Multistage modelling of a cardiovascular trial using saddlepoint approximation*

M. Hudson[2,†], S. Heritier[1] and S. Lo[1,2]

[1] The George Institute, Sydney University
[2] Macquarie University
[†] Support from the NHMRC Clinical Trials Centre, Australia.

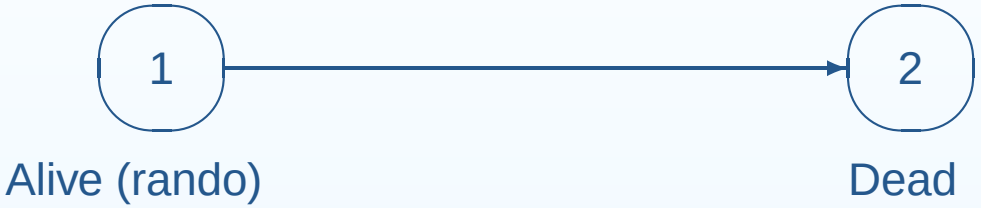
mhudson@george.org.au

November 2009

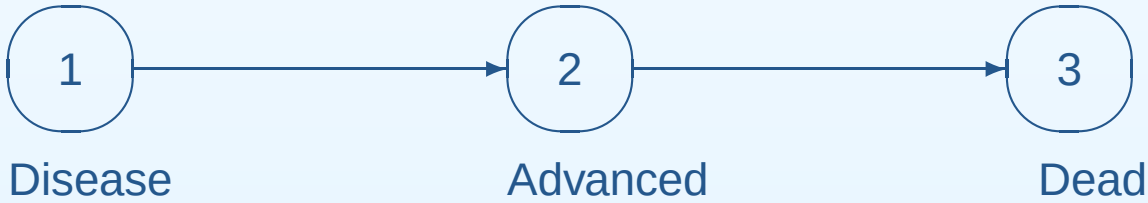
*Computational Statistics and Data Analysis (2008), doi:10.1016/j.csda.2008.09.003

Background :

- **Two-stage illness-death model**



- **Progressive illness disease model**

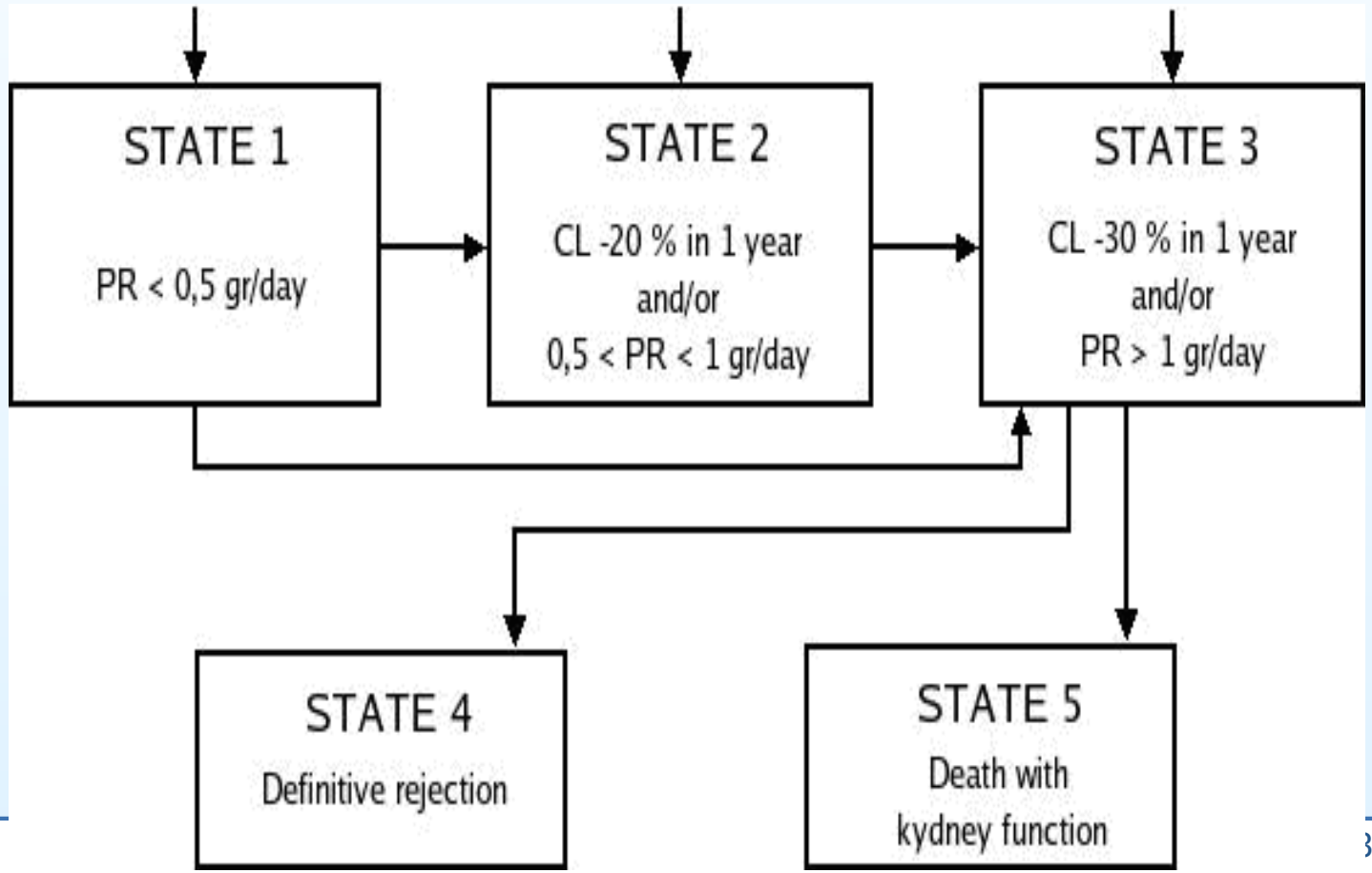


- Events : progression and death
- Outcomes : time to event T_{12}, T_{23} or $T = T_{12} + T_{23}$
- Std. methods : KM plots, Cox PH...
- Analyse each outcome separately

- **Diabetic retinopathy flowgraph** (with feedback)
  Yau and Huzurbazar, SIM (2003)

  - Type I diabetes for 5+ years
  - longitudinal observational study, n=277
  - 368 transitions, five year follow-up
  - no treatment group
  - fit based on an inverse Gaussian model (using saddlepoint method to estimate survival)

- **Kidney transplant recipients** (Competing Risks)
  see Foucher et al, SM (2007)
  - 3-levels of severity, form states based on Creatinine clearance (CL) & Proteinuria (PR)
  - 2-terminal states : chronic rejection of the kidney and death.



Standard approach

RCT reality : *multiple events, possibly recurrent, competing risks* e.g. for LIPID,[1]

- **Primary outcome** :
  - CHD death
- **Secondary outcomes** :
  - death from any cause
  - death from cardiovascular causes
  - death from CHD or nonfatal MI
  - MI

LIPID : RCT in patients with CHD history, NEJM (1998).

[1] LIPID study group, NEJM, 339 :1349–1357, 1998

Potential drawbacks

- Each outcome treated separately (multiplicity)
- Recurrent events omitted (time to *first* occurrence)[2]
- Competing risks issue (poorly dealt with / ignored ?)
- Not enough events $\longrightarrow$ combined endpoint

Potential drawbacks

- Each outcome treated separately (multiplicity)
- Recurrent events omitted (time to *first* occurrence)[2]
- Competing risks issue (poorly dealt with / ignored ?)
- Not enough events $\longrightarrow$ combined endpoint
- *Overuse of the Cox model*

Potential drawbacks

- Each outcome treated separately (multiplicity)
- Recurrent events omitted (time to *first* occurrence)[2]
- Competing risks issue (poorly dealt with / ignored ?)
- Not enough events $\longrightarrow$ combined endpoint
- *Overuse of the Cox model*

Alternatively

see the whole process as a multistage disease.

