Part IV – Extensions: Competing Risks Endpoints and Non-Parametric AUC(t) Estimation

- Patrick J. Heagerty PhD
- Department of Biostatistics
- University of Washington

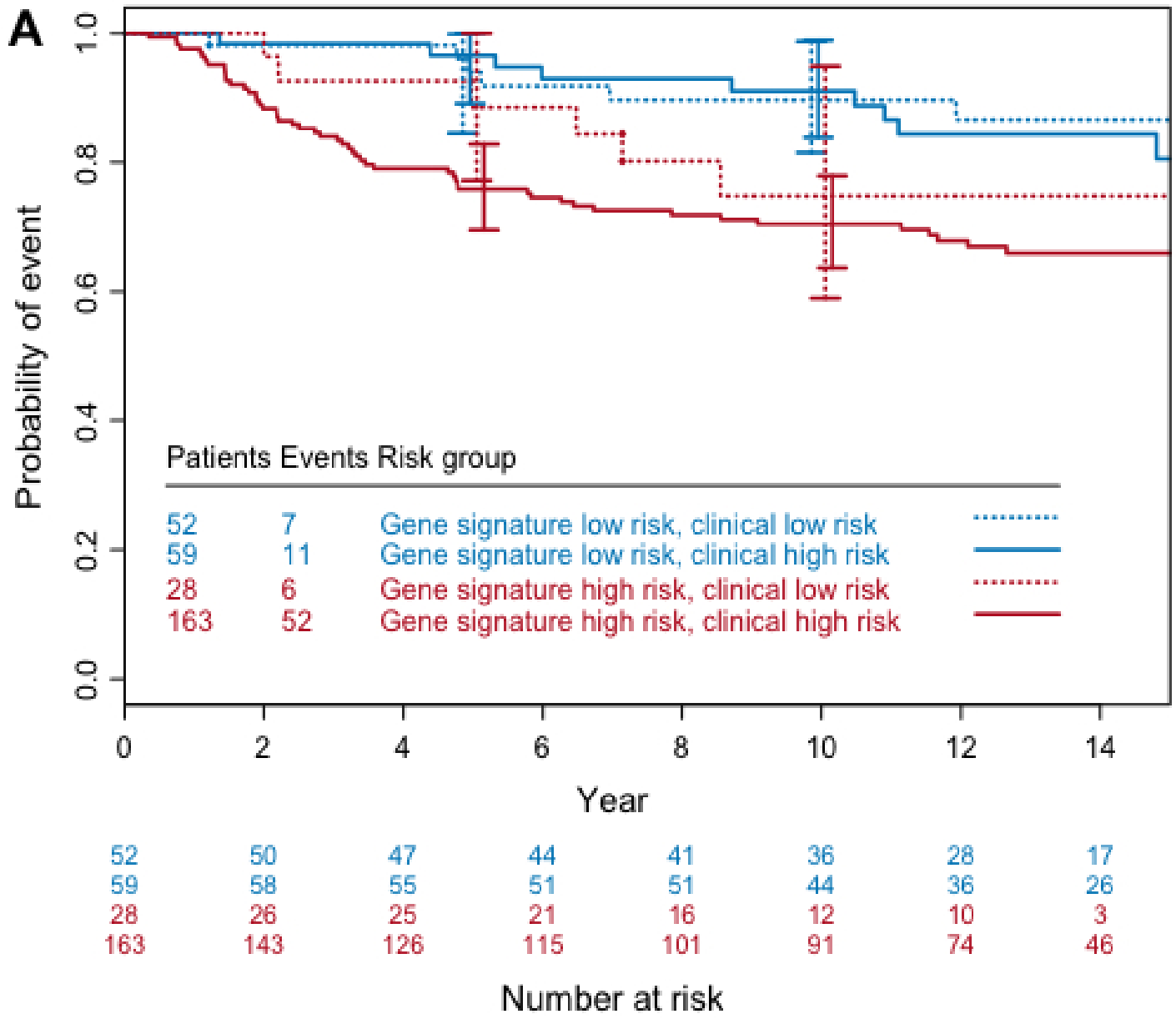
Session Four Outline

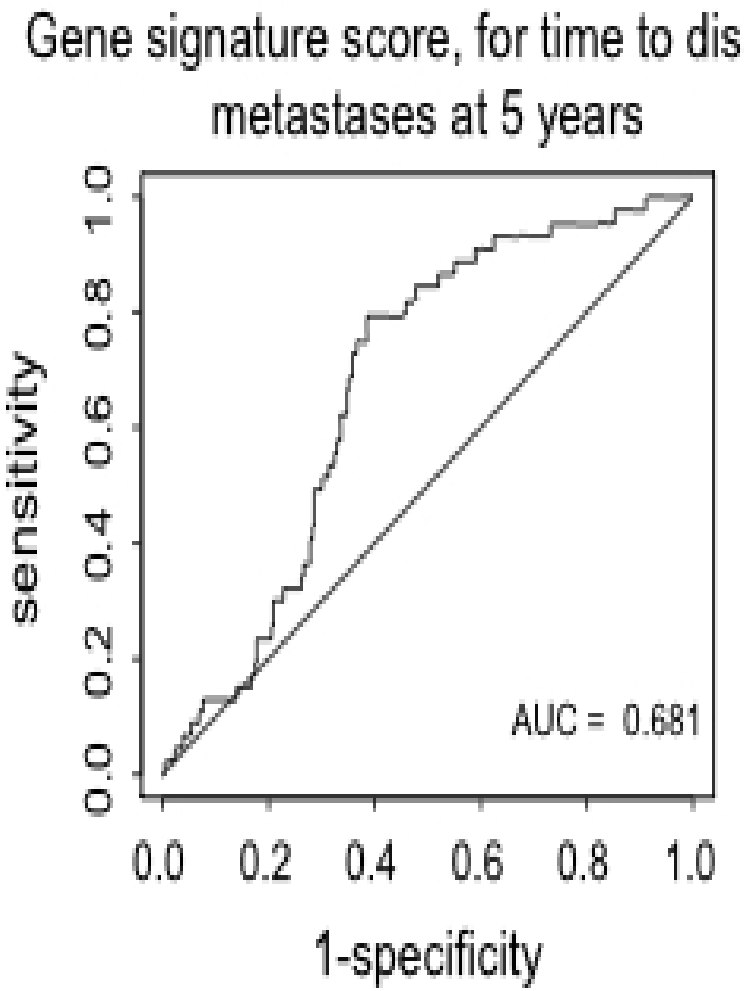
- Examples
  - ▷ Breast Cancer: 70 gene prediction / validation
