A comparison between landmarking and joint modeling for producing predictions using longitudinal outcomes

Magdalena Murawska, Dimitris Rizopoulos, Eleni-Rosalina Andrinopoulou, Emmanuel Lesaffre and Johanna J.M. Takkenberg
Department of Biostatistics, Erasmus Medical Center

mmurawsk@iu.edu

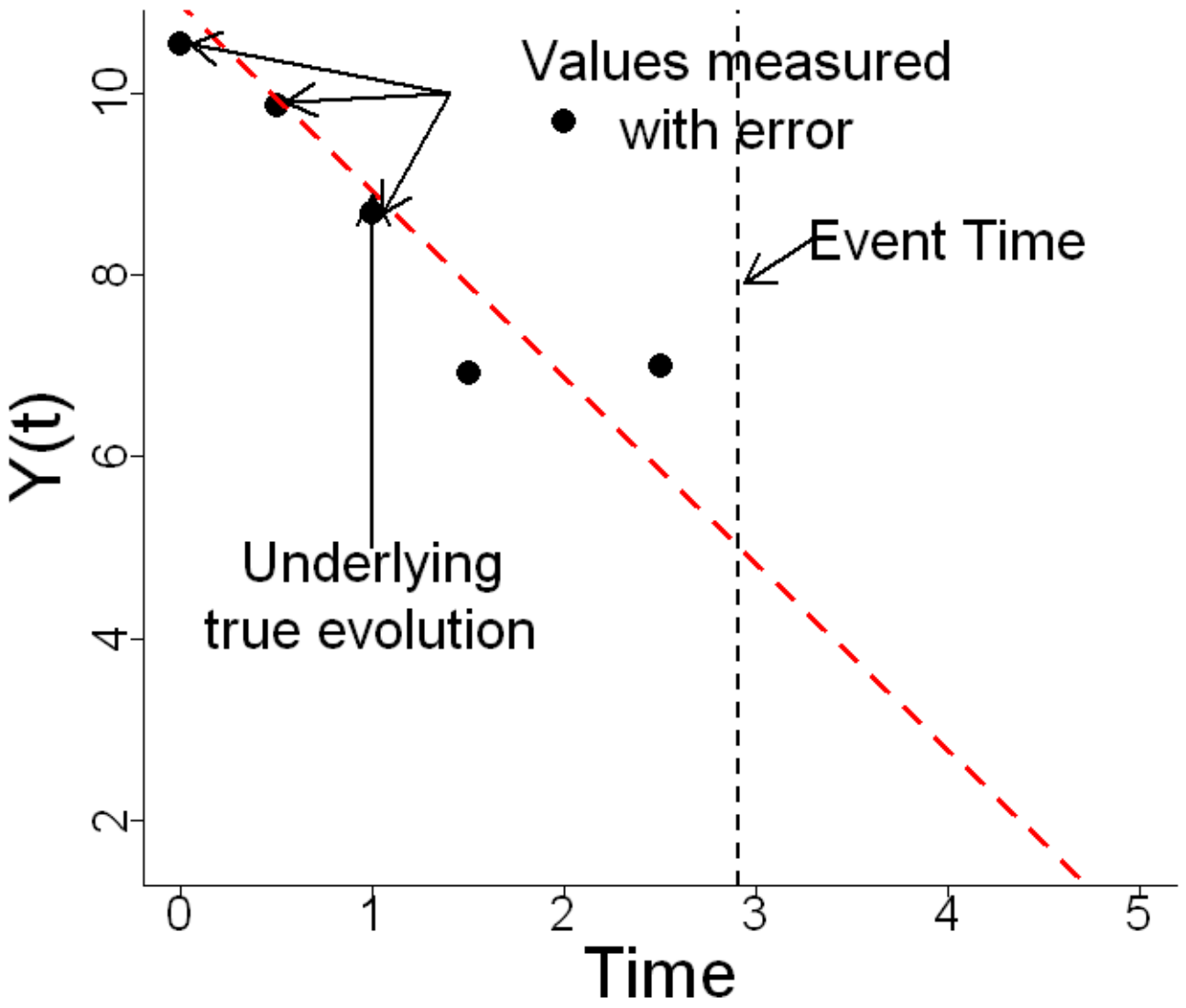
Indianapolis 2014, October 24, 2014

Dynamic Prediction

- Use repeated measurements of specific biomarkers to assess risk of death
- Example: CD4 in HIV study
- Dynamic prediction: update of survival probability as more measurements are available
- We compare two approaches for producing dynamic predictions of survival probabilities
  - landmarking (van Houwelingen and Putter, 2011)
  - joint modeling (Henderson et al.,2002, Yu et al., 2008, Rizopoulos, 2012)

Joint Modeling Approach

- Joint Modeling Approach:
  - reconstructs true evolution of biomarker
  - uses the true values of biomarker in survival model


Values measured
with error
Event Time
Underlying
true evolution
Y(t)
Time
10
8
6
4
2
0
1
2
3
4
5

Joint Modeling Approach

- Two submodels for longitudinal and survival processes
- For continuous longitudinal markers usually a linear mixed model is used:

$$y_i(t) = m_i(t) + \epsilon_i(t) = x_i^T(t)\beta + z_i^T(t)b_i + \epsilon_i(t)$$

$m_i(t)$ - true value of the longitudinal marker at time t

β - vector of the fixed-effects parameters

$b_i \sim N(0, D)$ -vector of random effects

$x_i(t)$ and $z_i(t)$ - design matrices for the fixed and random effects

$\epsilon_i(t)$ - measurement error, $\epsilon_i(t) \sim N(0, \sigma^2)$

Joint Modeling Approach

- For survival process standard relative risk model

  $\lambda_i(t) = \lambda_0(t) \exp(\alpha^T f(t, b_i) + \gamma^T v_i)$

- shares some common (time-dependent) term $f(t, b_i)$, with longitudinal model

  v_i - vector of baseline covariates, γ - vector of associated coefficients

  α - measure the strength of association between longitudinal and survival processes

Joint Modeling Approach

- Based on fitted model dynamic predictions for new subject k constructed
- We predict conditional probability of surviving time $u > t$ given that subject k has survived up to t:

$$S_k(u \mid t) = \mathsf{Pr}(T_k^* > u \mid T_k^* > t, Y_k(t))$$

  $Y_k(t)$ - longitudinal profile for subject k at time t, T^*- true survival time

- $S_k(u \mid t)$ can be written as Bayesian posterior expectation:

$$S_k(u \mid t) = \int \mathsf{Pr}\left(T_k^* > u \mid T_k^* > t, Y_k(t), \mathcal{S}_n; \theta\right) p(\theta \mid \mathcal{S}_n) d\theta \qquad (*)$$

  θ - vector of parameters from joint model, $\mathcal{S}_n$ - a sample of size n on which joint model was fitted

Joint Modeling Approach

- Let $f(b_i, t) = b_i$. First part of the integrant (*) can be written as:

$$\Pr\left(T_k^* > u \mid T_k^* > t, Y_k(t), \mathcal{S}_n; \theta\right)$$
$$= \int \Pr\left(T_k < u \mid T_k^* > t, b_k; \theta\right) \times p\left(b_k \mid T_k^* > t, Y_k(t), \theta\right) db_k$$

- Monte Carlo approach used to compute $S_k(u \mid t)$ for patient k and $S_k(u \mid t')$ updated for every time point $t' > t$

Joint Modeling Approach

- For each individual k given available longitudinal profile $Y_k(t)$:
- Step 1: sample $b_k^{(l)}$ from posterior $\{b_k \mid \mathcal{T}_k^*(t), Y_k(t); \theta\}$

Joint Modeling Approach

- For each individual k given available longitudinal profile $Y_k(t)$:
- Step 1: sample $b_k^{(l)}$ from posterior $\{b_k \mid \mathcal{T}_k^*(t), Y_k(t); \theta\}$
- Step 2: compute $S_k^{(l)}(u \mid t, b_k^{(l)})$

Joint Modeling Approach

- For each individual k given available longitudinal profile $Y_k(t)$:
- Step 1: sample $b_k^{(l)}$ from posterior $\{b_k \mid \mathcal{T}_k^*(t), Y_k(t); \theta\}$
- Step 2: compute $S_k^{(l)}(u \mid t, b_k^{(l)})$
- Repeat Steps 1-2, $l = 1, \ldots L$

Joint Modeling Approach

- For each individual k given available longitudinal profile $Y_k(t)$:
- Step 1: sample $b_k^{(l)}$ from posterior $\{b_k \mid \mathcal{T}_k^*(t), Y_k(t); \theta\}$
- Step 2: compute $S_k^{(l)}(u \mid t, b_k^{(l)})$
- Repeat Steps 1-2, $l = 1, \ldots L$
- Use median (and quantiles) of $S_k^{(l)}(u \mid t, b_k^{(l)})$

Joint Modeling Approach

- For each individual k given available longitudinal profile $Y_k(t)$:
- Step 1: sample $b_k^{(l)}$ from posterior $\{b_k \mid \mathcal{T}_k^*(t), Y_k(t); \theta\}$
- Step 2: compute $S_k^{(l)}(u \mid t, b_k^{(l)})$
- Repeat Steps 1-2, $l = 1, \ldots L$
- Use median (and quantiles) of $S_k^{(l)}(u \mid t, b_k^{(l)})$

Landmark Approach

- Landmark method simplifies the longitudinal history $Y_k(t)$ to the last value $y_k(t)$
- Dynamic predictions obtained by adjusting the risk set and refitting Cox model:
  - landmark time t_L chosen
  - for t_L landmark data set $\mathcal{L}_L$ constructed: selecting individuals at risk at t_L
  - Cox model fitted for $\mathcal{L}_L$
- Advantage of JM approach: possibility of defining different association structure between longitudinal and survival processes