Potential drawbacks

- Each outcome treated separately (multiplicity)
- Recurrent events omitted (time to *first* occurrence)[2]
- Competing risks issue (poorly dealt with / ignored ?)
- Not enough events $\longrightarrow$ combined endpoint
- *Overuse of the Cox model*

Alternatively

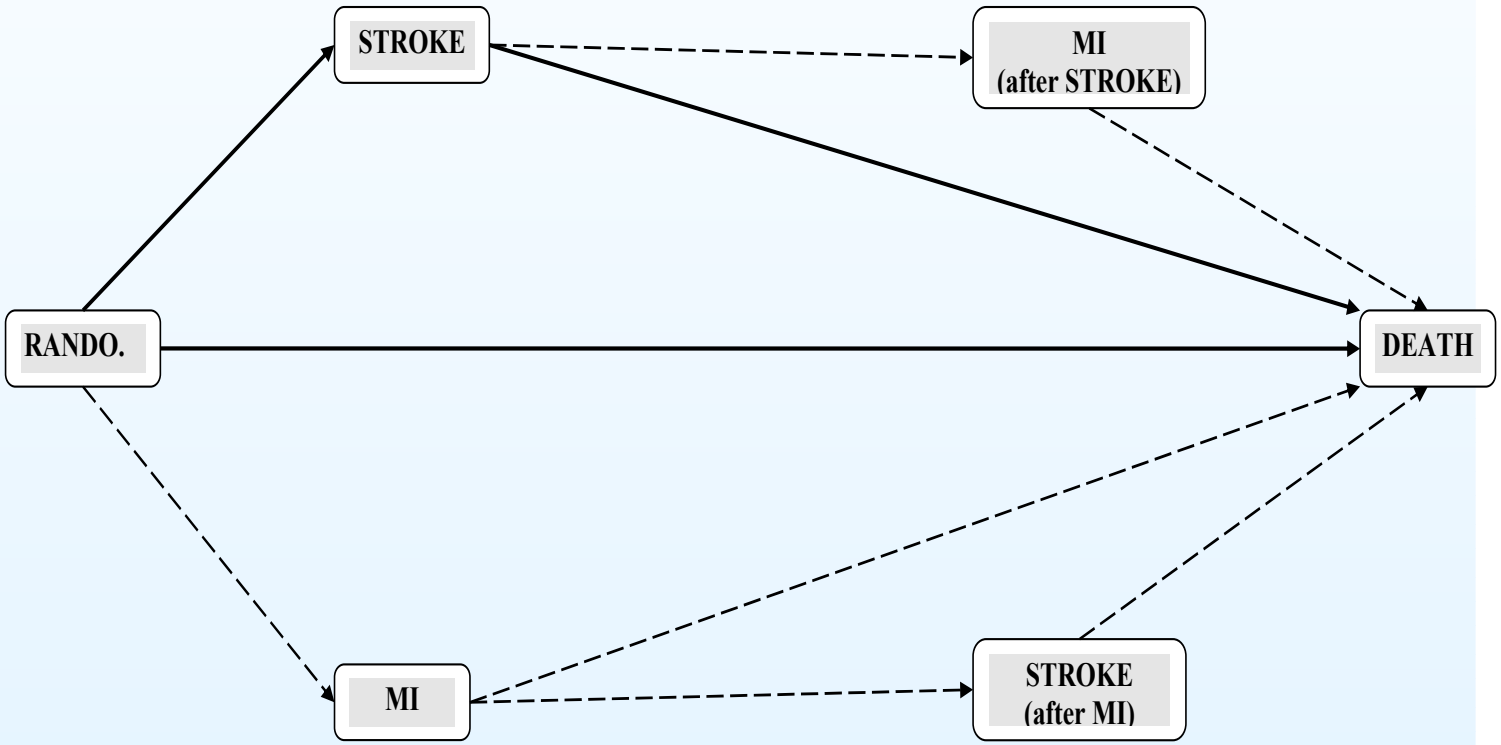
see the whole process as a multistage disease.

This requires

- a multistage model (here semi-Markov)
- a way to fit it / interpret the results

[2] It is possible to account for these (e.g. WLW analysis)

- **e.g. LIPID model based on the flowgraph below**



Semi-Markov Process (SMP)

- Less restrictive assumptions than a Markov model

Semi-Markov Process (SMP)

- Less restrictive assumptions than a Markov model
- transition times between an initial state and a next state are independent of history prior to first state,
- with distribution that depends only on the adjoining states

Semi-Markov Process (SMP)

- Less restrictive assumptions than a Markov model
- transition times between an initial state and a next state are independent of history prior to first state,
- with distribution that depends only on the adjoining states
- standard approach : model specified through transition intensity (or cause-specific hazards)

Semi-Markov Process (SMP)

- Less restrictive assumptions than a Markov model
- transition times between an initial state and a next state are independent of history prior to first state,
- with distribution that depends only on the adjoining states
- standard approach : model specified through transition intensity (or cause-specific hazards)
- Non-constant hazard distributions

Semi-Markov Process (SMP)

- Less restrictive assumptions than a Markov model
- transition times between an initial state and a next state are independent of history prior to first state,
- with distribution that depends only on the adjoining states
- standard approach : model specified through transition intensity (or cause-specific hazards)
- Non-constant hazard distributions
- Characterised by a *transmittance matrix* = the product of 2 matrices (elementwise) - Butler (2001-02)

Applications to RCT ?

Little work so far

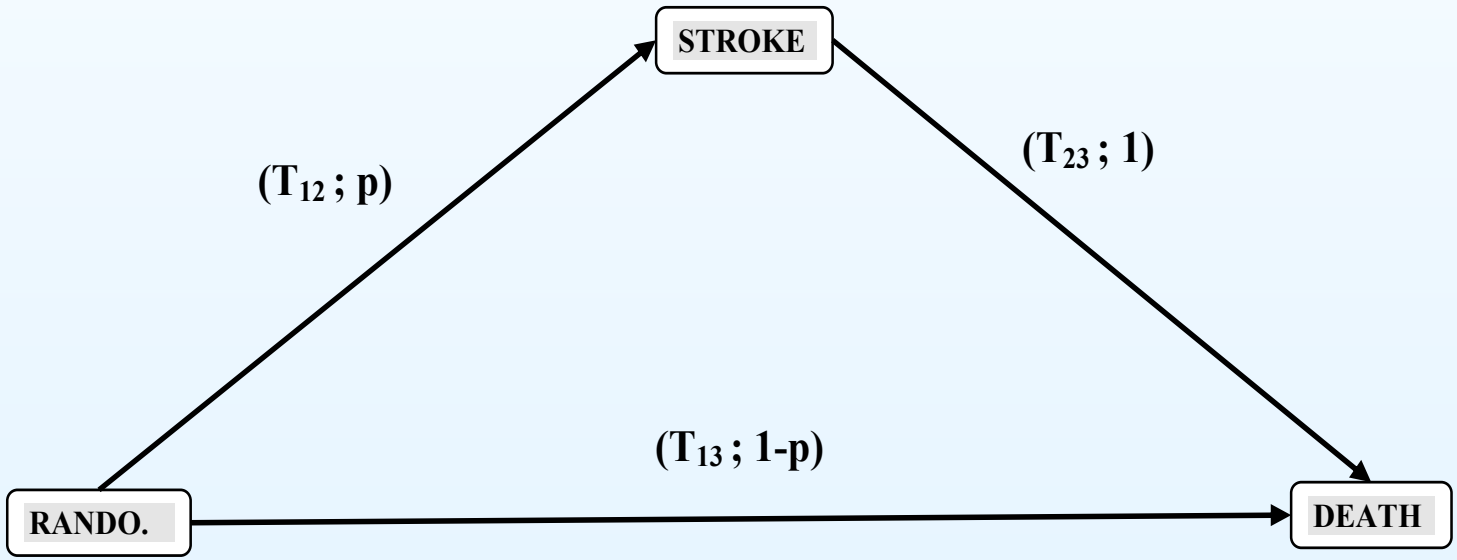
- more on observational data (disease pathways)
- difficult to estimate relevant quantities in RCT
- presence of censoring
- mathematical complexity $\longrightarrow$ Saddlepoint technique (SP)

