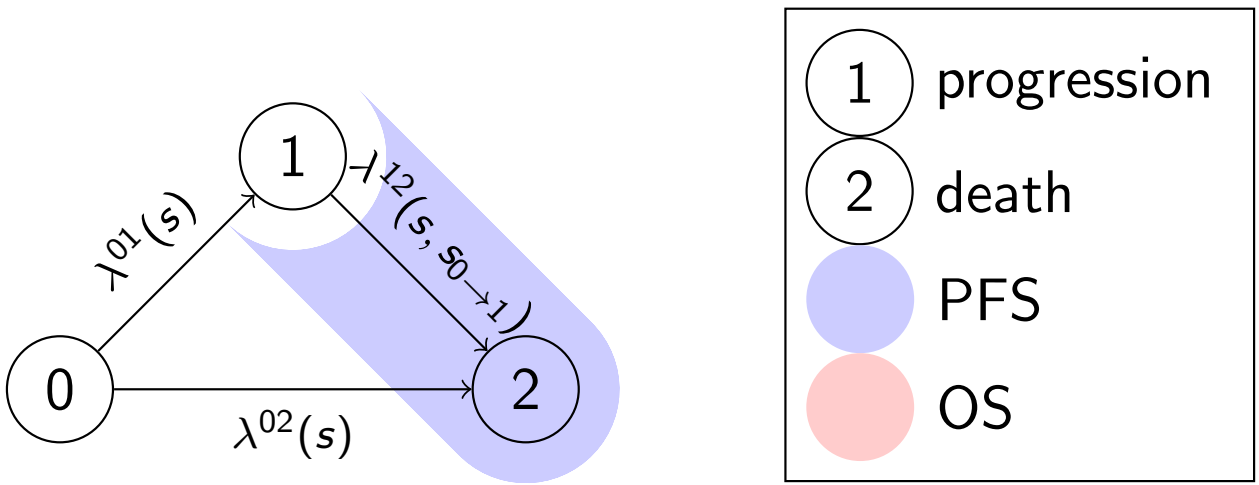


Markov model: $\lambda^{12}(s, s_{0\to1}) = \lambda^{12}_{M}(s)$

Semi-Markov model: $\lambda^{12}(s, s_{0\to1}) = \lambda^{12}_{SM}(s - s_{0\to1})$

Illness-death models

A multi-state model enables a more granular description of the course of disease



Markov model: $\lambda^{12}(s, s_{0\to1}) = \lambda^{12}_M(s)$

Semi-Markov model: $\lambda^{12}(s, s_{0\to1}) = \lambda^{12}_{SM}(s - s_{0\to1})$

Approach from Lin (1991)

Test of the hypothesis

$$H_{0,\text{marg}}\colon F_0^{\text{PFS}} = F_1^{\text{PFS}} \cap F_0^{\text{OS}} = F_1^{\text{OS}}$$

with the classical two-sample log-rank tests for the marginals

- $U^{\text{PFS}}(t) := \sum_{i=1}^{n} \int_0^t (Z_i - Y^{0,Z=1}(t,s)/Y^0(t,s)) N_i^{\text{PFS}}(t, ds)$
  ($\simeq$ martingale w.r.t. filtration generated by *PFS*-events)
- $U^{\text{OS}}(t) := \sum_{i=1}^{n} \int_0^t (Z_i - Y^{Z=1}(t,s)/Y(t,s)) N_i^{\text{OS}}(t, ds)$
  ($\simeq$ martingale w.r.t. filtration generated by *OS*-events)

and empirical estimation of covariance from asymptotically equivalent processes

Adjusted approach under Markov assumption

Test of the hypothesis

$$H_{0,\text{joint}}\colon F_0^{\text{PFS,OS}} = F_1^{\text{PFS,OS}} \qquad (\text{Note: } H_{0,\text{joint}} \supset H_{0,\text{marg}})$$