Motivating Data set

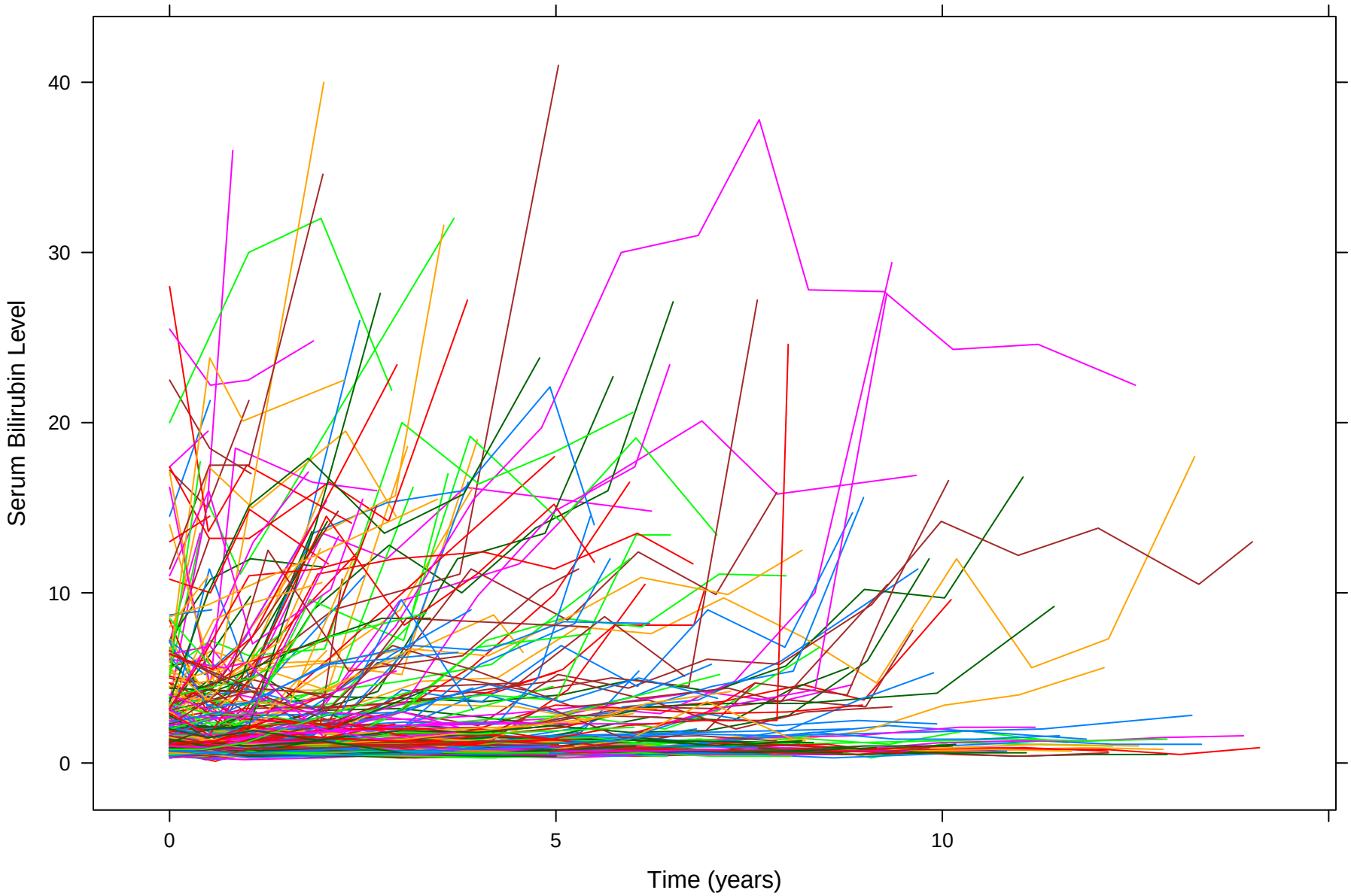
- PBC study

  conducted by Mayo Clinic between 1974 and 1984

- For patients with PBC serum bilirubin is known to be a good marker of progression

- Aim: find which characteristics of serum bilirubin profile are most predictive for death

- Longitudinal serum bilirubin level $Y_i(u)$ modeled by mixed effects model
  - natural cubic splines to account for nonlinear character of marker evolution
  - interaction terms between B-spline basis and treatment group to model different trajectories for 2 treatment groups


Serum Bilirubin Level
0
10
20
30
40
0
5
10
Time (years)

Motivating Data set

- For survival process standard relative risk model with different forms of the association structure:

$$
\begin{aligned}
\text{I} \quad \lambda_i(t) &= \lambda_0(t)\exp\{\gamma^T v_i + \alpha_1 m_i(t)\} \\
\text{II} \quad \lambda_i(t) &= \lambda_0(t)\exp\{\gamma^T v_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\} \\
\text{III} \quad \lambda_i(t) &= \lambda_0(t)\exp\left\{\gamma^T v_i + \alpha_1 \int_0^t m_i(s)ds\right\} \\
\text{IV} \quad \lambda_i(t) &= \lambda_0(t)\exp\{\gamma^T v_i + \alpha^T b_i\}.
\end{aligned}
\tag{1}
$$

Baseline hazard $\lambda_0(t)$ modeled parametrically using Weibull distribution, i.e: $\lambda_0(t) = \phi t^{\phi-1}$

Different parameterizations of Joint Model

- In J-M where only random effects are shared likelihood is of the (closed!) form:

$$p(T_i, \Delta_i \mid b_i, \theta, \beta) = \left[\lambda_0(T_i)\exp(\alpha^T b_i + \gamma^T v_i)\right]^{I(\Delta_i=1)} \times$$

$$\exp\left(-\int_0^{T_i} \lambda_0(s)\exp(\alpha^T b_i + \gamma^T v_i)ds\right)$$

  ▷ Dependence on s only through piecewise constant baseline hazards $\lambda_0(s)$

- Problem arises when time-dependent term shared:

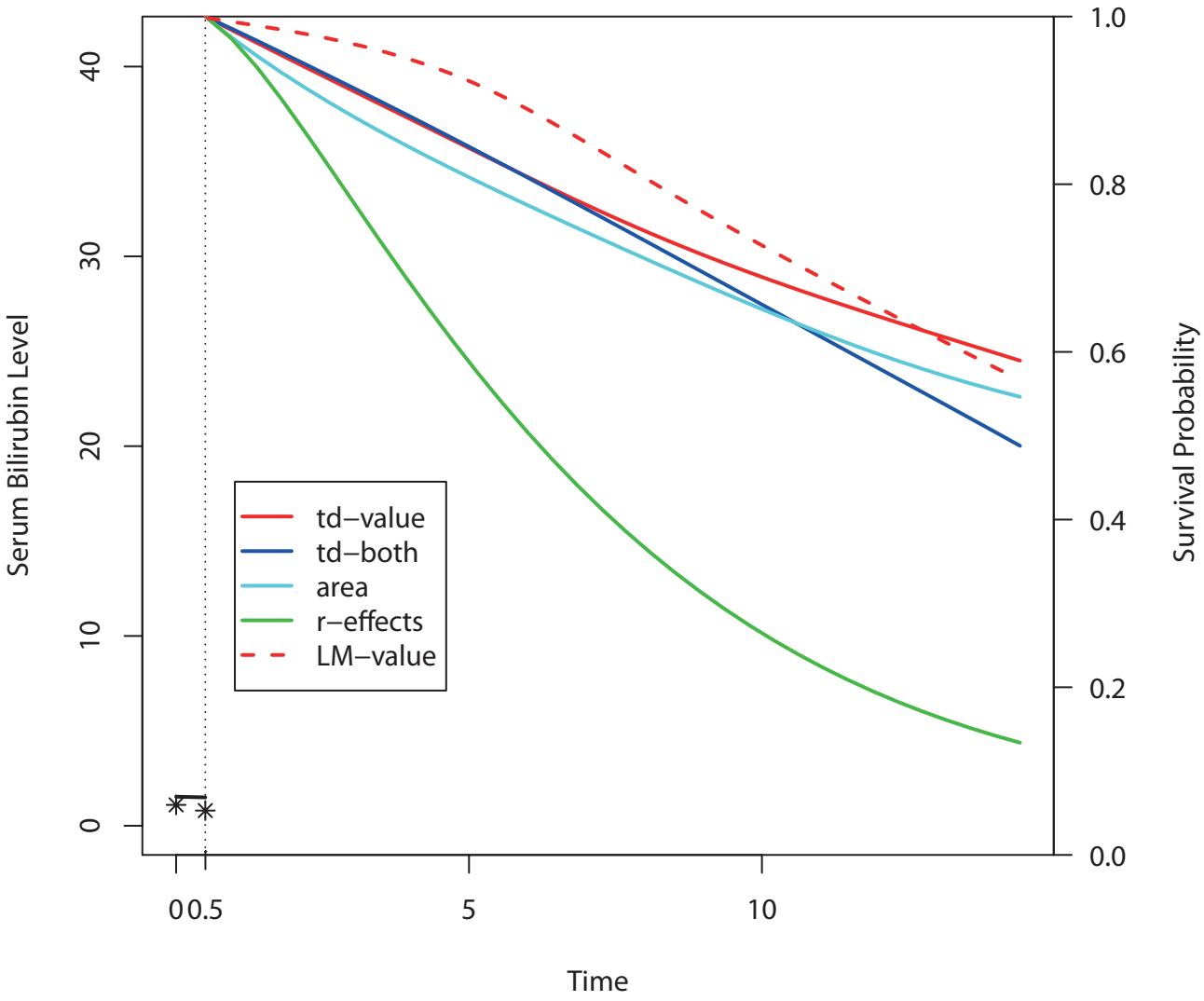
$$\int_0^{T_i} \lambda_0(s)\exp(\alpha^T f_i(s) + \gamma^T v_i)ds$$