Objectives of this work

- Develop SP methods (Butler's work) for use in RCTs. In particular, account for censoring
- Apply the new methodology to the LIPID data
- Model its primary outcome while accounting for prior events
- Develop its implementation and flexibility
- Comparison with standard approach (HR, survival...)
- Explore the link with competing risks methodology

The model

Let p_{ij} and T_{ij} for $i, j = 1, 2, 3$ be respectively the transition probability and the transition time from i to j.



With this simple pattern, the transmittance matrix is defined by,

$$\mathcal{Q}(s) = \begin{pmatrix} 0 & p_{12}.M_{12}(s) & p_{13}.M_{13}(s) \\ 0 & 0 & p_{23}.M_{23}(s) \\ 0 & 0 & 0 \end{pmatrix}$$

where $M_{ij}(s) = MGF$ of the transition time from i to j.

e.g. in the case of the *exponential distribution* with mean μ,

$$M(s) = \frac{1}{1 - \mu s}$$

Data summary

	placebo$(N = 4502)$	pravastatin $(N = 4512)$
Censoring at		
1	3728	3888
2	141	126
Transitions		
$1 \rightarrow 2$	185	151
$1 \rightarrow 3$	589	473
$2 \rightarrow 3$	44	25

Probability transition matrix for placebo group $(\mathcal{P}_P)$ and pravastatin group $(\mathcal{P}_T)$;

$$\mathcal{P}_P = \begin{pmatrix} 0 & 0.239 & 0.761 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \mathcal{P}_T = \begin{pmatrix} 0 & 0.242 & 0.758 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Fitting a SMP

General idea :

1. estimate parameters[3] within transition stages ;
2. form MGF of time through system algebraically from individual stage MGFs using the transmittance matrix ;
3. invert the MGF using saddle-point methods, providing the survival distribution (or hazard) to terminal event.

Butler's method assumes fully observed data (no censoring).
The presence of *censoring* creates an additional difficulty
$\Rightarrow$ we resort to parametric LL techniques to estimate the MGFs and therefore the transmittance matrix

[3] A nonparametric approach is difficult, but we have an approach !

MGF and saddlepoint

Butler's work [4] leads to explicit MGF for T, time through the system.

$$\begin{aligned}\mathcal{M}_{1m}(s) &:= \frac{(m;1)\text{-th cofactor } [I_m - \mathcal{Q}(s)]}{(m;m)\text{-th cofactor } [I_m - \mathcal{Q}(s)]} \\ &:= \frac{(-1)^{m+1}\Psi_{m1}(s)}{\Psi_{mm}(s)}\end{aligned}$$

MGF and saddlepoint

Butler's work [4] leads to explicit MGF for T, time through the system.

$$\begin{aligned}\mathcal{M}_{1m}(s) &:= \frac{(m;1)\text{-th cofactor } [I_m - \mathcal{Q}(s)]}{(m;m)\text{-th cofactor } [I_m - \mathcal{Q}(s)]} \\ &:= \frac{(-1)^{m+1}\Psi_{m1}(s)}{\Psi_{mm}(s)}\end{aligned}$$

where Ψ_{ij} is the (i,j)-th minor of $I_m - \mathcal{Q}()$

e.g. 3-node model : $M_T(s) = pM_{12}(s)M_{23}(s) + (1-p)M_{13}$

[4] proof : see Butler (2000,2001, 2006)

Let $\mathcal{K}(s) = \ln(\mathcal{M}_{1m})$ be the *cumulant generating fct* of T
The survival function of T is approximated by the **Lugannani-Rice** formula :

$$S_T(t) = 1 - \Phi(\hat{w}) - \varphi(\hat{w})(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}) \qquad t \neq E(T) = \mathcal{K}'(s)$$

Let $\mathcal{K}(s) = \ln(\mathcal{M}_{1m})$ be the *cumulant generating fct* of T
The survival function of T is approximated by the **Lugannani-Rice** formula :

$$S_T(t) = 1 - \Phi(\hat{w}) - \varphi(\hat{w})(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}) \qquad t \neq E(T) = \mathcal{K}'(s)$$

where Φ and φ are respectively the pdf and the cdf of the standard normal distribution, $\hat{w} = \hat{w}(\hat{s})$ and $\hat{u} = \hat{u}(\hat{s})$ depend on $\hat{s}$ according to
$\hat{w} = sign(\hat{s})\sqrt{2\{\hat{s}t - \mathcal{K}(\hat{s})\}}$,
$\hat{u} = \hat{s}\sqrt{\mathcal{K}''(\hat{s})}$,
and the saddlepoint $\hat{s}$ solves $\mathcal{K}'(\hat{s}) = t$ for $t > 0$.

$\mathcal{K}'$ and $\mathcal{K}''$ given by a simple algebraic formulas see (Butler 2007).

Likelihood function

Uncensored intervals Partial likelihood

Transitions

Censored state

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \rightarrow 2$	$L_{12} = \prod_{u=1}^{N_{12}} pf_{12}(x_{u12}, \theta_{12})$
Censored state	

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \rightarrow 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \rightarrow 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
Censored state	

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \to 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \to 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
$2 \to 3$	$L_{23} = \prod_{u=1}^{N_{23}} f_{23}(x_{u23}, \theta_{23})$
Censored state	

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \to 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \to 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
$2 \to 3$	$L_{23} = \prod_{u=1}^{N_{23}} f_{23}(x_{u23}, \theta_{23})$
Censored state	
1	$L_1 = \prod_{c=1}^{N_{11}} [p S_{12}(x^*_{c1}) + (1-p) S_{13}(x^*_{c1})]$

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \to 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \to 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
$2 \to 3$	$L_{23} = \prod_{u=1}^{N_{23}} f_{23}(x_{u23}, \theta_{23})$
Censored state	
1	$L_1 = \prod_{c=1}^{N_{11}} [pS_{12}(x^*_{c1}) + (1-p)S_{13}(x^*_{c1})]$
2	$L_2 = \prod_{c=1}^{N_{22}} S_{23}(x^*_{c2}) = \prod_{c=1}^{N_{22}} [1 - F_{23}(x^*_{c2})]$

see Huzurbazar (2004)

Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \to 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \to 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
$2 \to 3$	$L_{23} = \prod_{u=1}^{N_{23}} f_{23}(x_{u23}, \theta_{23})$
Censored state	
1	$L_1 = \prod_{c=1}^{N_{11}} [pS_{12}(x^*_{c1}) + (1-p)S_{13}(x^*_{c1})]$
2	$L_2 = \prod_{c=1}^{N_{22}} S_{23}(x^*_{c2}) = \prod_{c=1}^{N_{22}} [1 - F_{23}(x^*_{c2})]$

$$L(\theta, p) = L_{12} \cdot L_{23} \cdot L_{13} \cdot L_1 \cdot L_2$$

see Huzurbazar (2004)