  - ▷ HIV: markers of disease progression
- Competing Risks Data
- $TP^{\mathbb{C}}$ and cause-specific endpoints / Estimation (non-parametric)
- $TP^{\mathbb{I}}$ and cause-specific endpoints / Estimation (semi-parametric)
- Illustration / Software

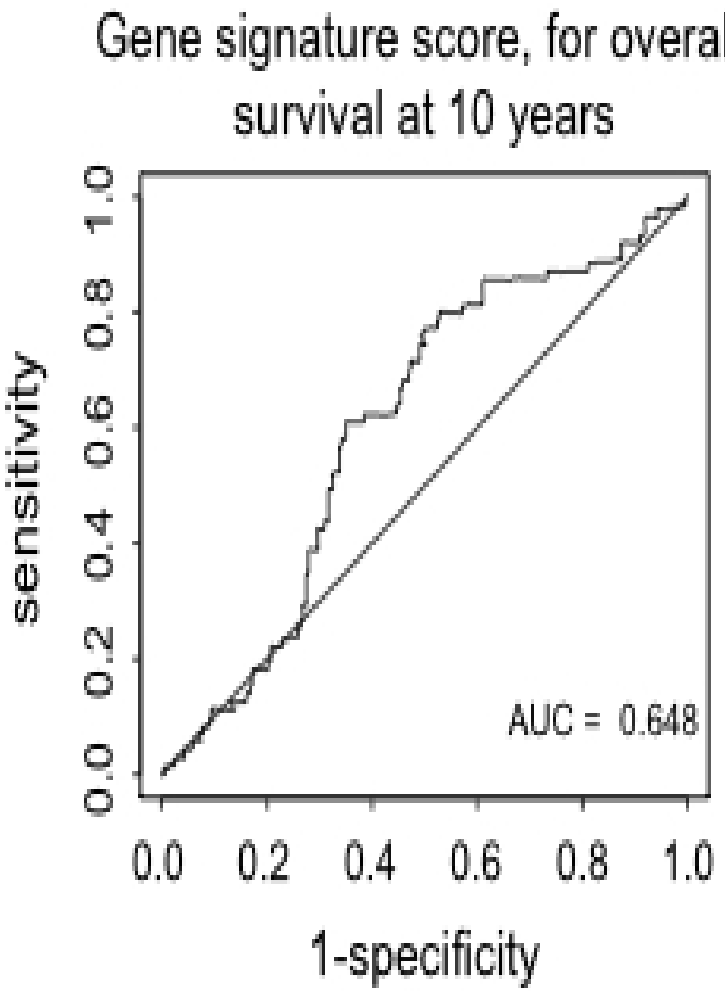
Example: BC and 70-gene Signature among Node-negative