with the multivariate process

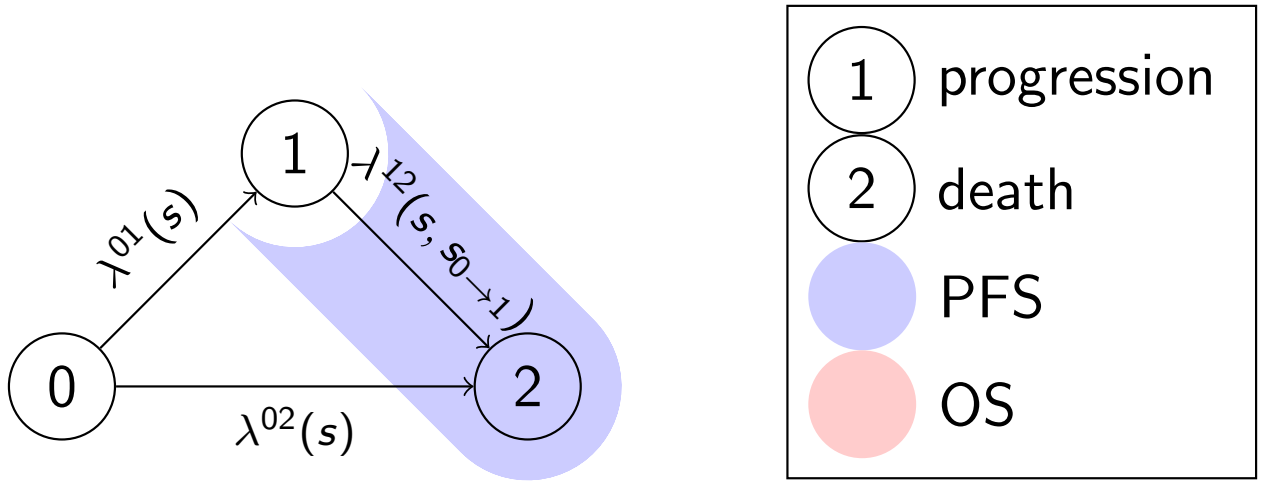
$$\mathbf{U}(t) := \begin{pmatrix} U^{\text{PFS}}(t) \\ U^{\text{OS}}(t) \end{pmatrix}$$

No correction for first component neccessary, i.e.

$$U^{\text{PFS}}(t) := \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{Y^{0,Z=1}(t,s)}{Y^0(t,s)} \right) N_i^{\text{PFS}}(t, ds)$$

Second component requires consideration of origin of transition!

Distinguishing events by transitions



Decompose $Y_i(t,s) = Y_i^0(t,s) + Y_i^1(t,s)$ where

- $Y_i^0(t,s) = 1$ if at calendar time t patient i is known to be healthy at trial time s
- $Y_i^1(t,s) = 1$ if at calendar time t patient i is known to be ill at trial time s

Adjusted approach under Markov assumption (contd.)

Test of the hypothesis

$$H_{0,\text{joint}}\colon F_0^{\text{PFS,OS}} = F_1^{\text{PFS,OS}} \qquad (\text{Note: } H_{0,\text{joint}} \supset H_{0,\text{marg}})$$

with the multivariate process

$$\mathbf{U}(t) := \begin{pmatrix} U^{\text{PFS}}(t) \\ U^{\text{OS}}(t) \end{pmatrix}$$

Adjusted second component:

$$U^{\text{OS}}(t) := \sum_{i=1}^{n} \int_0^t \left(Z_i - \mathbb{1}_{\{Y_i^0(t,s-)=1\}} \cdot \frac{Y^{Z=1,0}(t,s)}{Y^0(t,s)} \right.$$
$$\left. - \mathbb{1}_{\{Y_i^1(t,s-)=1\}} \cdot \frac{Y^{Z=1,1}(t,s)}{Y^1(t,s)} \right) N_i^{\text{OS}}(t, ds)$$

Asymptotics

Asymptotic equivalence:

$$\frac{1}{\sqrt{n}}(\mathbf{U}(t) - \tilde{\mathbf{U}}(t)) \xrightarrow{\mathbb{P}} 0 \qquad \forall t \geq 0$$

where $(\tilde{\mathbf{U}}(t))_{t\geq 0}$ is a martingale w.r.t. the filtration generated by PFS- and OS-events!