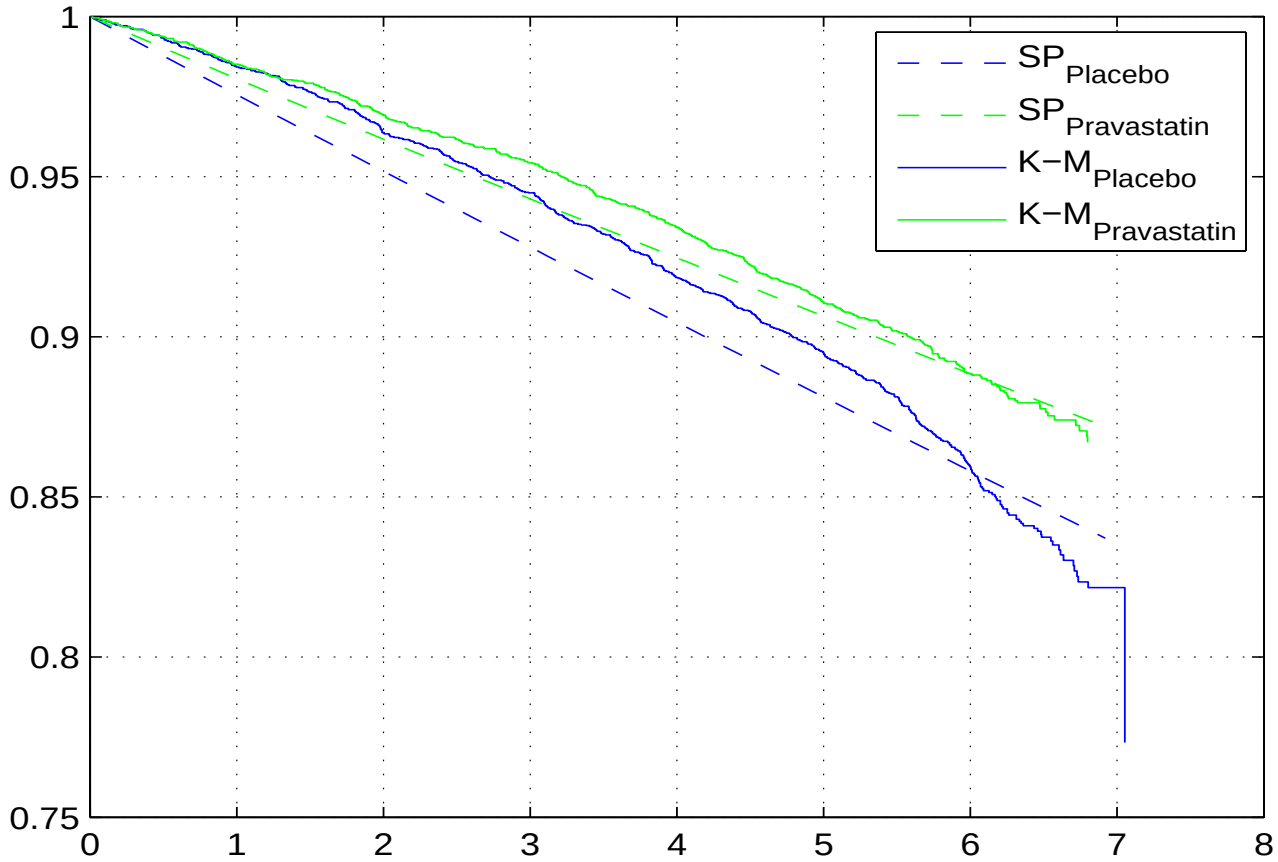
Likelihood function

Uncensored intervals	Partial likelihood
Transitions	
$1 \to 2$	$L_{12} = \prod_{u=1}^{N_{12}} p f_{12}(x_{u12}, \theta_{12})$
$1 \to 3$	$L_{13} = \prod_{u=1}^{N_{13}} (1-p) f_{13}(x_{u13}, \theta_{13})$
$2 \to 3$	$L_{23} = \prod_{u=1}^{N_{23}} f_{23}(x_{u23}, \theta_{23})$
Censored state	
1	$L_1 = \prod_{c=1}^{N_{11}} [pS_{12}(x^*_{c1}) + (1-p)S_{13}(x^*_{c1})]$
2	$L_2 = \prod_{c=1}^{N_{22}} S_{23}(x^*_{c2}) = \prod_{c=1}^{N_{22}} [1 - F_{23}(x^*_{c2})]$

$$L(\theta, p) = L_{12} \cdot L_{23} \cdot L_{13} \cdot L_1 \cdot L_2$$

see Huzurbazar (2004)

Markov model : KM survival and SP model with exponential distributions of time in each state



LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**
- failures rates increase due to **aging** or **wear out**.

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**
- failures rates increase due to **aging** or **wear out**.

hazard function	survival function
$h(t) = \alpha$	$S(t) = \exp\left[-t\ \alpha \qquad\right]$

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**
- failures rates increase due to **aging** or **wear out**.

hazard function	survival function
$h(t) = \alpha + \frac{1}{\beta^2}t$	$S(t) = \exp\left[-t\ \alpha \qquad\right]$

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**
- failures rates increase due to **aging** or **wear out**.

hazard function	survival function
$h(t) = \alpha + \frac{1}{\beta^2}t$	$S(t) = \exp\left[-t(\alpha + \frac{1}{2\beta^2}t)\right]$

LHRD

Linear hazard rate distribution (LHRD)

popular for modelling the life-length of a system or component (reliability theory)

particularly when

- failures **occur at random**
- failures rates increase due to **aging** or **wear out**.

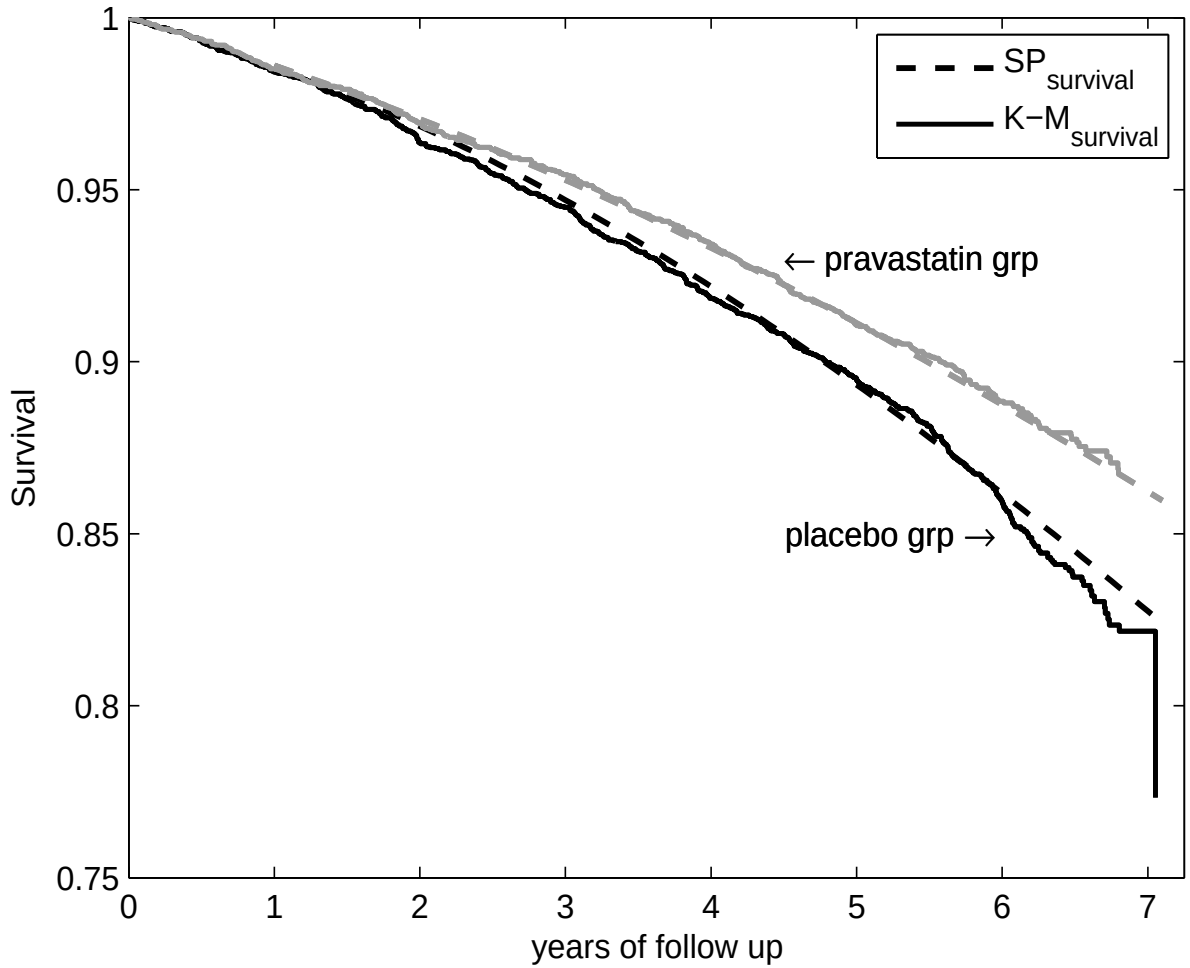
hazard function	survival function
$h(t) = \alpha + \frac{1}{\beta^2}t$	$S(t) = \exp\left[-t(\alpha + \frac{1}{2\beta^2}t)\right]$