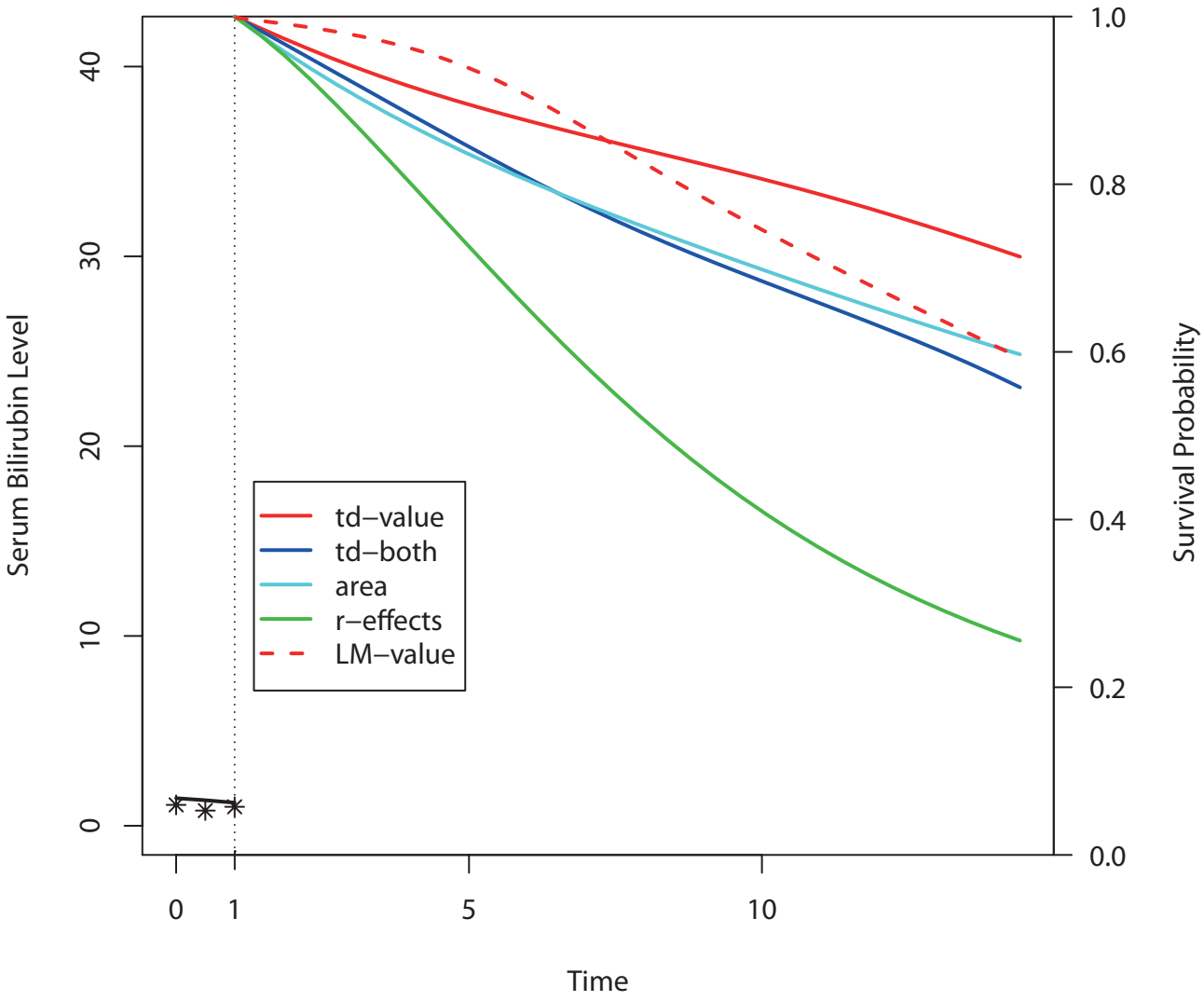
  ▷ Solution: use quadrature points to approximate the integral

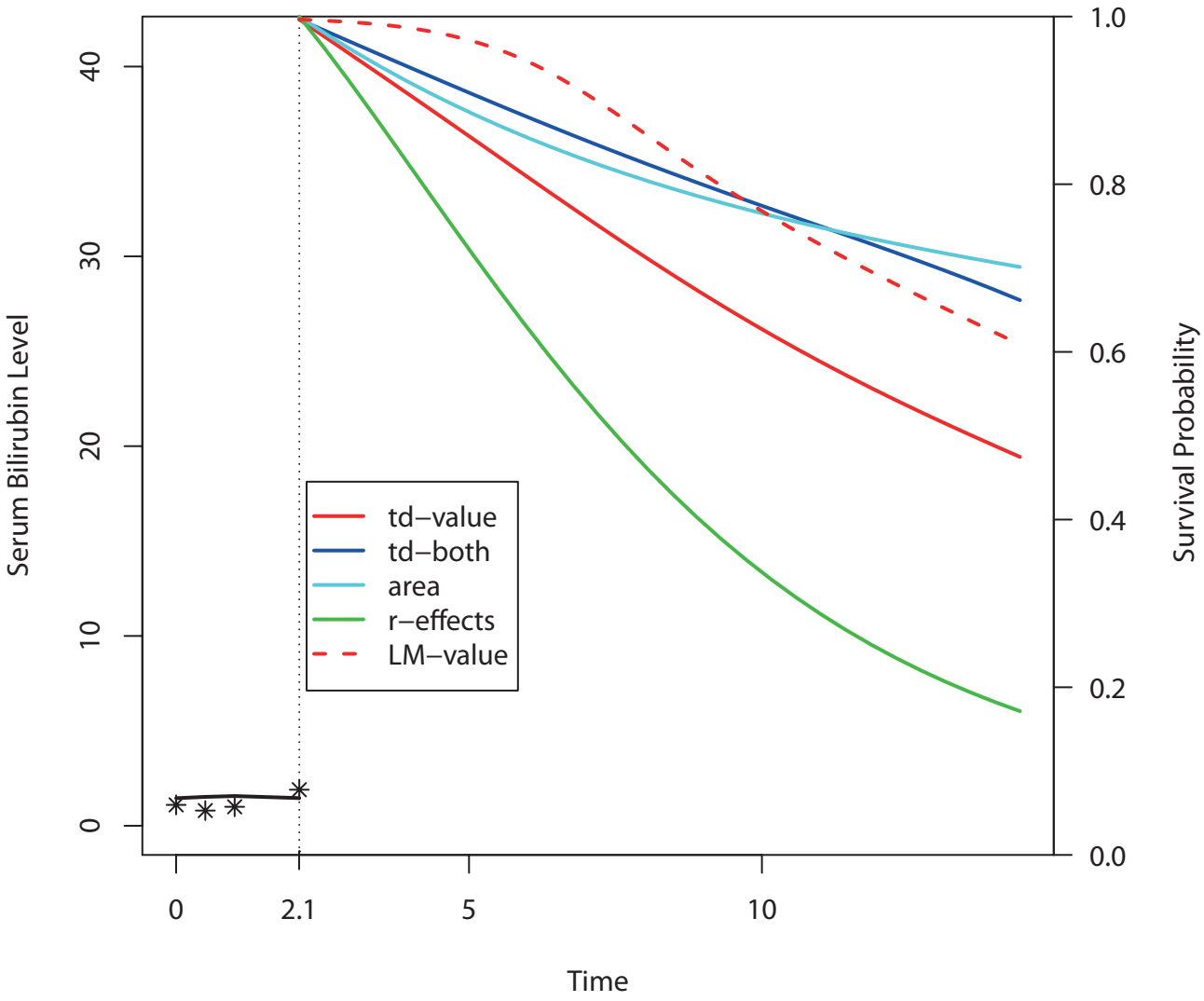
PBC Data

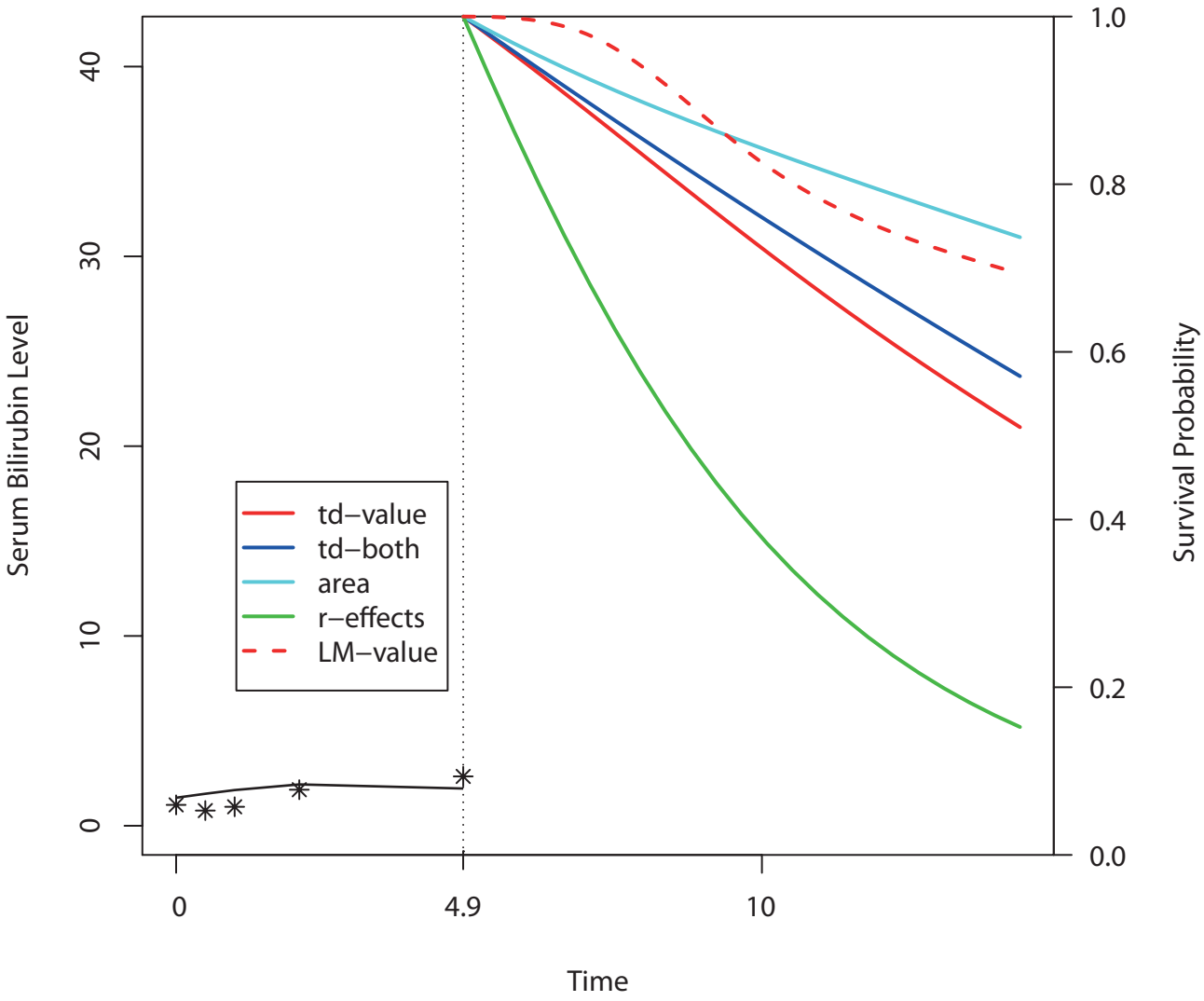
- Differences between prediction from joint models I-IV and landmark approach observed