Central limit theorem yields:

- $\frac{1}{\sqrt{n}}(\tilde{\mathbf{U}}(t)-\tilde{\mathbf{U}}(s)) \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{n}}(\tilde{\mathbf{U}}_\infty(t)-\tilde{\mathbf{U}}_\infty(s)) \sim \mathcal{N}(0, \mathbf{V}(t)-\mathbf{V}(s))$
- asymptotically independent increments

Sequential testing procedure

For analysis dates $0 =: t_0 < t_1 < t_2$ we get standardized stagewise test statistics

$$\mathbf{Z}_1 = \hat{\mathbf{C}}_1^{-1} \begin{pmatrix} U^{\text{PFS}}(t_1) \\ U^{\text{OS}}(t_1) \end{pmatrix} \text{ resp. } \mathbf{Z}_2 = \hat{\mathbf{C}}_2^{-1} \begin{pmatrix} U^{\text{PFS}}(t_2) - U^{\text{PFS}}(t_1) \\ U^{\text{OS}}(t_2) - U^{\text{OS}}(t_1) \end{pmatrix}$$

which...

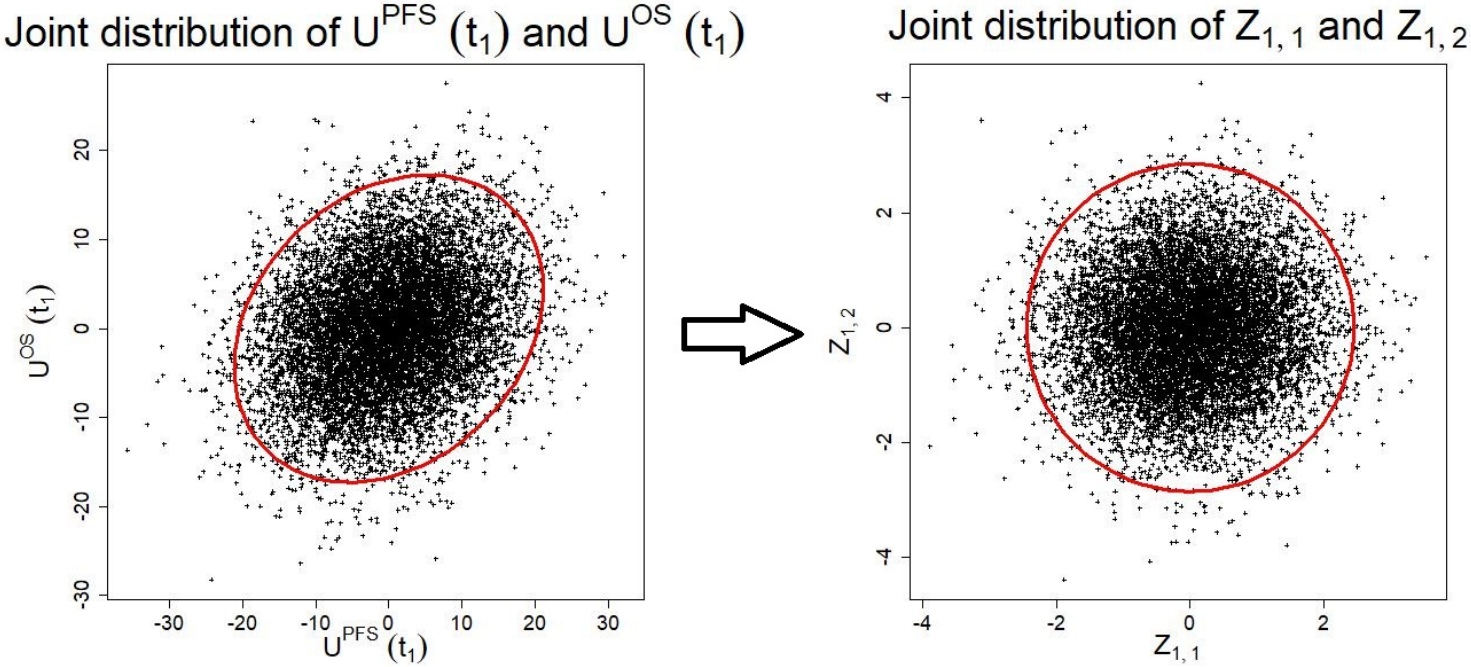
- ...asymptotically follow a bivariate standard normal distribution
- ...are asymptotically independent