MGF has a close form

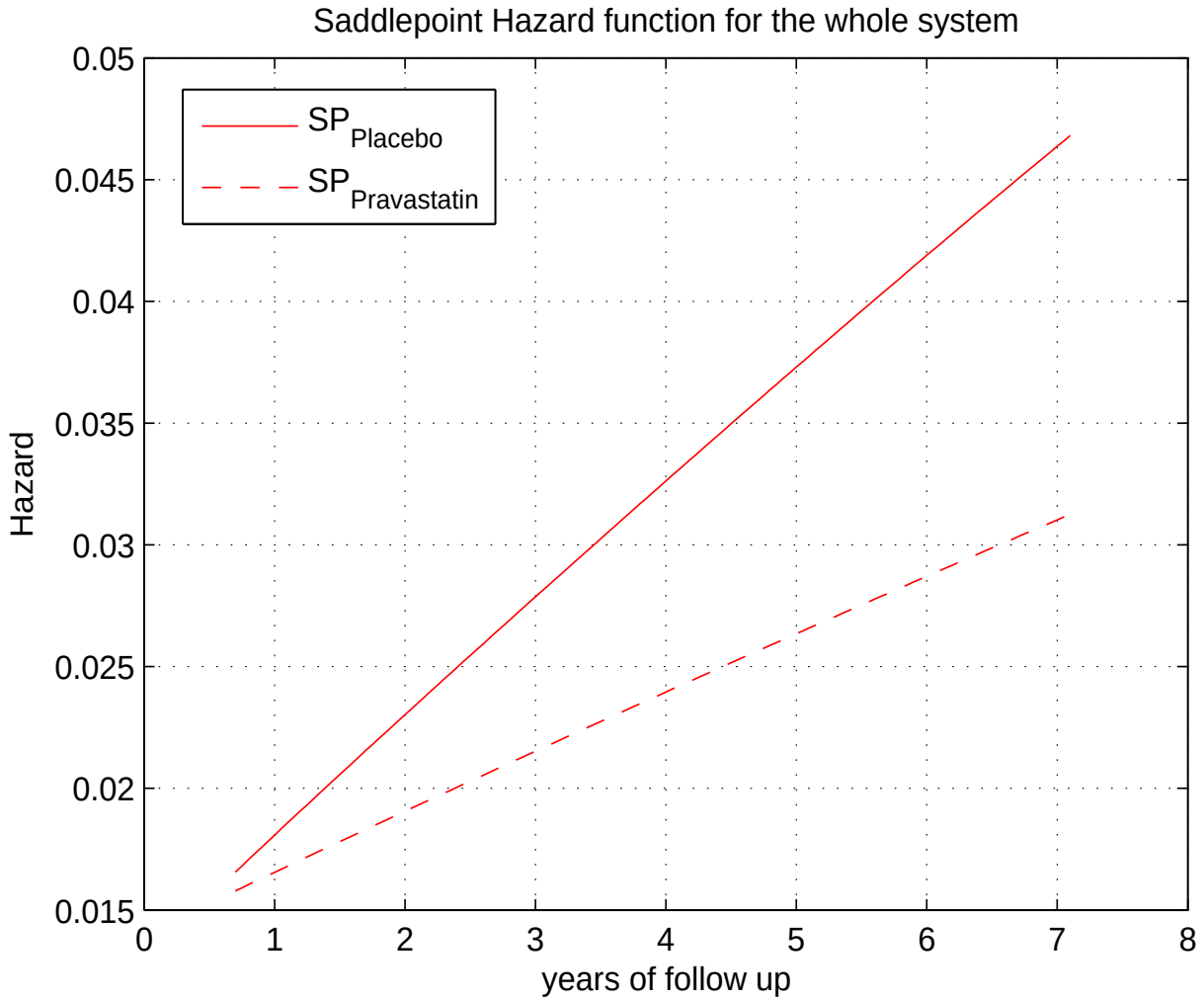
Results

- Background:
- Semi-Markov Process (SMP)
- The model
- Data summary
- MGF and saddlepoint
- Likelihood function
- LHRD
- Results
- Assumptions
- Conclusions