Breast Cancer Prediction

- $N = 307$ women from (5) Euro Centers
- Endpoint(s):
  - time-until-distant-metastases (next slide)
  - disease-free-survival
- Predictive measurements:
  - Clinicopathologic risk assessment
  - 70-gene Signature
- Goal: validate (added) utility of "signature"
- Buyse et al. (2006) $JNCI$


A
Probability of event
1.0
0.8
0.6
0.4
0.2
0.0
Patients Events Risk group
52 7 Gene signature low risk, clinical low risk
59 11 Gene signature low risk, clinical high risk
28 6 Gene signature high risk, clinical low risk
163 52 Gene signature high risk, clinical high risk
0 2 4 6 8 10 12 14
Year
52 50 47 44 41 36 28 17
59 58 55 51 51 44 36 26
28 26 25 21 16 12 10 3
163 143 126 115 101 91 74 46
Number at risk

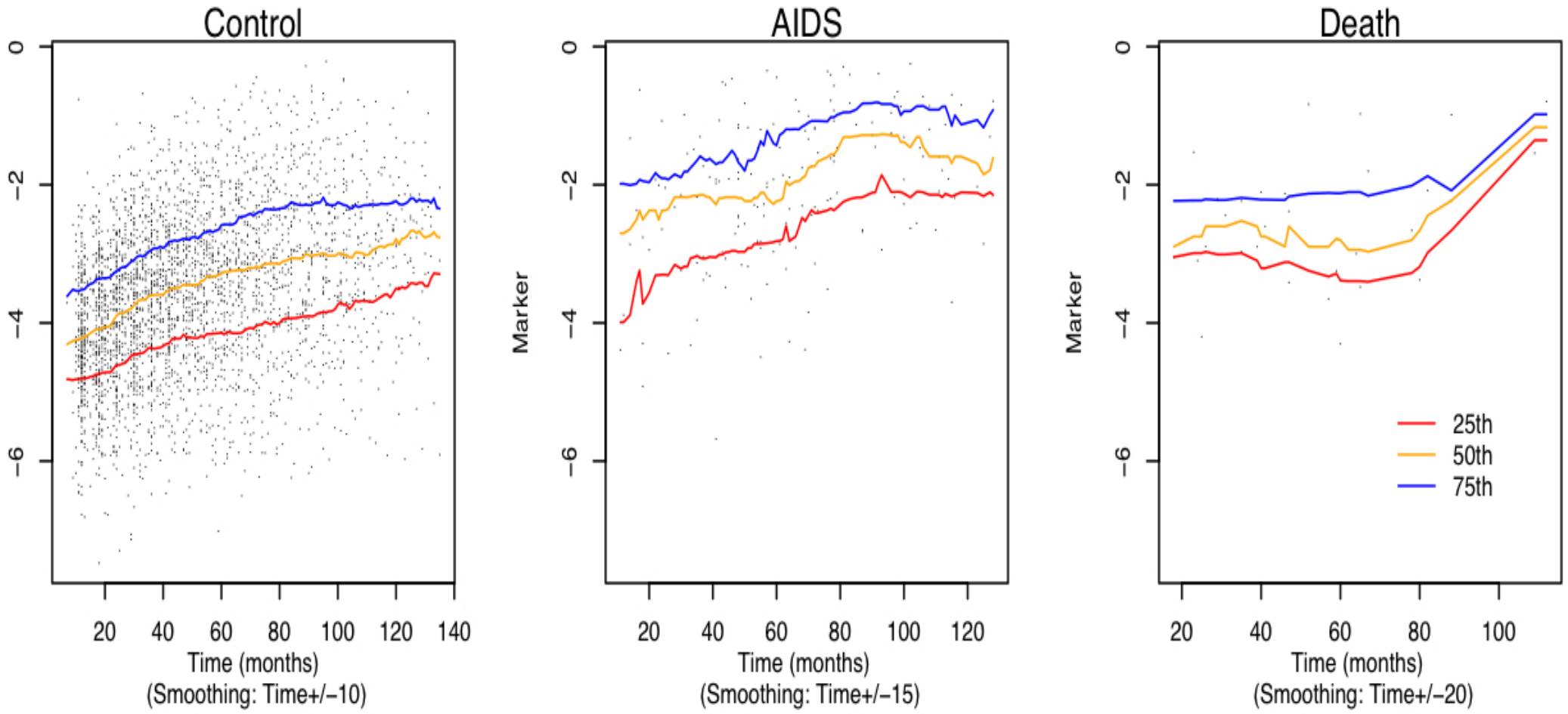

Gene signature score, for time to distant metastases at 5 years
sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.681