Agreement with asymptotic distribution is already observed for small sample sizes

Stagewise p-values

Obtain stagewise p-values by taking a norm of the stagewise test statistic:

L^2-norm $|Z_1|_2^2$ and $|Z_1|_2^2$ are χ^2_2-distributed (as in Wei and Lachin (1984))



Can be combined e.g. by inverse normal method

Sample size calculation

Alternative hypotheses in terms of the transition hazards of the model

In the case of our example:

- Does the therapy prevent from progressions?
- Is it more or less toxic and thus leading to direct deaths?
- Is it a good salvage therapy?

Use e.g. transition-wise hazard ratios δ^{01}, δ^{02} and δ^{12}

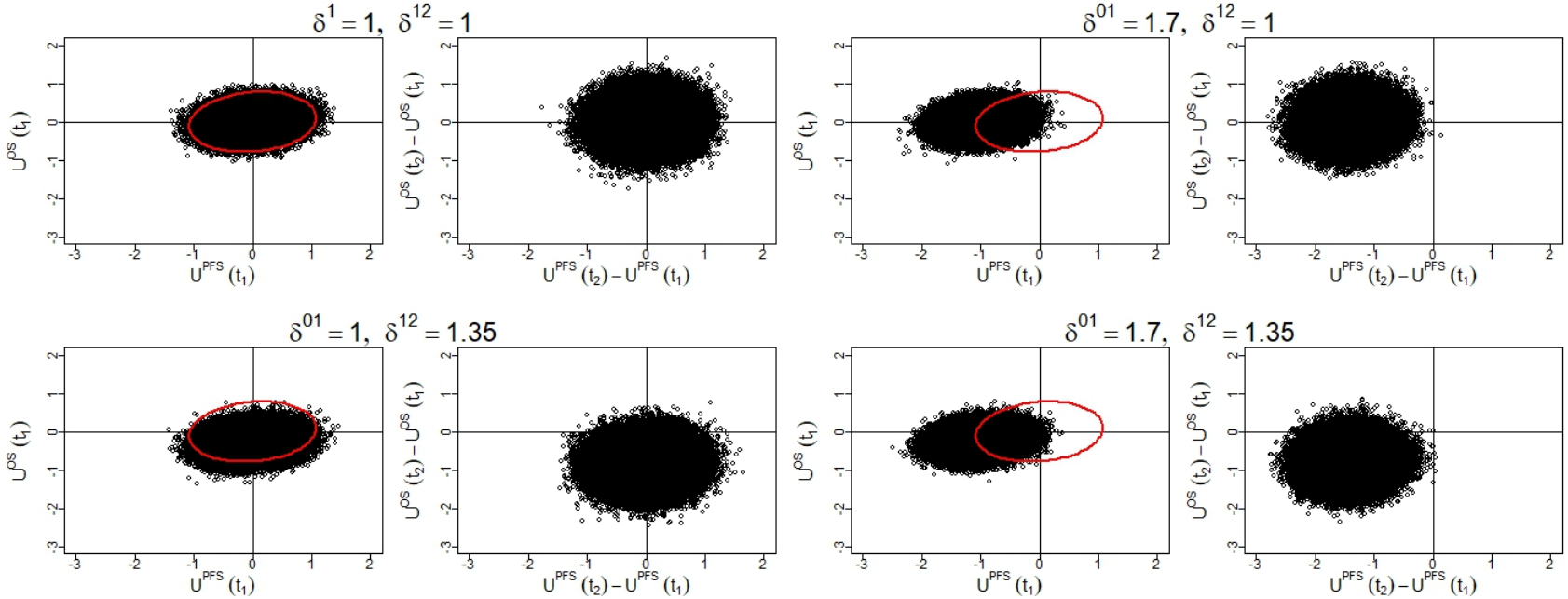
Hazard ratios imply drift function

$$n \cdot d(t; \delta^{01}, \delta^{02}, \delta^{12})$$

in calendar time t.