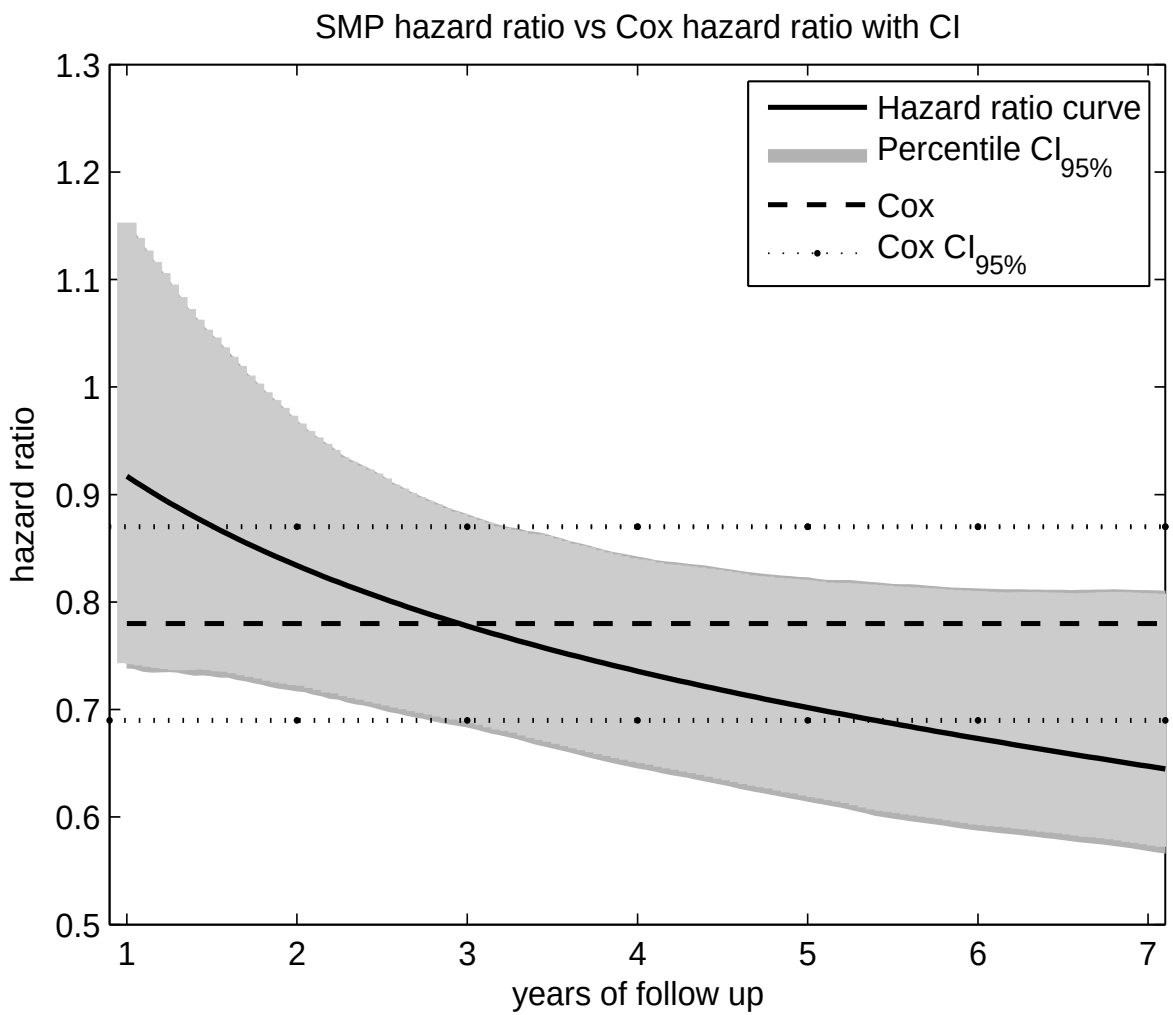
Overall survival curve with stroke as intermediate endpoint



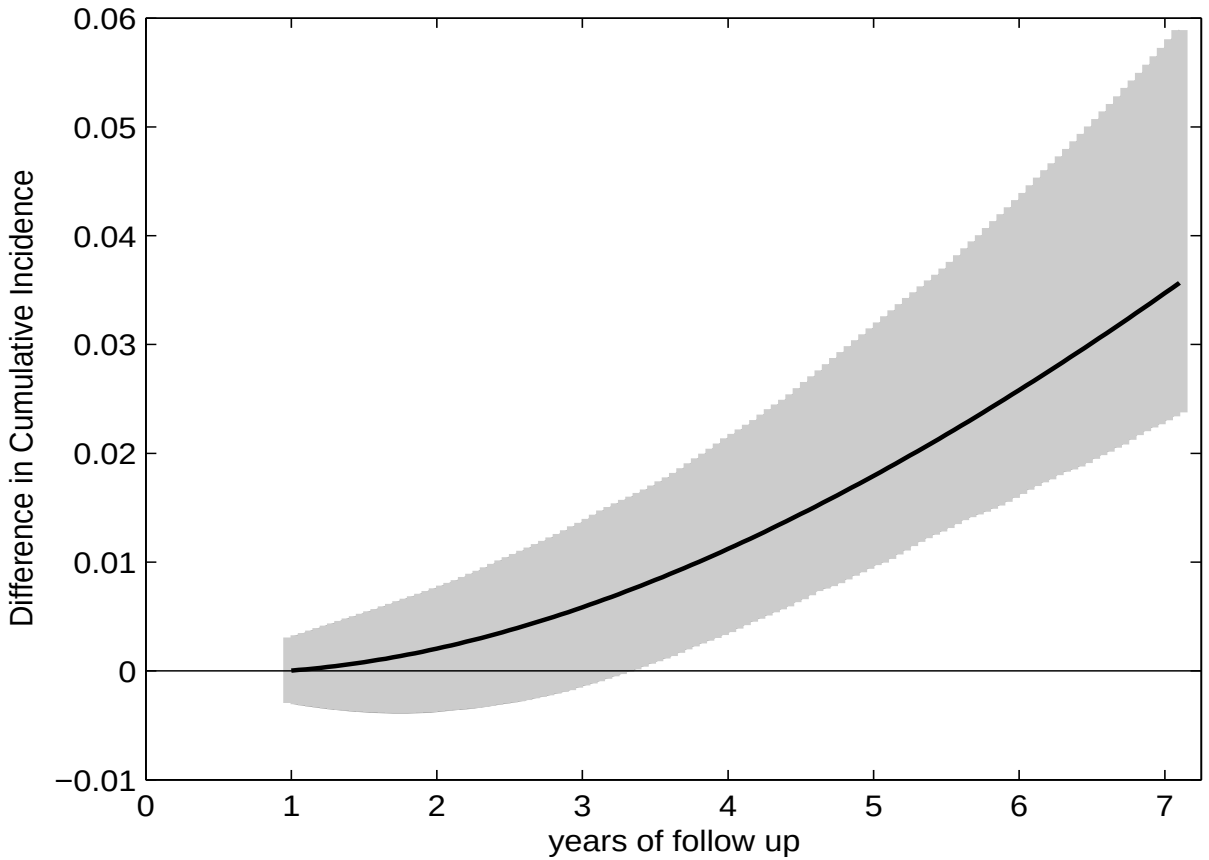
Hazards for T (irrespective of the path)



Overall hazard ratio function



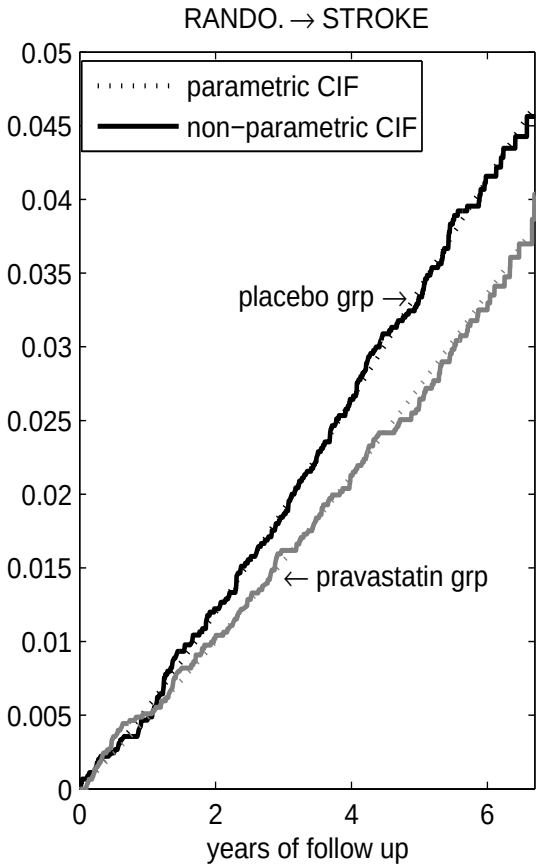
Absolute treatment benefit (Lagakos et al. NEJM 2006)