Gene signature score, for overall survival at 10 years
sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.648

Example: Immune markers and disease progression

Multicenter AIDS Cohort Study

- $N = 447$ men observed to seroconvert
- Endpoint(s):
  - time-until-AIDS
  - time-until-death
- Predictive measurements:
  - CD4, CD8 at "baseline"
  - CD4, CD8 measured every 6 months
- Goal: evaluate markers as predictors of disease-progression
- Saha and Heagerty (2010)


Control
AIDS
Death
Marker
0
−2
−4
−6
20
40
60
80
100
120
140
Time (months)
(Smoothing: Time+/−10)
(Smoothing: Time+/−15)
(Smoothing: Time+/−20)
25th
50th
75th

Competing Risks Endpoints

- Observed time-until-event, and **type** of event.
  - Death, cause = (BC, other)
- "Derived" time-until-first-event, and type of event
  - Time until progression or Death (first event, type)
  - e.g. metastases, death (without metastases first)
  - e.g. AIDS, death (without AIDS first)
- **Representation**
  - (T_i^*, δ_i) where $\delta_i = 0, 1, 2, \ldots, C$
  - δ_i: censored = 0; types = 1, 2, ... C

Sensitivity and Specificity for Survival (again!)

Let T denote the survival time, and let $N(t)$ denote the counting process for the uncensored outcome:

$$N(t) = 1(T \leq t)$$

Possible definitions:

$$\text{CASE}(t) \quad : \quad \begin{cases} \textbf{Cumulative} & N(t) = 1 \\ \textbf{Incident} & dN(t) = 1 \end{cases}$$

$$\text{CONTROL}(t) \quad : \quad \begin{cases} \text{Static} & N(t^\star) = 0 \\ \textbf{Dynamic} & N(t) = 0 \end{cases}$$

- Where $t^\star$ is a fixed "large" time, $t^\star >> t$.

Sensitivity and Specificity for Cause-specific Survival

Define:

$$\text{sensitivity}^{\mathbb{C}}(c, t; \mathbf{d}) \quad : \quad P(M > c \mid T \leq t; \delta = \mathbf{d})$$

$$\text{specificity}^{\mathbb{D}}(c, t) \quad : \quad P(M \leq c \mid T > t)$$

- "Cases" are broken into finer groups based on the **type** of case.
- e.g. high marker given metastases by time t (d=1)
- e.g. high marker given death w/o metastases by time t (d=2)

Sensitivity and Specificity for Cause-specific Survival

Example: d=1, 2

$$\begin{aligned}
\text{Case } \mathbf{1} &: \; T_i \le t,\ \delta = \mathbf{1} \\
\text{Case } \mathbf{2} &: \; T_i \le t,\ \delta = \mathbf{2} \\
\text{Control} &: \; T_i > t,\ \delta = [1, 2]
\end{aligned}$$

$$\begin{aligned}
TP_t^{\mathbb{C}}(c, \mathbf{1}) &= P(M > c \mid T_i \le t,\ \delta = \mathbf{1}) \\
TP_t^{\mathbb{C}}(c, \mathbf{2}) &= P(M > c \mid T_i \le t,\ \delta = \mathbf{2}) \\
FP_t^{\mathbb{D}}(c) &= P(M > c \mid T_i > t,\ \delta = [1, 2])
\end{aligned}$$

Estimation: Using "local" Cumulative Incidence

- **Cause-specific Cumulative Incidence**
  - $C_d(t) = P(T \leq t; \delta = d)$
  - Percent of population with event of type **d** by time t.

- Non-parametric estimation (K&P 1980, p. 168)

$$\widehat{C}_d(t) = \sum_{s \leq t} \widehat{S}(s-) \cdot \widehat{\lambda}_d(s)$$

Estimation: Using "local" Cumulative Incidence

- Cumulative incidence estimator can handle censoring.
- Parallel the estimation of HLP(2000) using:

$$P(M > c \mid T \leq t, \delta = d) \quad = \quad \frac{P(M > c, T \leq t, \delta = d)}{C_d(t)}$$