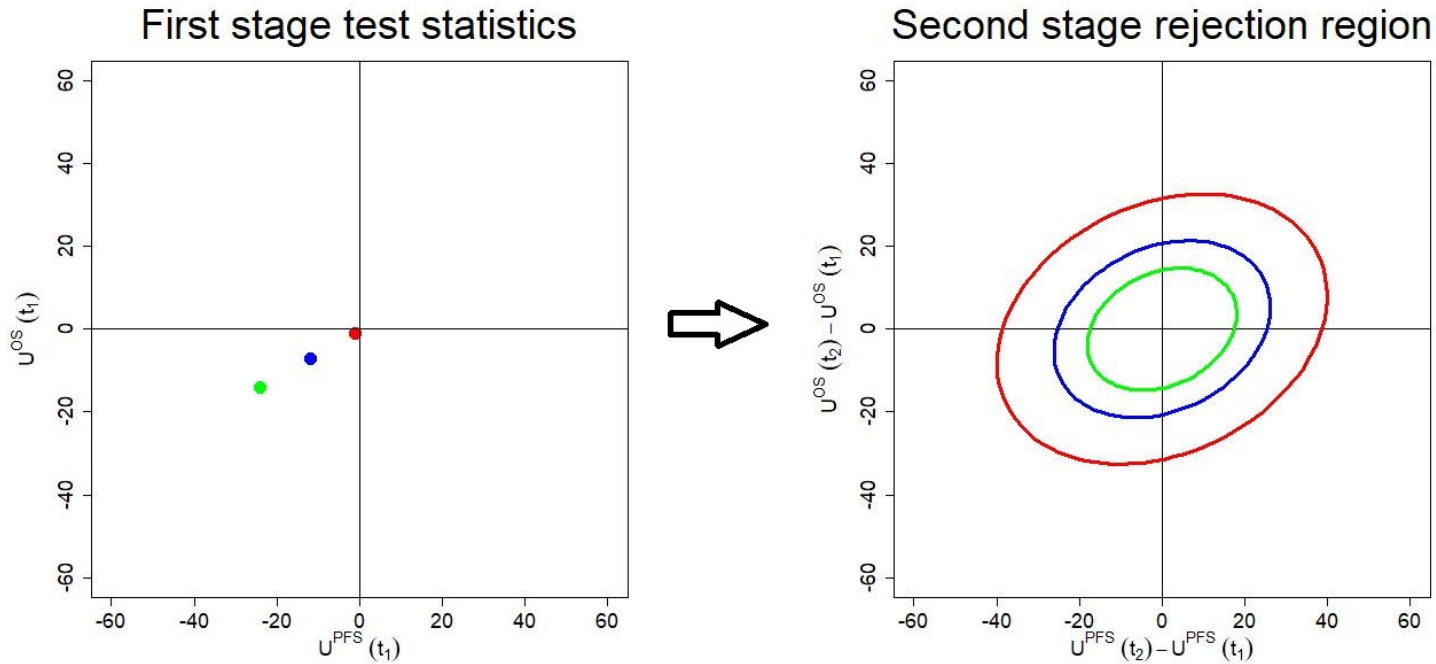
Drift under alternatives

Hazard ratios cause shift of increments of multivariate process **U** in corresponding direction:



Adaptations in a two-stage design

First-stage p-value p_1 implies rejection region $\mathcal{R}_2(p_1)$ of second stage



Separate re-assessment of hazard ratios δ^{01}, δ^{02} and δ^{12}
$\Rightarrow$ Manipulation of external parameters of drift function yields required conditional power