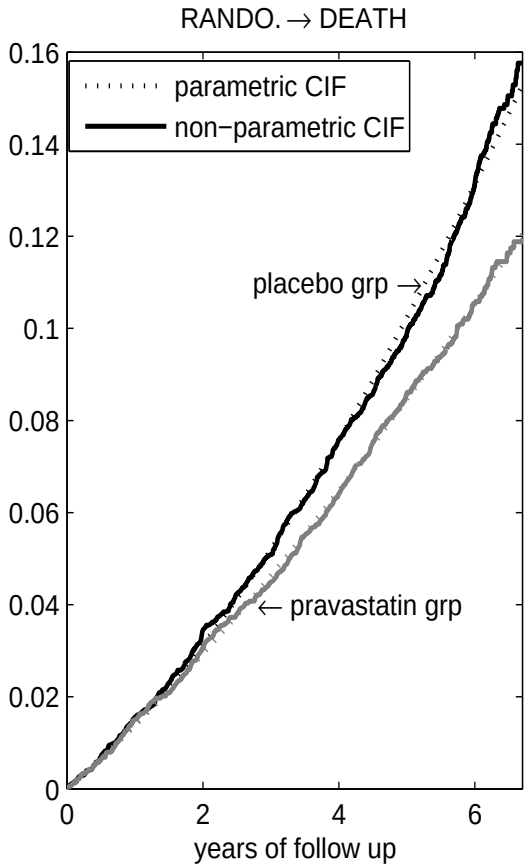


Link with CR

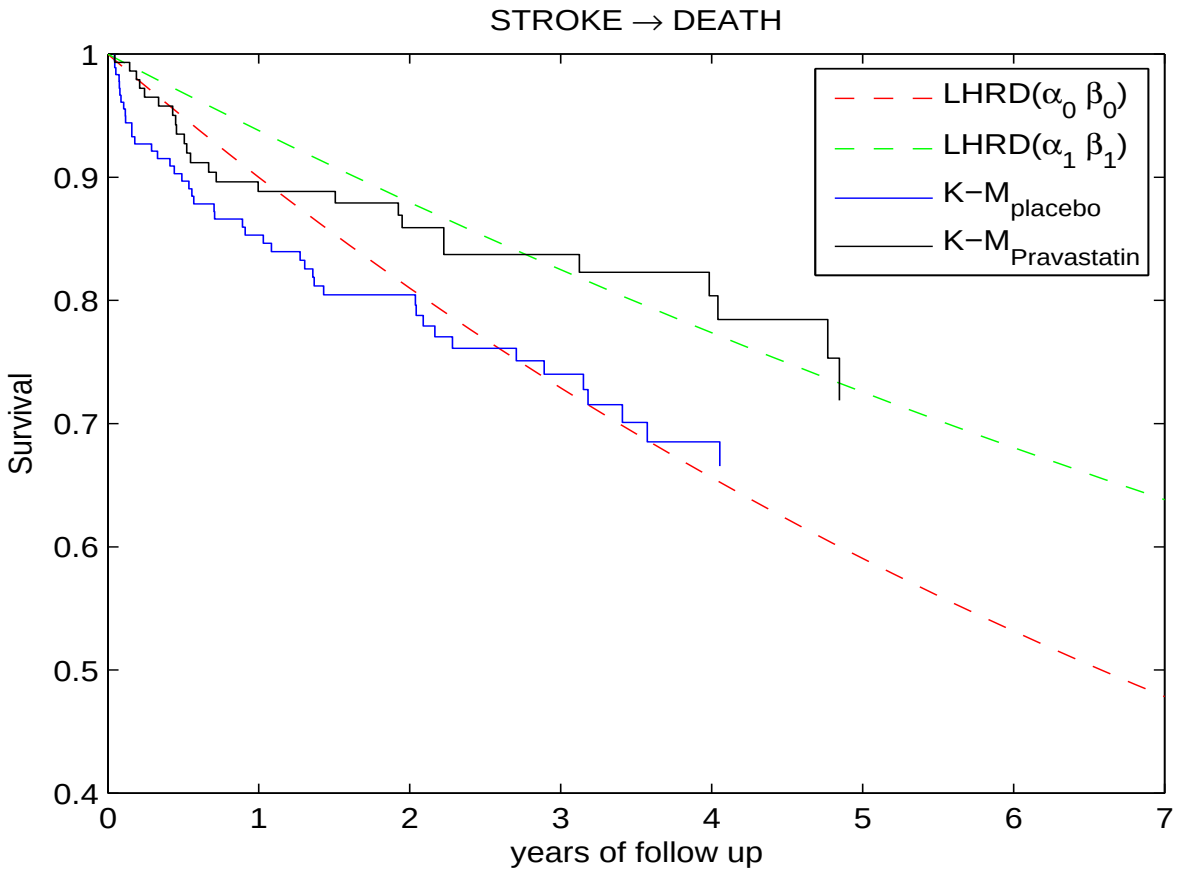
- Background:
- Semi-Markov Process (SMP)
- The model
- Data summary
- MGF and saddlepoint
- Likelihood function
- LHRD
- Results
- Assumptions
- Conclusions