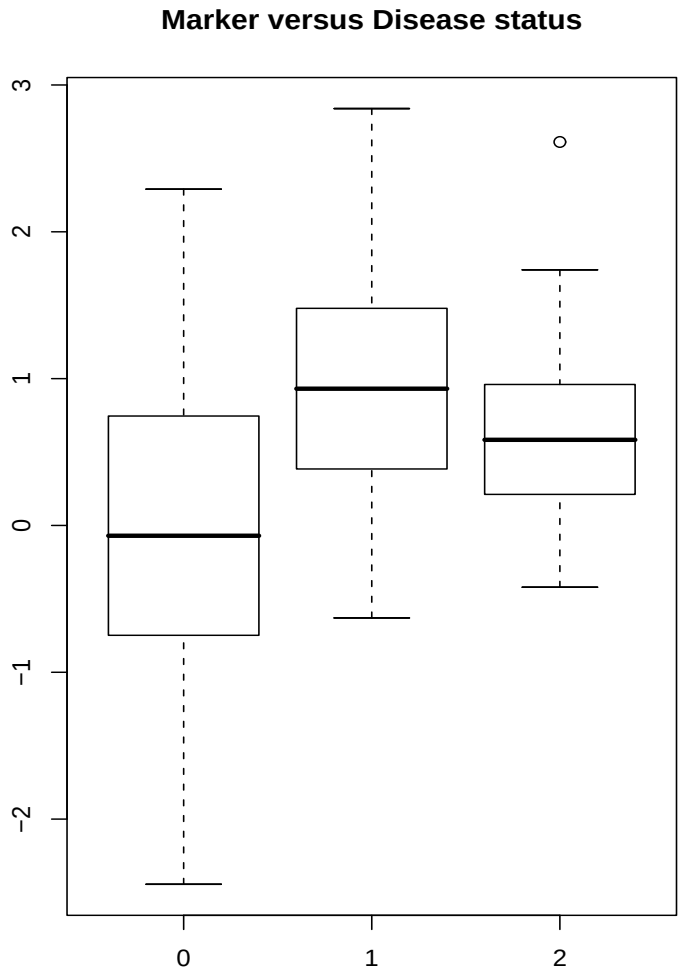
$$\begin{aligned} \text{numerator} \quad &= \quad \int_c^\infty P(T \leq t, \delta = d \mid M = m) \cdot P(M = m)\, dm \\ &= \quad \int_c^\infty C_d(t \mid M = m) \cdot P(M = m)\, dm \end{aligned}$$

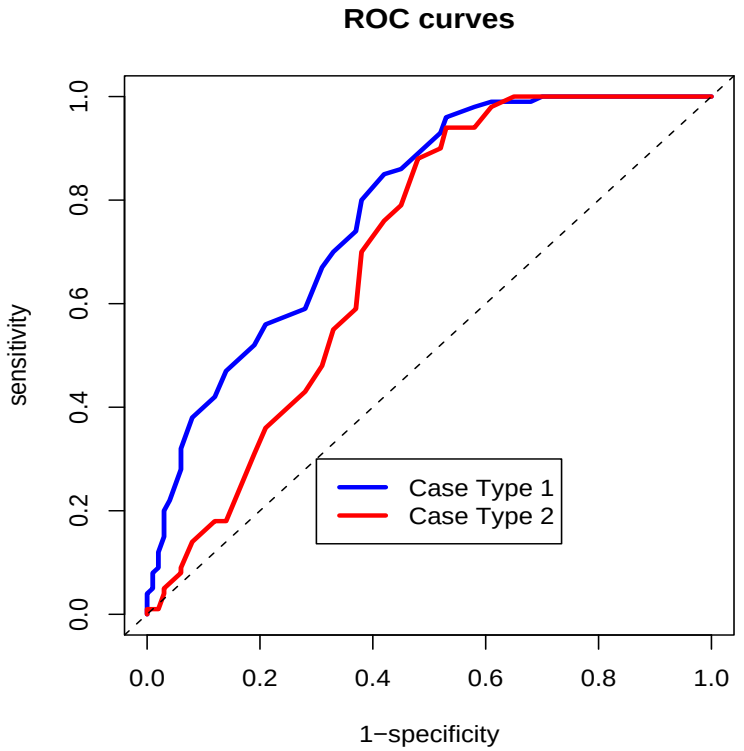
Estimation: Using "local" Cumulative Incidence

- Use local cause-specific cumulative incidence to estimate $C_d(t \mid M = m)$ and use empirical for $P(M = m)$.
- **Note**:

$$P(T > t \mid M = m) = 1 - \sum_d P(T \leq t, \delta = d \mid M = m)$$

- Use above to estimate $FP_t^{\mathbb{D}}(c)$ such that joint distribution is proper.