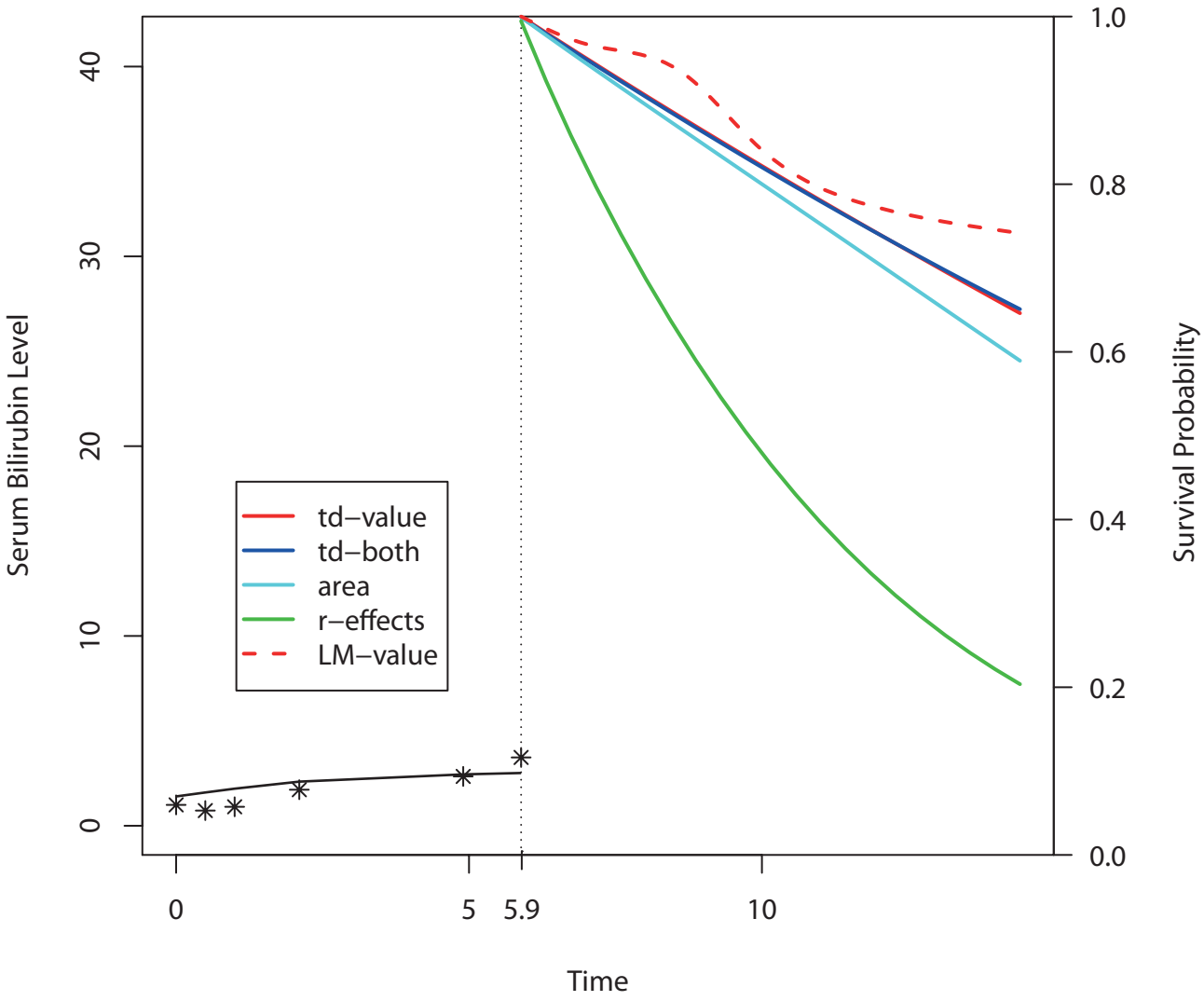
- Different joint models compared using DIC criterion $\rightarrow$ best Model I (td-value)


Serum Bilirubin Level
0
10
20
30
40
Survival Probability
0.0
0.2
0.4
0.6
0.8
1.0
0 0.5
5
10
Time
td−value
td−both
area
r−effects
LM−value


Serum Bilirubin Level
0
10
20
30
40
Survival Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
1
5
10
Time
td−value
td−both
area
r−effects
LM−value

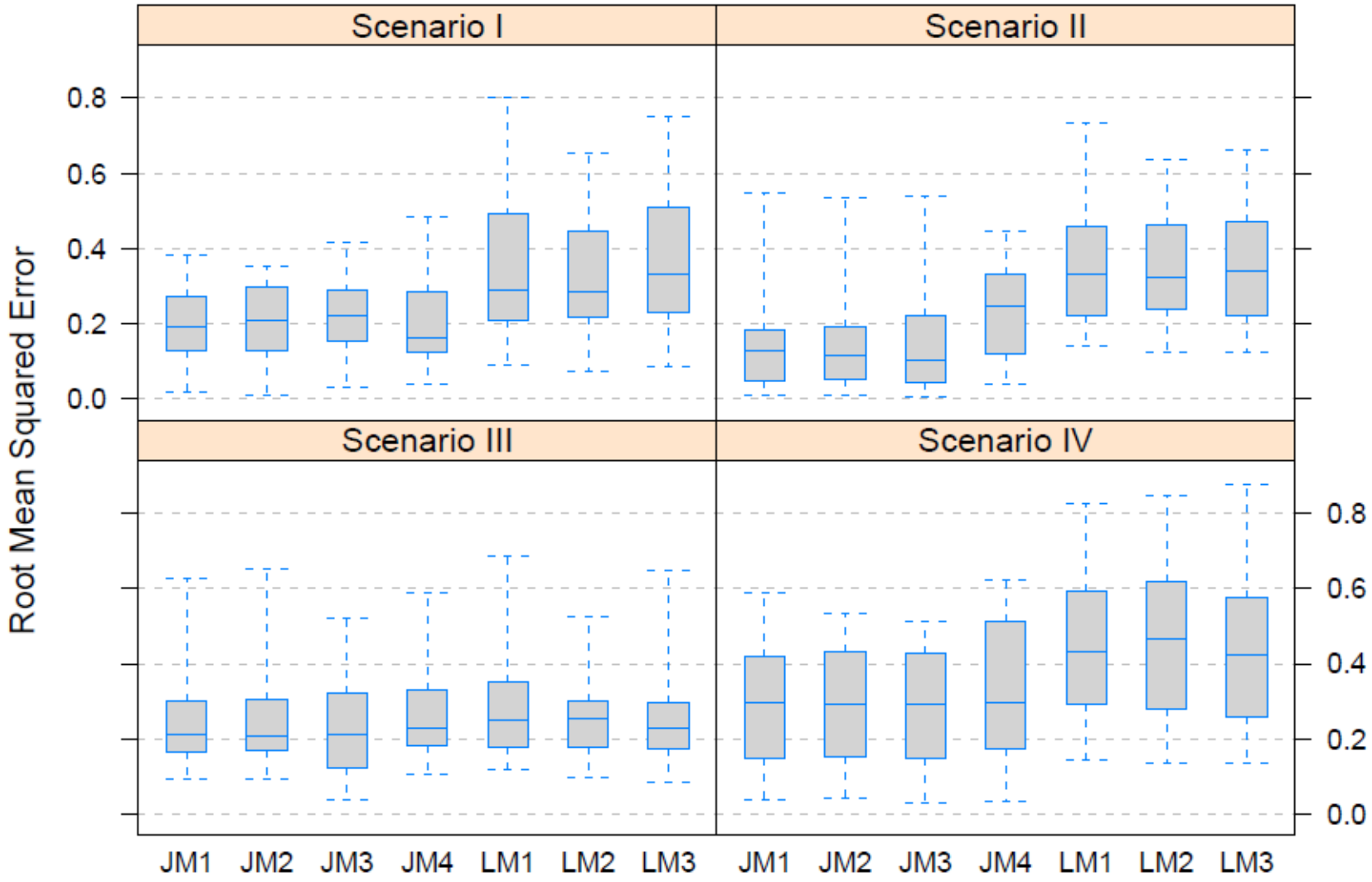

Serum Bilirubin Level
Survival Probability
Time
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
0
2.1
5
10
td−value
td−both
area
r−effects
LM−value