Components $1 \longrightarrow 2$ and $1 \longrightarrow 3$

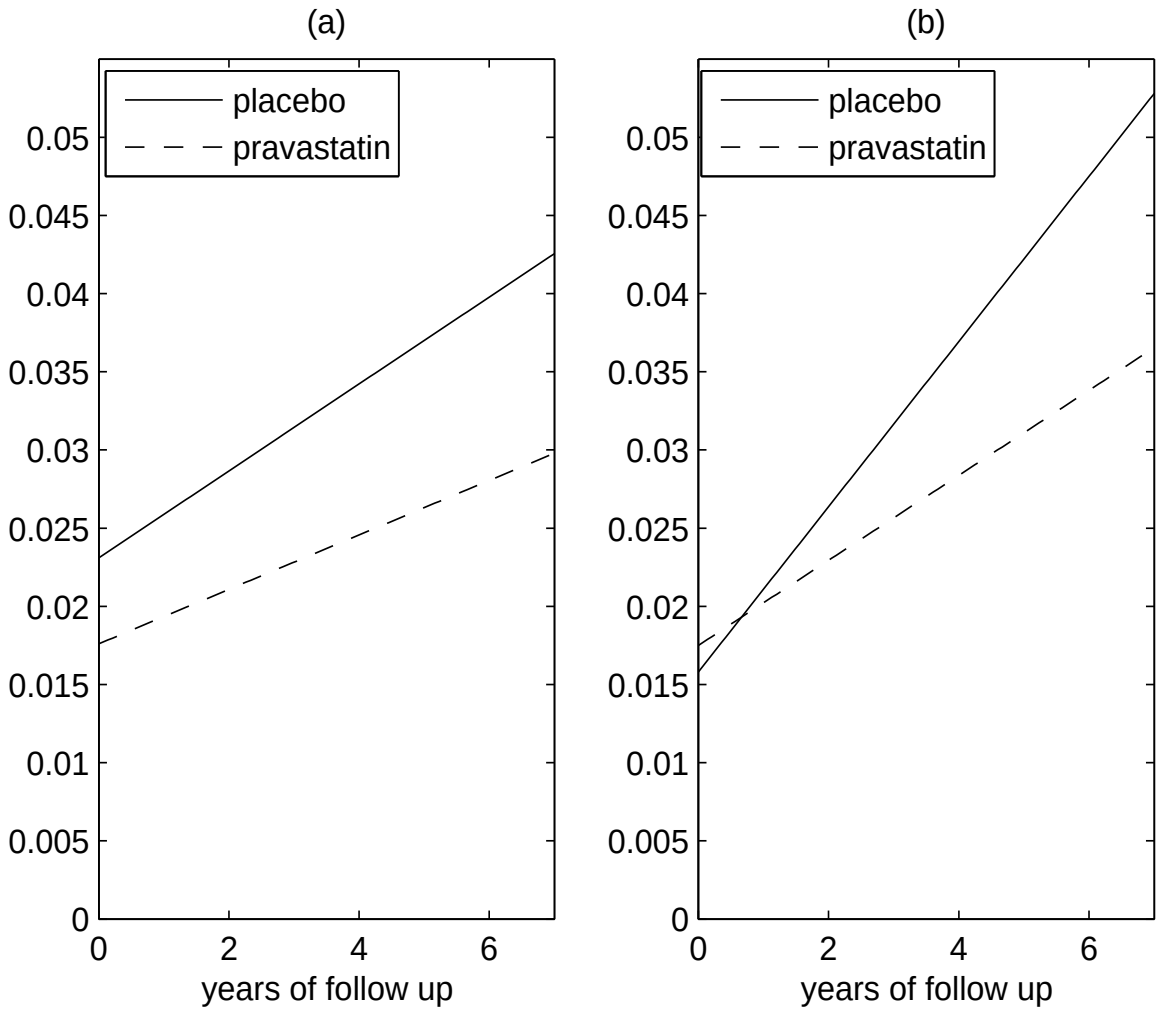




Component $2 \longrightarrow 3$



Cause-specific hazards



Assumptions

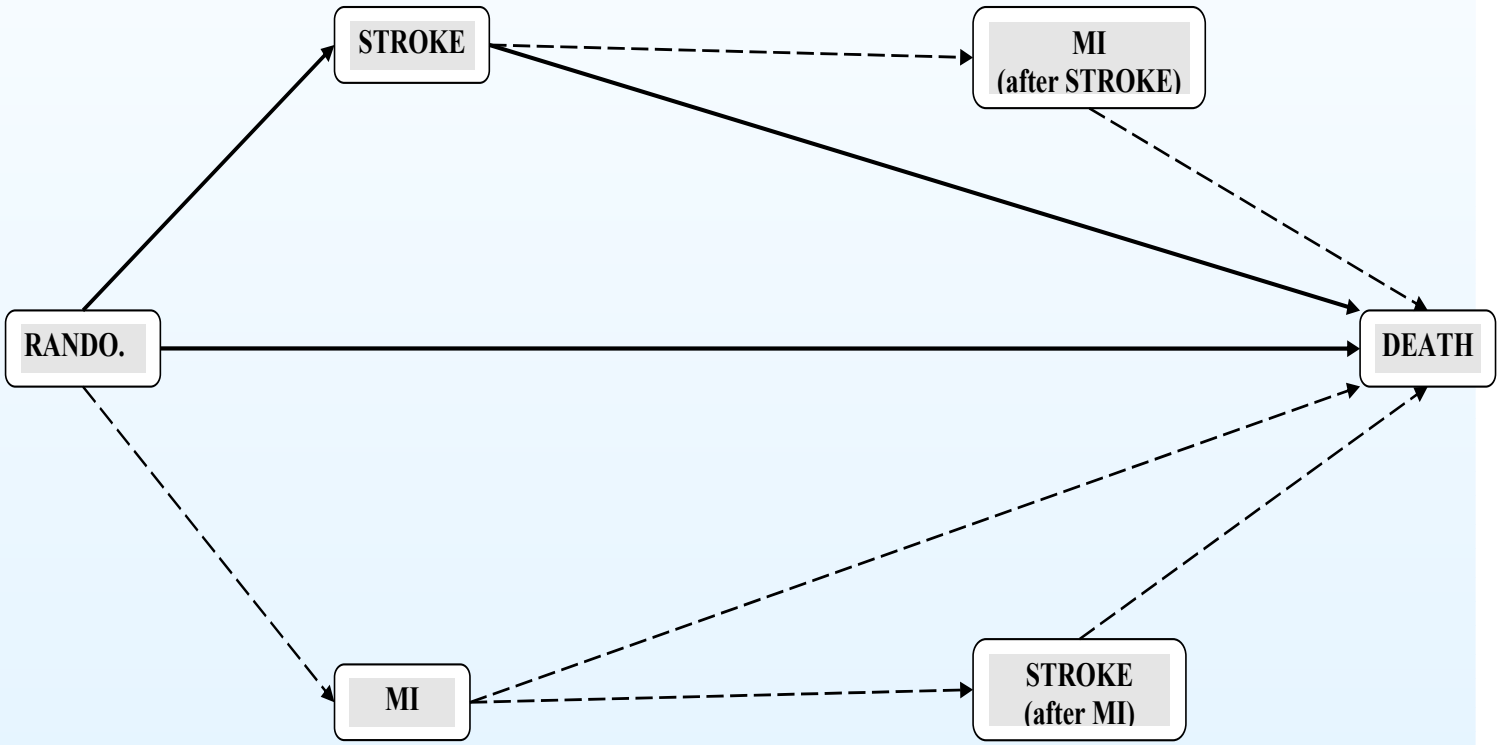
- Censored observations in a known stage have the same distribution of time to the next event as individuals not censored (**Uninformative censoring**)

- Information on all states visited

- Waiting times in different stages are **independent**

- Waiting times to future events depend only on the current health state, not on the history leading to this health state. (**lack of memory of past history**)

- **e.g. LIPID model based on the flowgraph below**



Conclusions

- We have used SP methods in a trial context to aggregate information from a structured flowgraph model.

Conclusions

- We have used SP methods in a trial context to aggregate information from a structured flowgraph model.
- By use of a SMP in LIPID data we found evidence of a **cumulating benefit of treatment** with time, not consistent with a fixed proportional hazard ratio model.

Conclusions

- We have used SP methods in a trial context to aggregate information from a structured flowgraph model.
- By use of a SMP in LIPID data we found evidence of a **cumulating benefit of treatment** with time, not consistent with a fixed proportional hazard ratio model.
- Standard approaches to survival analysis **had not identified this cumulative** effect.

Conclusions

- We have used SP methods in a trial context to aggregate information from a structured flowgraph model.
- By use of a SMP in LIPID data we found evidence of a **cumulating benefit of treatment** with time, not consistent with a fixed proportional hazard ratio model.
- Standard approaches to survival analysis **had not identified this cumulative** effect.
- **Interpretation** and **inference** is **possible** with SP methodology in flowgraph data for RCTs.

Conclusions

- We have used SP methods in a trial context to aggregate information from a structured flowgraph model.
- By use of a SMP in LIPID data we found evidence of a **cumulating benefit of treatment** with time, not consistent with a fixed proportional hazard ratio model.
- Standard approaches to survival analysis **had not identified this cumulative** effect.
- **Interpretation** and **inference** is **possible** with SP methodology in flowgraph data for RCTs.
- **Potential** in many applications *(models can be developed in a blinded fashion !)*

Future work

- Background:
- Semi-Markov Process (SMP)
- The model
- Data summary
- MGF and saddlepoint
- Likelihood function
- LHRD
- Results
- Assumptions
- Conclusions

- non-parametric version

Future work

- Background:
- Semi-Markov Process (SMP)
- The model
- Data summary
- MGF and saddlepoint
- Likelihood function
- LHRD
- Results
- Assumptions
- Conclusions

- non-parametric version
- adjustment for covariates

Future work

- non-parametric version
- adjustment for covariates
- other applications

Future work

- non-parametric version
- adjustment for covariates
- other applications
- 2+ final states (competing risks)

Future work

- non-parametric version
- adjustment for covariates
- other applications
- 2+ final states (competing risks)

- Background:
- Semi-Markov Process (SMP)
- The model
- Data summary
- MGF and saddlepoint
- Likelihood function
- LHRD
- Results
- Assumptions
- Conclusions

Open questions

- more complex models including
  - more stages ?
  - loops ?
  - competing risks ?
- Stroke - Death : time difference ?
- prediction beyond FU ?

Future work

- application to other trials
- health economics
- theory ?
- covariates
- non-parametric estimation of the MGF (with censoring)