Generalized setting

Non-recurrent multi-state model with state space $S = \{0, \ldots, k\}$

Transition times $T^{\{1\}}, \ldots, T^{\{k\}}$

Clinically relevant composite events $E_1, \ldots, E_d \subset \{1, \ldots, k\}$ with event times

$$T^{E_m} := \min_{j \in E_m} T^{\{j\}}$$

Component m of test statistic

$$U^{E_m}(t) := \sum_{i=1}^{n} \int_0^{t \wedge T_i^{E_m}} \left(Z_i - \sum_{l \notin E_m} \mathbb{1}_{\{Y_i^l(t,s-)=1\}} \frac{Y^{Z=1,l}(t,s)}{Y^l(t,s)} \right) N_i^{E_m}(t, ds)$$

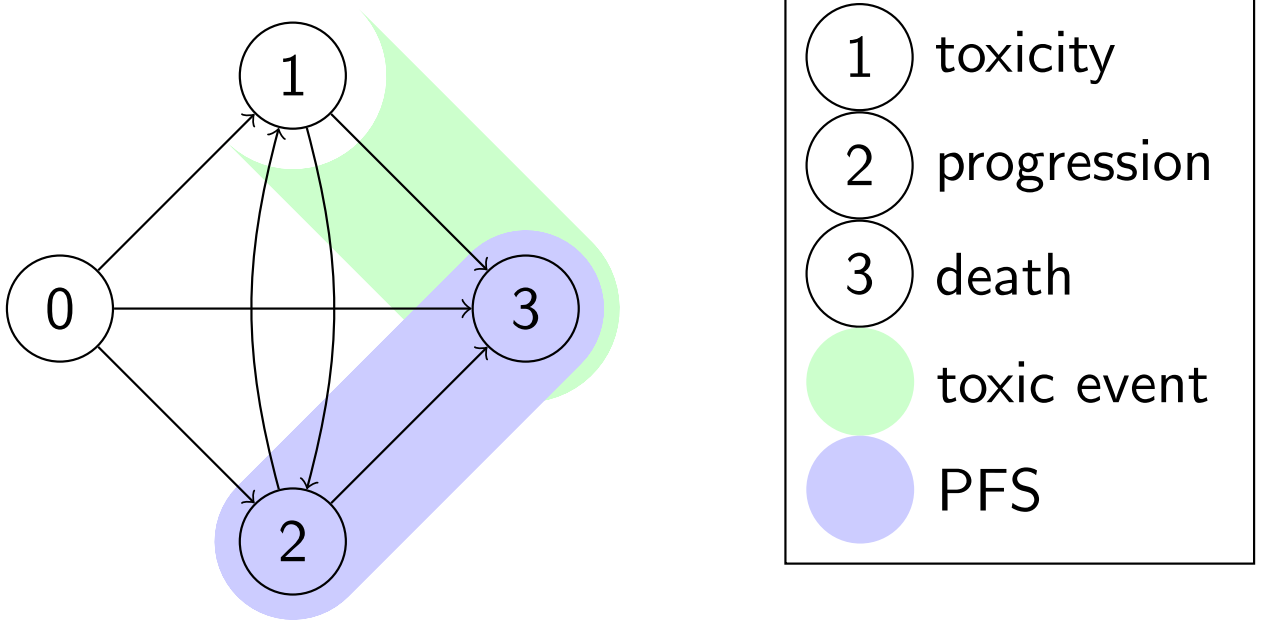
In particular, we generalize

- the standard log-rank test ($k = 1$, $d = 1$ and $E_1 = \{1\}$)
- the procedure of Lin (1991) in a competing risks setting

Further application

Simultaneous consideration of efficacy and toxicity

Acounting for death as a competing event



$k = 3$, $d = 2$, $E_1 = \{2, 3\}$ (PFS), $E_2 = \{1, 3\}$ (TFS)

Beyond the Markov assumption

Cox-Markov model:

- Transition intensity may depend on additional parameters
- E.g. $t^{\{l\}}$ be the time of transition into current state l

$$\lambda_j(u|\mathcal{F}_i(u)) = \lambda_{lj}(u)\exp(\beta_{lj}t^{\{l\}})$$

- Estimate and plug in regression parameters as in log-rank test with covariates

Semi-Markov model:

- Transition intensities given by

$$\lambda_j(u|\mathcal{F}_i(u)) = \lambda_{lj}(u - t^{\{l\}})$$

- Requires different definition of "at risk sets" and individual counting processes

References I

Bauer P and Posch M (2004). Letter to the editor: Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine* **23**(8), 1333–1335.