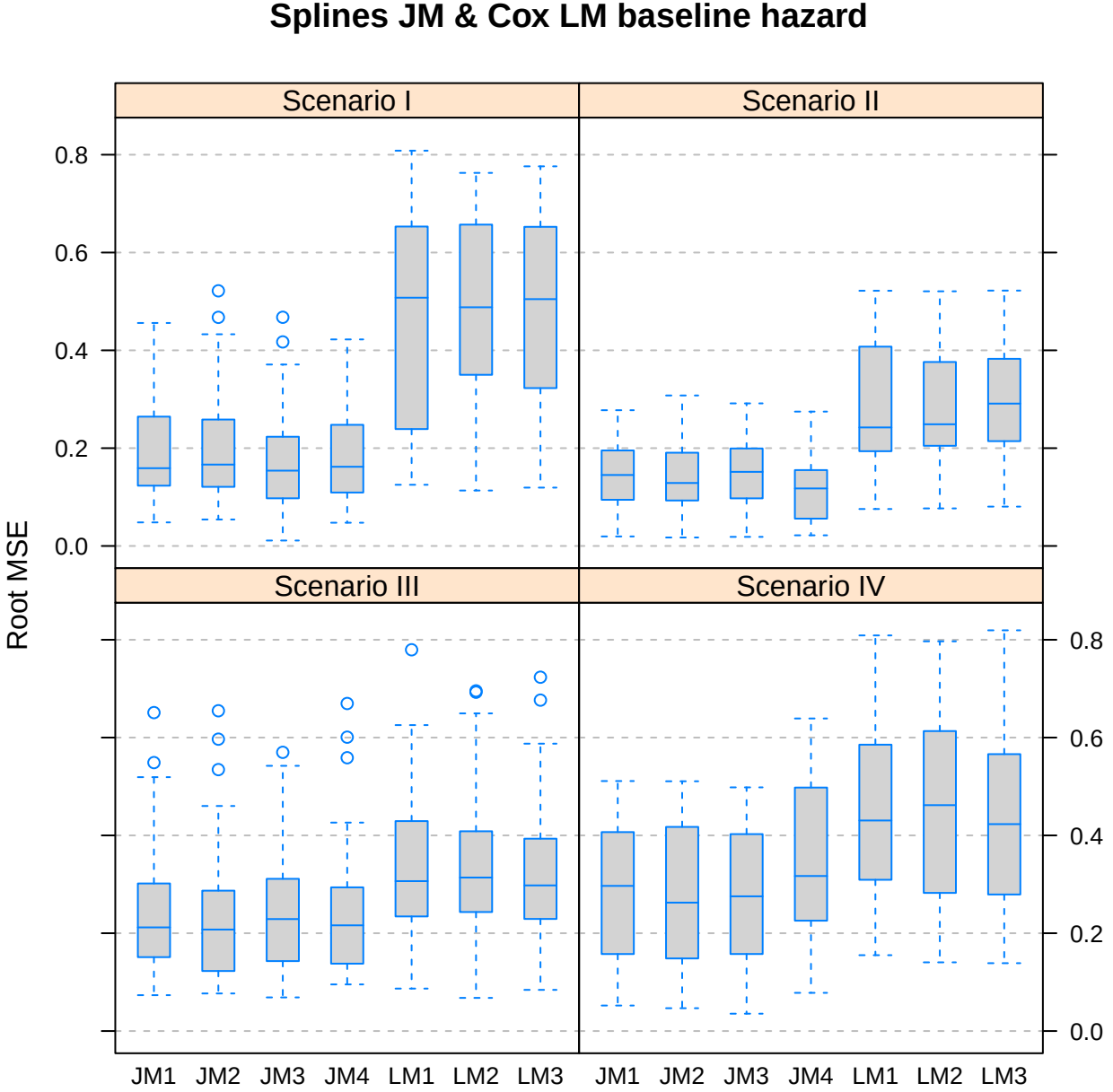

Serum Bilirubin Level
Survival Probability
Time
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
0
4.9
10
td−value
td−both
area
r−effects
LM−value


Serum Bilirubin Level
Survival Probability
Time
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
0
5
5.9
10
td−value
td−both
area
r−effects
LM−value

Simulation Study

- Data simulated data using joint models with different association structure I-IV
  - Baseline hazard simulated using Weibull distribution
  - Censoring kept at 40-50%
- In each scenario 10 censored pts excluded randomly from each simulated data set
- For remaining patients joint models I-IV fitted
- For excluded patients predictions from joint models I-IV and landmarking compared at 10 equidistant time points to predictions from gold standard model (model with true parametrization and true values of parameters)
- Standard landmark model extended: current value+slope (LM2), current value+area (LM3)


Scenario I
Scenario II
Scenario III
Scenario IV
Root Mean Squared Error
0.0
0.2
0.4
0.6
0.8
JM1
JM2
JM3
JM4
LM1
LM2
LM3


Splines JM & Cox LM baseline hazard
Scenario I
Scenario II
Scenario III
Scenario IV
Root MSE
0.0
0.2
0.4
0.6
0.8
JM1
JM2
JM3
JM4
LM1
LM2
LM3

Discrimination. Calibration

- Compare calibration and discrimination between two approaches in a simulation study using:
  - Expected Prediction Error (Henderson et al 2002) (PE)
  - Integrated Prediction Error (Schemper and Henderson 2000) (IPE)
  - AUC and dynamic concordance index $C_{dyn}^{\Delta t}$

Discrimination

- Focus on time interval when the occurence of event is of interest $(t, t + \Delta t]$
- Based on the model we would like to dicriminate between patients who are going to exprience the event in that interval from patients who will not
- For the first group physiscian can take action to improve survival during $(t, t + \Delta t]$
- For c in $[0, 1]$ we define $S_k(u \mid t) \leq c$ as success and $S_k(u \mid t) > c$ as failure
- Then sensitivity is defined as:

$$\Pr\{S_k(u \mid t) \leq c \mid T_k^* \in (t, t + \Delta t]\}$$

- And specificity as:

$$\Pr\{S_k(u \mid t) > c \mid T_k^* > t + \Delta t\}$$

Discrimination

- For random pair of subjects i, j that have measurments up to t discrimination capability of joint model can be assesed by area under ROC curve (AUC) obtained by varying c:

$$AUC(t, \Delta t) = \mathsf{Pr}[S_i(u \mid t) < S_j(u \mid t) \mid \{T_i^* \in (t, t + \Delta t]\} \cup \{T_j^* > t + \Delta t\}]$$

- Model will assign higher probability of surviving longer that $t + \Delta t$ for subject j who did not experience event

- To summarize model discrimination power weigthed average of AUCs used:

$$\mathsf{C}_{dyn}^{\Delta t} = \int_0^\infty AUC(t, \Delta t\} \mathsf{Pr}\{\mathcal{E}(t)\} dt \Big/ \int_0^\infty \mathsf{Pr}\{\mathcal{E}(t)\} dt \text{ (dynamic concordance index)}$$

$$\mathcal{E}(t) = [\{T_i^* \in (t, t + \Delta t]\} \cup \{T_j^* > t + \Delta t\}]$$

$\mathsf{Pr}\{\mathcal{E}(t)\}$-probability that pair $\{i, j\}$ comparable at t

Discrimination

- $\mathsf{C}_{dyn}^{\Delta t}$ depends on Δt
- In practice:

$$\hat{\mathsf{C}}_{dyn}^{\Delta t} = \frac{\sum_{q=1}^{15} \omega_q A\hat{U}C(t_q, \Delta t) \times \hat{\mathsf{Pr}}\{\mathcal{E}(t_q)\}}{\sum_{q=1}^{15} \omega_q \hat{\mathsf{Pr}}\{\mathcal{E}(t_q)\}}$$

ω_q-weights for 15 Gauss-Kronrod quadrature points on $(0, t_{max})$

$\hat{\mathsf{Pr}}\{\mathcal{E}(t_q)\} = \{\hat{S}(t_q) - \hat{S}(t_q + \Delta t)\}\hat{S}(t_q + \Delta t)$

$\hat{S}(\cdot)$-Kaplan-Meier estimator of marginal survival function $S(\cdot)$

Discrimination

- AUC is estimated as:

$$\hat{AUC}(t_q, \Delta t) = \frac{\sum_{i=i}^{n} \sum_{j=1, j \neq i}^{n} I\{\hat{S}_i(t + \Delta t \mid t) < \hat{S}_j(t + \Delta t \mid t)\} \times I\{\Omega_{ij}(t)\}}{I\{\sum_{i=i}^{n} \sum_{j=1, j \neq i}^{n} \Omega_{ij}(t)\}}$$

- Comparable pairs are those that satisfy:

$\Omega_{ij}(t) = [\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 1\}] \cap \{T_j > t + \Delta t\}$ or

$\Omega_{ij}(t) = [\{T_i \in (t, t + \Delta t]\} \cap \{\delta_i = 1\}] \cap [\{T_j = t + \Delta t\} \cap \{\delta_j = 0\}]$

Calibration

- Expected Prediction Error (Henderson et al 2002):

$$PE(u \mid t) = E[L\{N_i(u) - S_i(u \mid t)\}]$$

$$N_i(u) = I(T_i^* > u)$$

$L(\cdot)$-loss function (absolute or square loss)

$$\hat{PE}(u \mid t) = \{\mathcal{R}(t)\}^{-1} \sum_{i:T_i \geq t} I(T_i > u)L\{1 - \hat{S}(u \mid t)\} + \delta_i I(T_i < u)L\{0 - \hat{S}(u \mid t)\}$$
$$+(1 - \delta_i)I(T_i < u)[\hat{S}_i(u \mid T_i)L\{1 - \hat{S}(u \mid t)\} + \{1 - \hat{S}(u \mid T_i)\}L\{0 - \hat{S}(u \mid t)\}]$$

$\mathcal{R}(t)$-number of subjects at risk at t

Calibration

- $PE(u \mid t)$ measures predictive accuracy only at u using longitudinal information up to time t
- To summarize predictive accuracy for interval $[t, u]$ and take into account censoring weighted average of $PE(s \mid t)$, $t < s < u$ considered, similar to $\hat{\mathrm{C}}_{dyn}^{\Delta t}$
- Integrated Prediction Error (Schemper and Henderson 2000):

$$IPE(u \mid t) = \frac{\sum_{i:u \leq T_i \leq t} \delta_i \{\hat{S}_C(t)/\hat{S}_C(T_i)\} \hat{PE}(u \mid t)}{\sum_{i:u \leq T_i \leq t} \delta_i \{\hat{S}_C(t)/\hat{S}_C(T_i)\}}$$