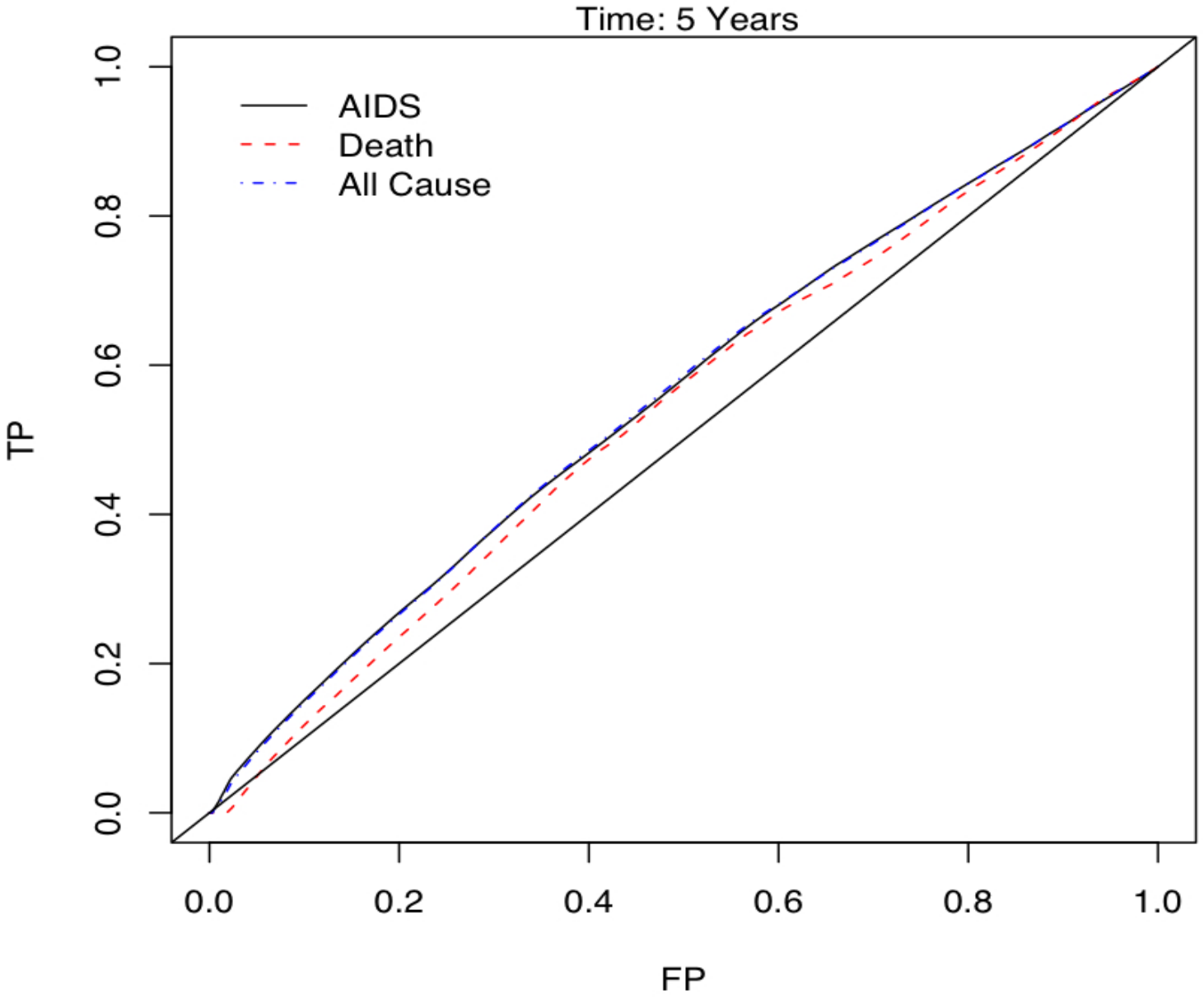

Marker versus Disease status
3
2
1
0
−1
−2
0
1
2


ROC curves
sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1−specificity
Case Type 1
Case Type 2

Software / Illustration

- **Software**: CRAN package for R called `survivalROC` – we have extended this to implement the competing risks calculations. (P. Saha)

- **MACS Data**
  - Baseline (e.g. seroconversion time) values of CD4 and CD8
  - Linear combination based on Cox regression
  - Case Type 1 = AIDS
  - Case Type 2 = death before AIDS
  - Time for cumulative case status = 5 years