Danzer MF, Terzer T, Berthold F, Faldum A and Schmidt R (2022). Confirmatory adaptive group sequential designs for single-arm phase II studies with multiple time-to-event endpoints. *Biometrical Journal* **64**(2), 312–342.

Lachin JM (2014). Applications of the Wei-Lachin Multivariate One-Sided Test for Multiple Outcomes on Possibly Different Scales. *PLOS ONE* **9**(10), e108784.

Lin DY (1991). Nonparametric sequential testing in clinical trials with incomplete multivariate observations. *Biometrika* **78**(1), 123–131.

Perlman MD and Wu L (2004). A note on one-sided tests with multiple endpoints. *Biometrics* **60**(1), 276–280.

Pocock SJ, Geller NL, Tsiatis AA (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* **43**(3), 487–498.

Tamhane AJ and Logan BR (2004). A superiority-equivalence approach to one-sided tests on multiple endpoints in clinical trials. *Biometrika* **91**(3), 715–727.

References II

Tang DI, Gnecco C, Geller NL (1989) An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. *Biometrika* **76**(3), 577–583.

Wei LJ and Lachin JM (1984). Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations. *Journal of the American Statistical Association* **79**(387), 653–661.

Estimation of covariance matrix $\mathbf{V}$

Non-zero elements of covariance matrix $\Leftrightarrow$ events occur simultaneously

In our case: Deaths without prior illness, i.e.

$$
\begin{aligned}
&\hat{V}_{12}(t) \\
:=&\widehat{\mathrm{Cov}}\left(\frac{1}{\sqrt{n}}U^{\mathrm{PFS}}(t), \frac{1}{\sqrt{n}}U^{\mathrm{OS}}(t)\right) \\
=&\frac{1}{n}\sum_{i=1}^{n}\int_{[0,t]}\mathbb{1}_{\{Y_i^0(t,s-)=1\}}\frac{Y^{Z=1,0}(t,s)}{Y^0(t,s)}\left(1-\frac{Y^{Z=1,0}(t,s)}{Y^0(t,s)}\right)N_i^{\mathrm{OS}}(t,ds)
\end{aligned}
$$

Compute lower triangular Cholesky factor $\hat{\mathbf{C}}$ of estimate $\hat{\mathbf{V}} = \hat{\mathbf{C}}\hat{\mathbf{C}}^T$

Simulation results: Empirical type I errrors

4 different simulation scenarios:

1. time-homogeneous Markov model
2. time-inhomogeneous Markov model
3. Semi-Markov model
4. Frailty model (baseline intensities from 2. with joint Gamma frailty)

Combination of p-values from l^2 norm via inverse normal with equal weights and O'Brien-Fleming rejection bounds

Empirical type I errors ($\alpha = 0.05$, 200,000 runs):

Scenario	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
1	0.0581	0.0538	0.0517	0.0504	0.0496
2	0.0592	0.0550	0.0516	0.0502	0.0503
3	0.0527	0.0484	0.0472	0.0456	0.0458
4	0.0587	0.0547	0.0527	0.0507	0.0500

One-sided testing

Previously shown methods can only reject $H_{0,\text{joint}}$ in favour of

$$H_{1,\text{joint}}: F_0^{E_1,\ldots,E_k} \neq F_1^{E_1,\ldots,E_k}$$

Makes further interpretation of results difficult!

Possible choices for one-sided stagewise p-values:

- Least squares approaches (GLS or OLS, see e.g. Pocock et al. (1987) or Lachin (2014))
- Approximate Likelihood Ratio test (see Tang et al. (1989))
- Non-inferiority for all and superiority for at least one endpoint (see Tamhane and Logan (2004) and Perlman and Wu (2004))