$\hat{S}_C(\cdot)$- Kaplan-Meier estimator of censoring distribution

	$\widehat{\text{PE}}(9\|7)$	$\widehat{\text{IPE}}(9\|7)$	$\widehat{\text{AUC}}(9\|7)$	$\widehat{\text{C}}_{dyn}^{\Delta t=2}$
JM_1: value	0.201	0.118	0.787	0.854
JM_2: value+slope	0.197	0.114	0.793	0.855
JM_3: area	0.191	0.112	0.758	0.809
JM_4: shared RE	0.191	0.108	0.807	0.840
Cox_{LM}	0.229	0.130	0.702	0.811

- Results for PBC data set will indicate different best model than DIC

Extensions

- Different types of longitudinal outcome (binary, categorical)
- Multiple longitudinal outcomes
- Multiple event times (Competing risk setting)

Motivating Data Set 2 : Heart Data

- Data from Eurotransplant Heart recipient waiting list (2921 recipients)
- During follow-up patients are evaluated as:
  - ▷ Transplantable (T)
  - ▷ Urgent (U)
  - ▷ High-Urgent (HU)
  - ▷ Non-Transplantable (NT)
- Patient is excluded from the list when:
  - ▷ Death (D)
  - ▷ Transplanted (TT)
  - ▷ Removed (from other reasons than transplantation) (R)

Heart Data cont.

- Different evaluation points
  - ▷ First evaluation time point at the moment of entering on the waiting list (time 0)
  - ▷ Next evaluation time points depend on the previous state
- At baseline (time 0) patient characteristics available:
  - ▷ age
  - ▷ country : 7 centers categorized in IConsent and Non-IConsent
  - ▷ blood group (A, B, AB, 0)
- **Aim**: predict patient's urgency status and asses risk of D/R/TT

  using available history & adjusting for baseline covariates

Joint Modeling Approach

- Modeling transient states : U, HU, T and NT as categorical longitudinal response
- Modeling the risk of final events: R, D or TT
- Categorical response cannot be ordered (due to NT state)
- Competing risks (D,TT,R)
- Similar procedure as above to update conditional CIF dynamically

Joint Model (J-M). Submodels specification

- Longitudinal submodel:

  multinomial logit mixed model to model probabilities of states $s = U, HU, T, NT$

  $$logit(P(Y_i(t) = s_r)) = x_i^T(t)a_r + z_i^T(t)b_{ir},\ r = 1, 2, \ldots, R-1, \quad i = 1, 2, \ldots N$$

  $$b_{ir}^T = (b_{i1}^T, b_{i2}^T, \ldots, b_{ir}^T),\ b_{ir} \sim N(0, \Sigma_r)$$

  $x_i(t)$ -vector of covariates

  $z_i(t)$ - design vector for random effects

Joint Model. Submodels specification

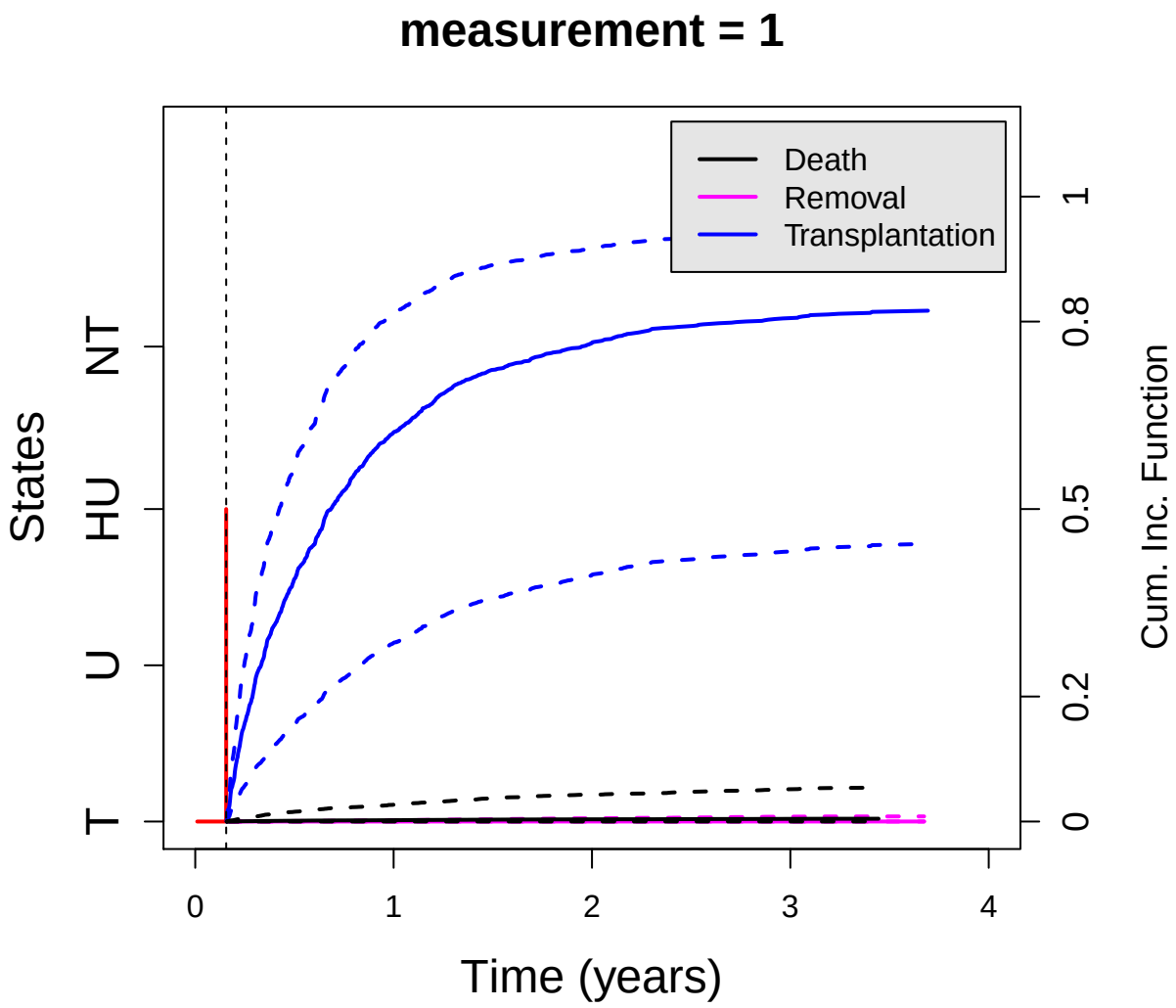
- Let $T_{i1}^*, T_{i2}^*, \ldots, T_{iK}^*$ - true failure times for individual i
- We observe only $T_i = \min(T_{i1}^*, T_{i2}^*, \ldots, T_{iK}^*, C_i)$, C_i -censoring time, Δ_i -failure ind.
- Relative risk submodel for each cause of failure k:

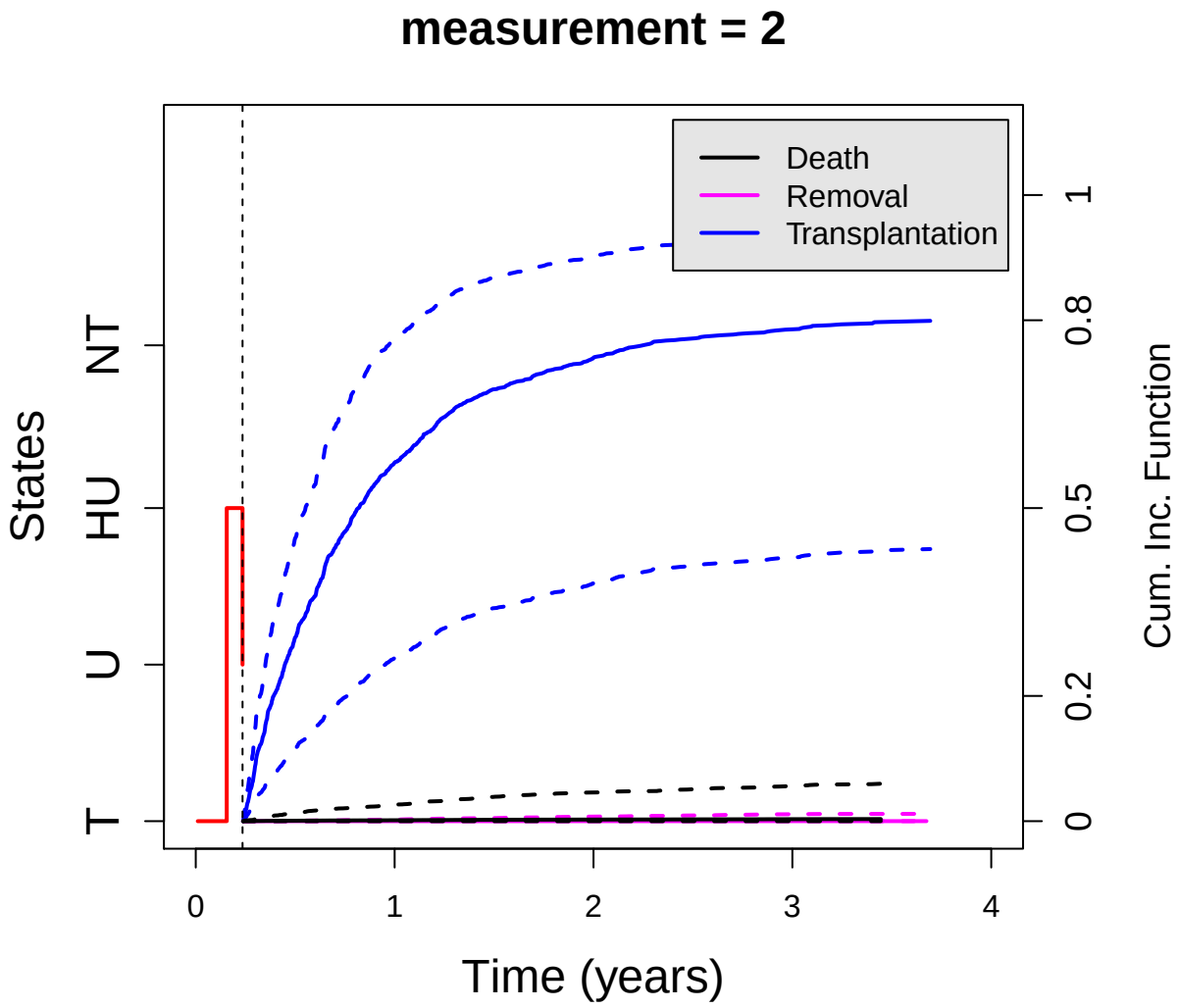
$$\lambda_{ik}(t) = \lim_{s \to 0} \mathsf{P}(t \leq T_i^* < t + s, \Delta_i = k \mid T_i^* \geq t)/s =$$