Time: 5 Years
AIDS
Death
All Cause
TP
FP
0.0
0.2
0.4
0.6
0.8
1.0

Review: Sensitivity and Specificity for Survival

Define: Heagerty and Zheng (2005) / Saha and Heagerty (2010)

$$\text{sensitivity}^{\mathbb{I}}(c,t) \quad : \quad P[\mathbf{M(t)} > c \mid T{=}t]$$
$$P[\mathbf{M(t)} > c \mid dN(t) = 1]$$

$$\text{specificity}^{\mathbb{D}}(c,t) \quad : \quad P[\mathbf{M(t)} \leq c \mid T > t]$$
$$P[\mathbf{M(t)} \leq c \mid N(t) = 0]$$

$$TP_t^{\mathbb{I}}(c) \quad = \quad P[\mathbf{M(t)} > c \mid \mathbf{dN(t){=}1}]$$

$$FP_t^{\mathbb{D}}(c) \quad = \quad P[\mathbf{M(t)} > c \mid N(t) = 0]$$

Sensitivity and Specificity for Cause-specific Survival

Define:

$$\text{sensitivity}^{\mathbb{I}}(c, t; \mathbf{d}) \quad : \quad P(M > c \mid T{=}t; \delta = \mathbf{d})$$

$$\text{specificity}^{\mathbb{D}}(c, t) \quad : \quad P(M \leq c \mid T > t)$$

- "Cases" are broken into finer groups based on the **type** of case.
- e.g. high marker given metastases **at** time t (d=1)
- e.g. high marker given death w/o metastases **at** time t (d=2)

Sensitivity and Specificity for Cause-specific Survival

Example: d=1, 2

$$\begin{aligned} \text{Case } \mathbf{1} &: T_i{=}t,\ \delta = \mathbf{1} \\ \text{Case } \mathbf{2} &: T_i{=}t,\ \delta = \mathbf{2} \\ \text{Control} &: T_i > t,\ \delta = [1,2] \end{aligned}$$

$$\begin{aligned} TP_t^{\mathbb{I}}(c,\mathbf{1}) &= P(M > c \mid T_i{=}t,\ \delta = \mathbf{1}) \\ TP_t^{\mathbb{I}}(c,\mathbf{2}) &= P(M > c \mid T_i{=}t,\ \delta = \mathbf{2}) \\ FP_t^{\mathbb{D}}(c) &= P(M > c \mid T_i > t,\ \delta = [1,2]) \end{aligned}$$

Estimation: Hazard as Bridge

A general definition for the cause-specific hazard is

$$\lambda^{(d)}(t \mid M_i) = \frac{P(T_i = t, \delta_i = d \mid M_i)}{P(T_i \geq t \mid M_i)}$$

Then using a little algebra yields

$$P(M_i = m \mid T_i = t, \delta_i = d) \quad \propto \quad \underbrace{\lambda^{(d)}(t \mid M_i = m)} \cdot \underbrace{P(M_i = m \mid T_i \geq t)}$$

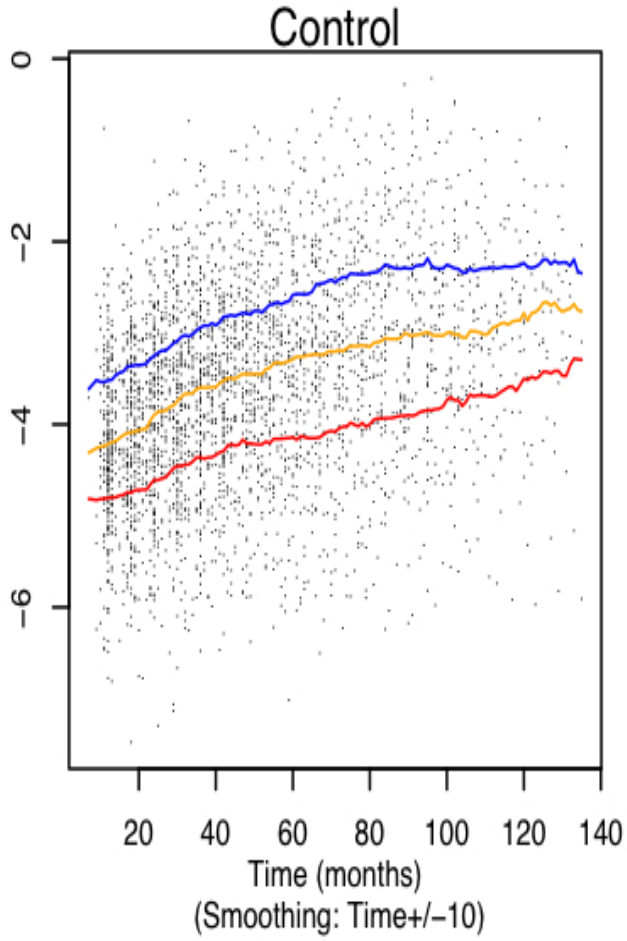
$$\text{Estimate} \quad \Longleftarrow \quad \text{Smooth model} \quad + \quad \text{Empirical}$$