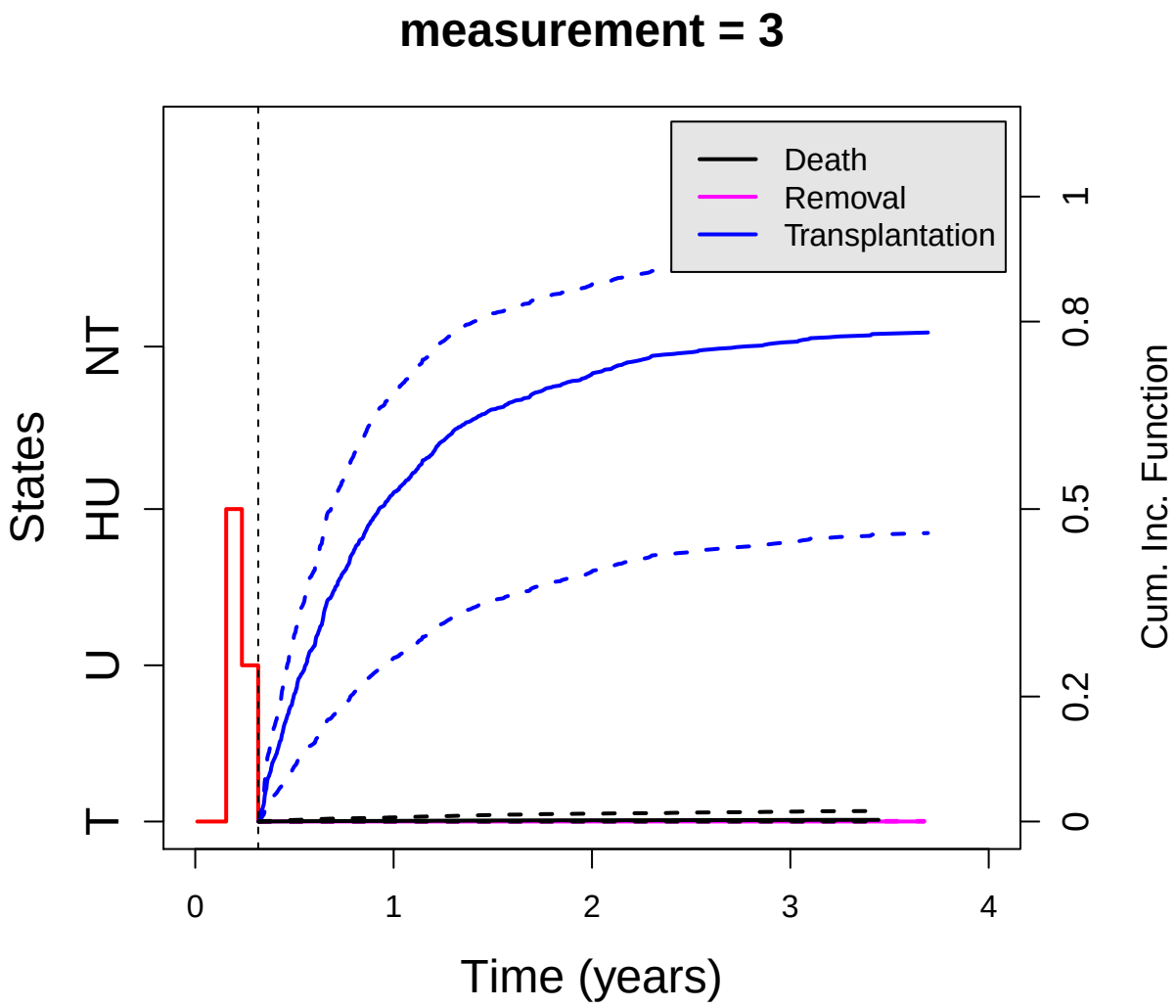
$$= \lambda_{0k}(t) \exp(\gamma_k^T b_i + \beta_k^T v_i),\ k = 1, \ldots, K, \quad b_i^T = (b_{i1}^T, b_{i2}^T, \ldots, b_{ir}^T)$$

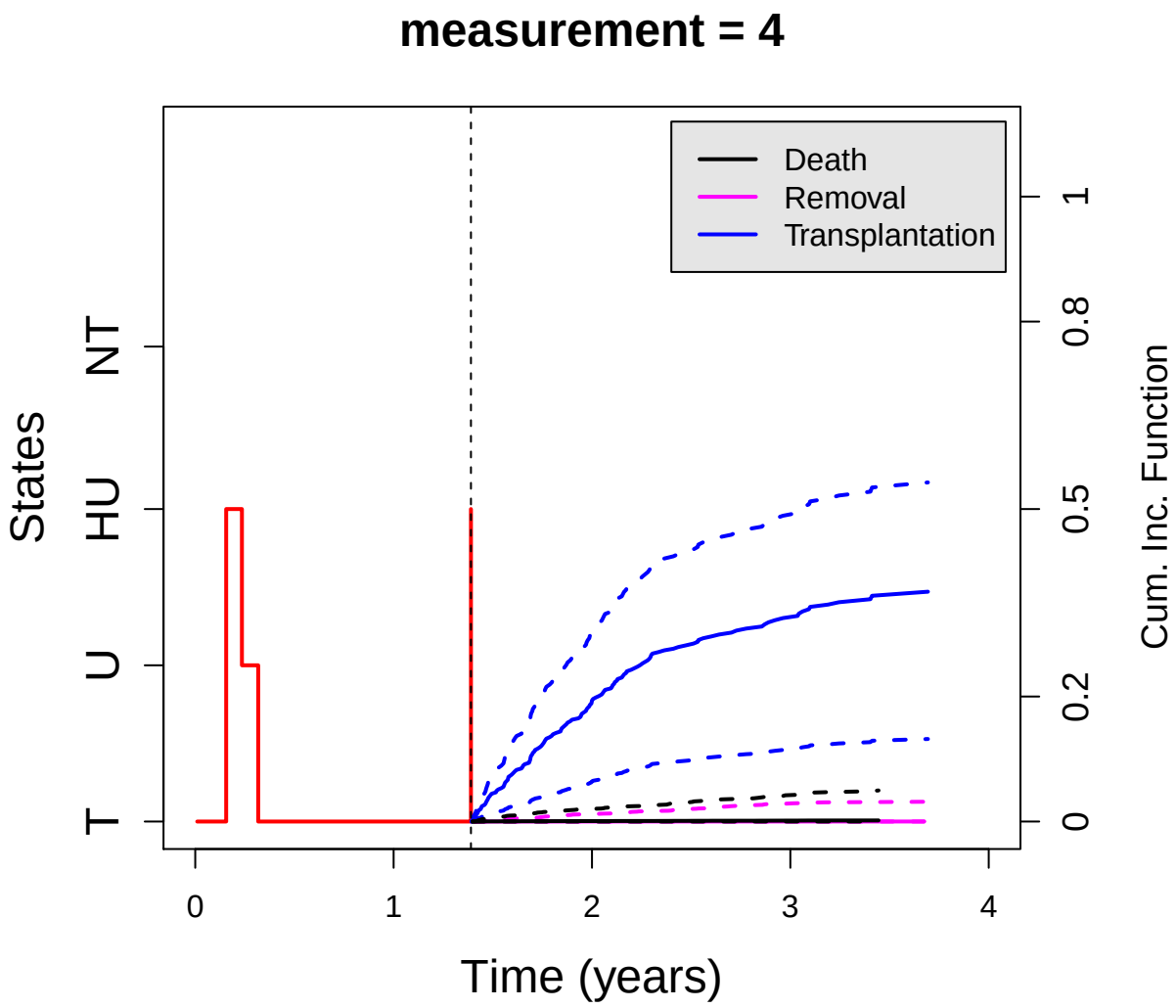
v_i - baseline covariates

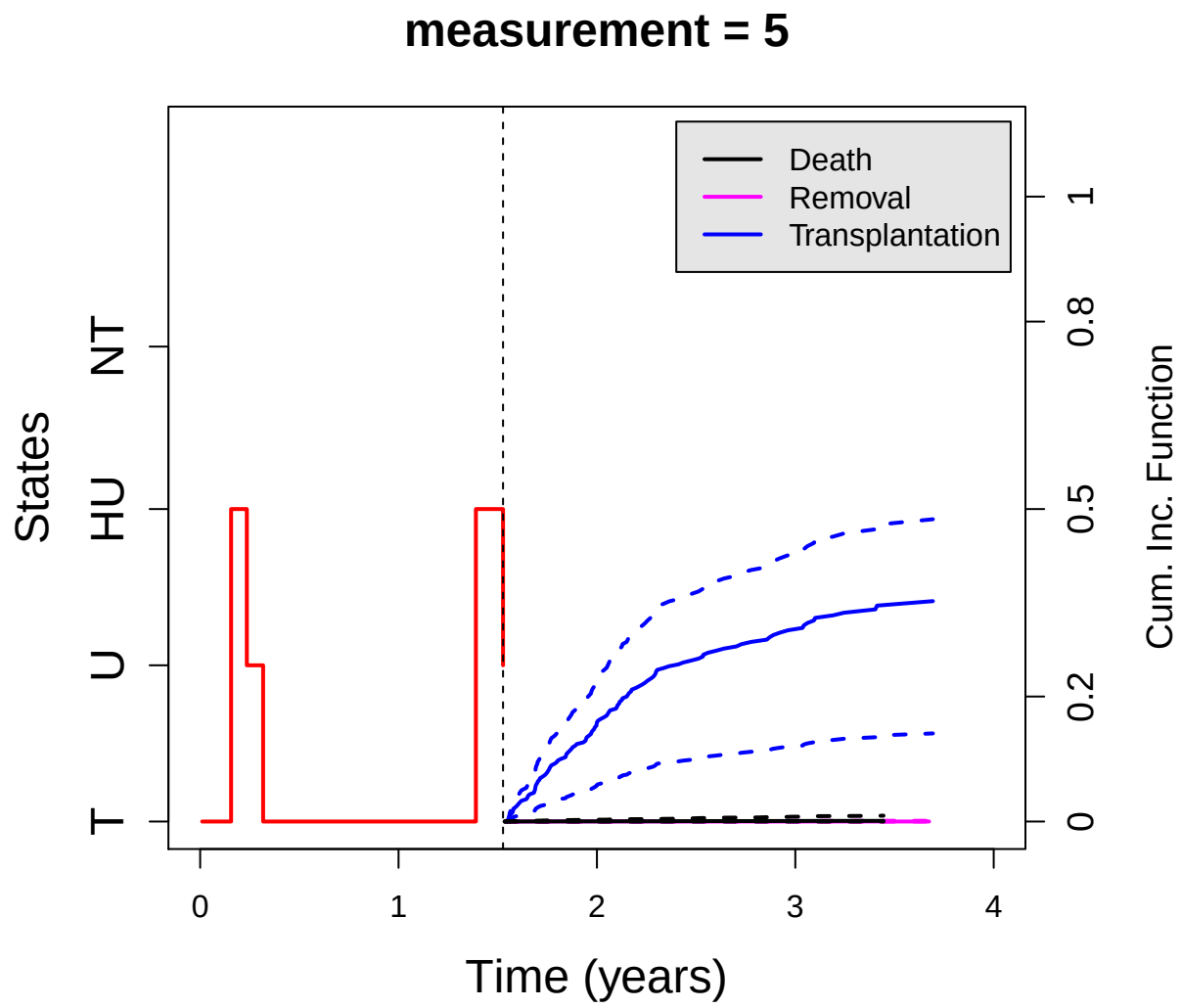
  - ▷ sharing all random effects b_i with multinomial logit model
  - ▷ cause-specific baseline hazards $\lambda_{0k}(t)$ modeled as piecewise constant function
  - ▷ γ - measure of strength of association between longitudinal and survival processes

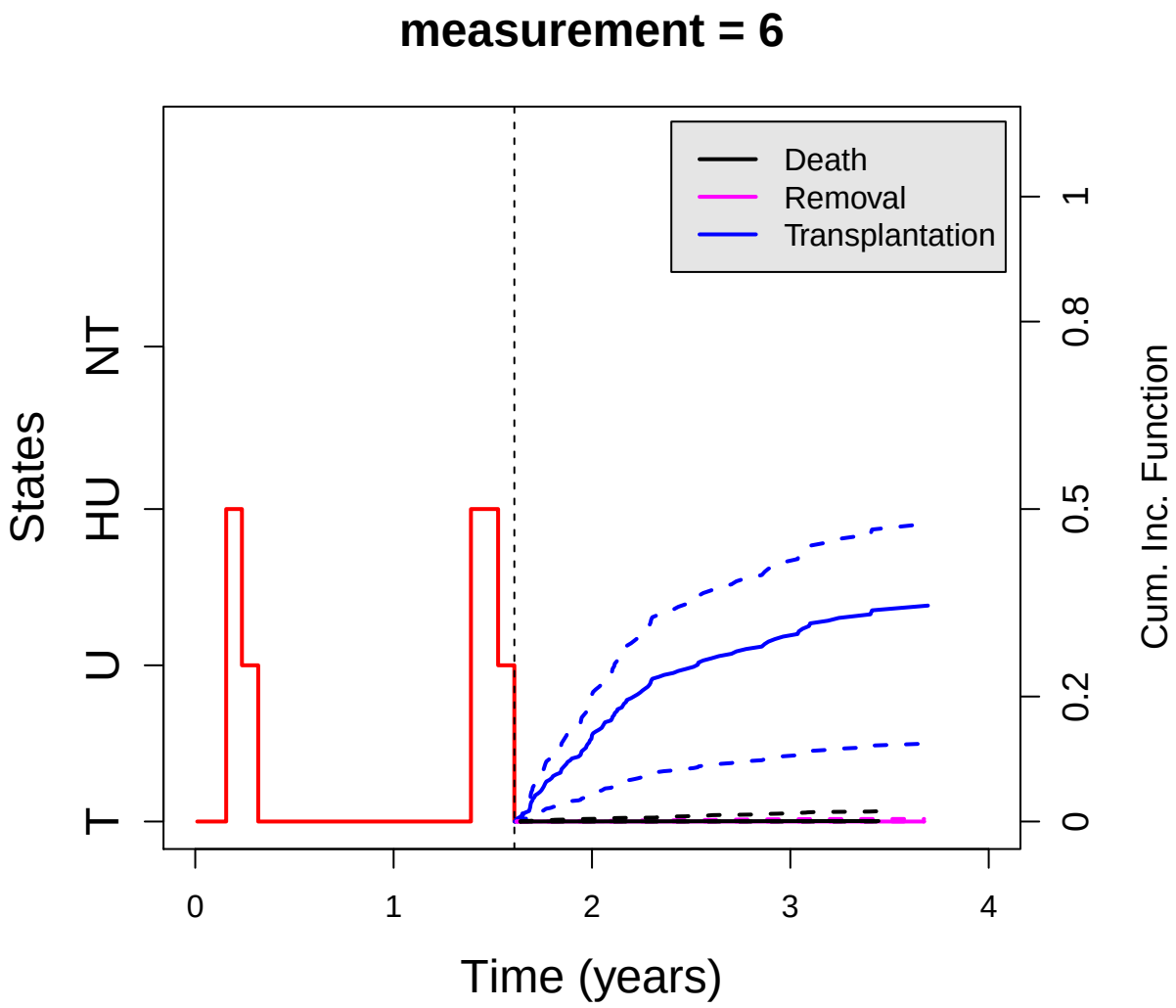

measurement = 1
Death
Removal
Transplantation
States
NT
HU
U
T
0
1
2
3
4
Time (years)
0
0.2
0.5
0.8
1
Cum. Inc. Function


measurement = 2
Death
Removal
Transplantation
States
NT
HU
U
T
0
1
2
3
4
Time (years)
0
0.2
0.5
0.8
1
Cum. Inc. Function


measurement = 3
Death
Removal
Transplantation
States
NT
HU
U
T
0
0.2
0.5
0.8
1
Cum. Inc. Function
0
1
2
3
4
Time (years)


measurement = 4
Death
Removal
Transplantation
States
NT
HU
U
T
0
0.2
0.5
0.8
1
Cum. Inc. Function
0
1
2
3
4
Time (years)


measurement = 5
Death
Removal
Transplantation
States
NT
HU
U
T
0
1
2
3
4
Time (years)
0
0.2
0.5
0.8
1
Cum. Inc. Function


measurement = 6
Death
Removal
Transplantation
States
NT
HU
U
T
Cum. Inc. Function
0
0.2
0.5
0.8
1
0
1
2
3
4
Time (years)

Extensions

- Landmark approach can be also extended using causes-specific hazards
- Fine-Gray type approach combined with landmarking(Cortese and Andersen (2010))
- Pseudo-values approach

Final Comments/Current Work

- In context of time-dependent ROC curves Heagerty et al.(2005) proposed several definitions of cases and controls
- Saha and Heagerty (2010) and Zheng et al. (2012) extended definition for competing risks setting
- Explore different methods of classifying subjects and use similar sampling procedure to estimate ROC in joint modeling framework
- This extension could be applied to fully Bayesian model for competing risks presented above
- Joint models for continuous longitudinal outcome implemented in **JM** and **JMBayes**
- Landmark approach : **dynpred**


Thank you!

References

Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3**, 33-50.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67**, 819-829.

van Houwelingen, H. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34**, 70-85.

Yu, M., Taylor, J., and Sandler, H. (2008). Individualized prediction in prostate cancer studies using a joint longitudinal-survival-cure model. *Journal of the American Statistical Association* **103**, 178-187.

Additional Slides

	$\widehat{\mathsf{PE}}(11\|9)$	$\widehat{\mathsf{IPE}}(11\|9)$	$\widehat{\mathsf{AUC}}(11\|9)$	$\widehat{\mathsf{C}}_{dyn}^{\Delta t=2}$
JM_1	0.05379544	0.1299842	0.5977982	0.6174411
JM_2	0.05287227	0.1276448	0.5966166	0.6243637
JM_3	0.05163110	0.1245160	0.5578209	0.5820453
JM_4	0.07623901	0.1852797	0.5595386	0.5766765
LM_1	0.06042414	0.1225664	0.6204565	0.6408248
LM_2	0.06036801	0.1223102	0.6234404	0.6472022
LM_3	0.06038512	0.1224577	0.6220706	0.6315361

	Scenario			
	I	II	III	IV
γ_0	−6.73	−6.73	−6.73	−6.73
γ_1	0.41	0.41	0.41	0.41
α_1	0.7	0.05	0.08	−0.3
α_2		3.3		−0.8
α_3				0.3
α_4				0.8
σ_t	1.65	1.65	1.65	1.60