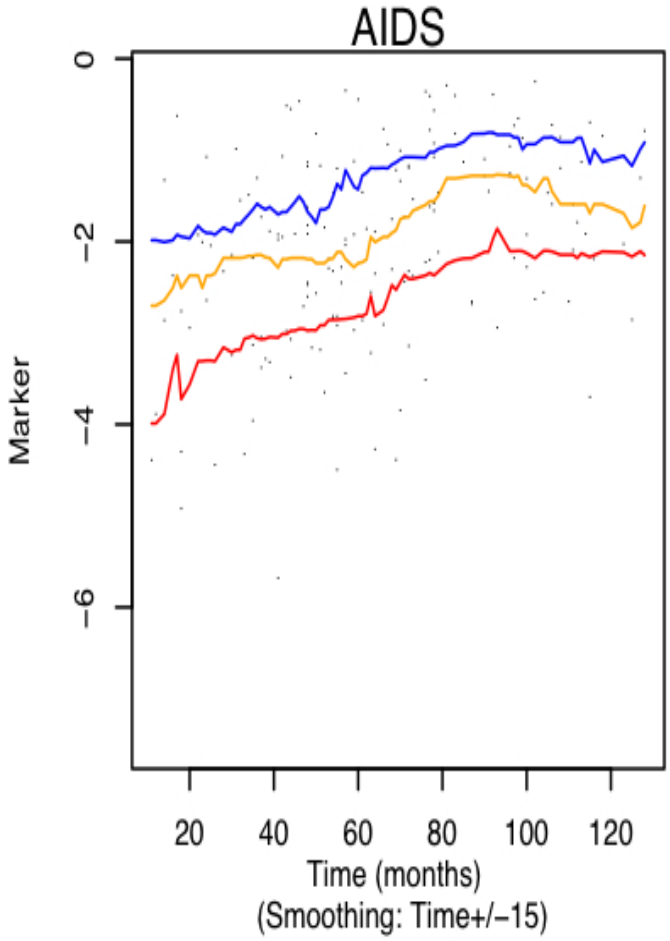
Note: direct (easy) generalization of the HZ(2005) methods.

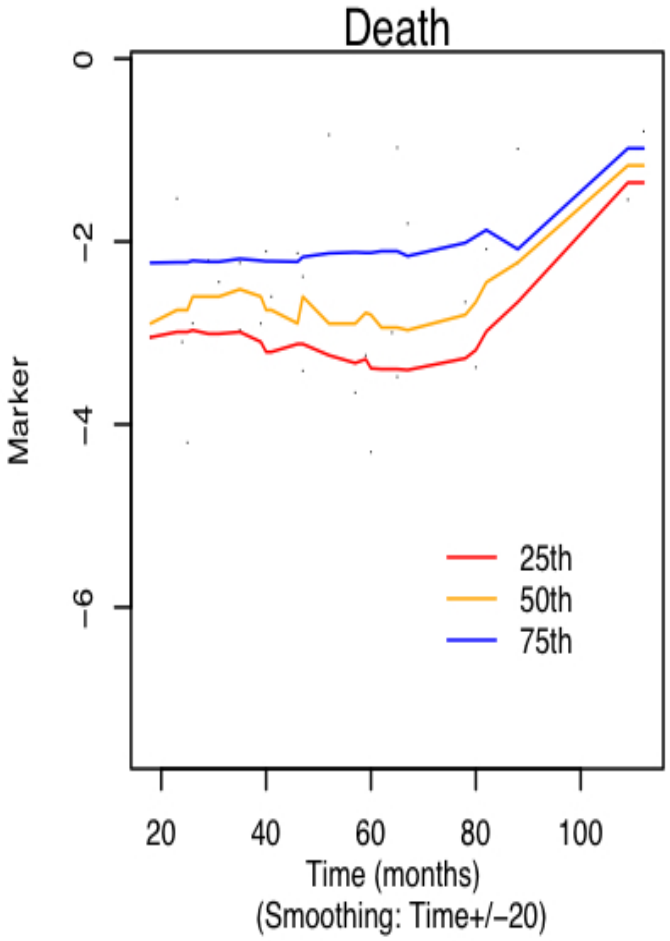
Software / Illustration

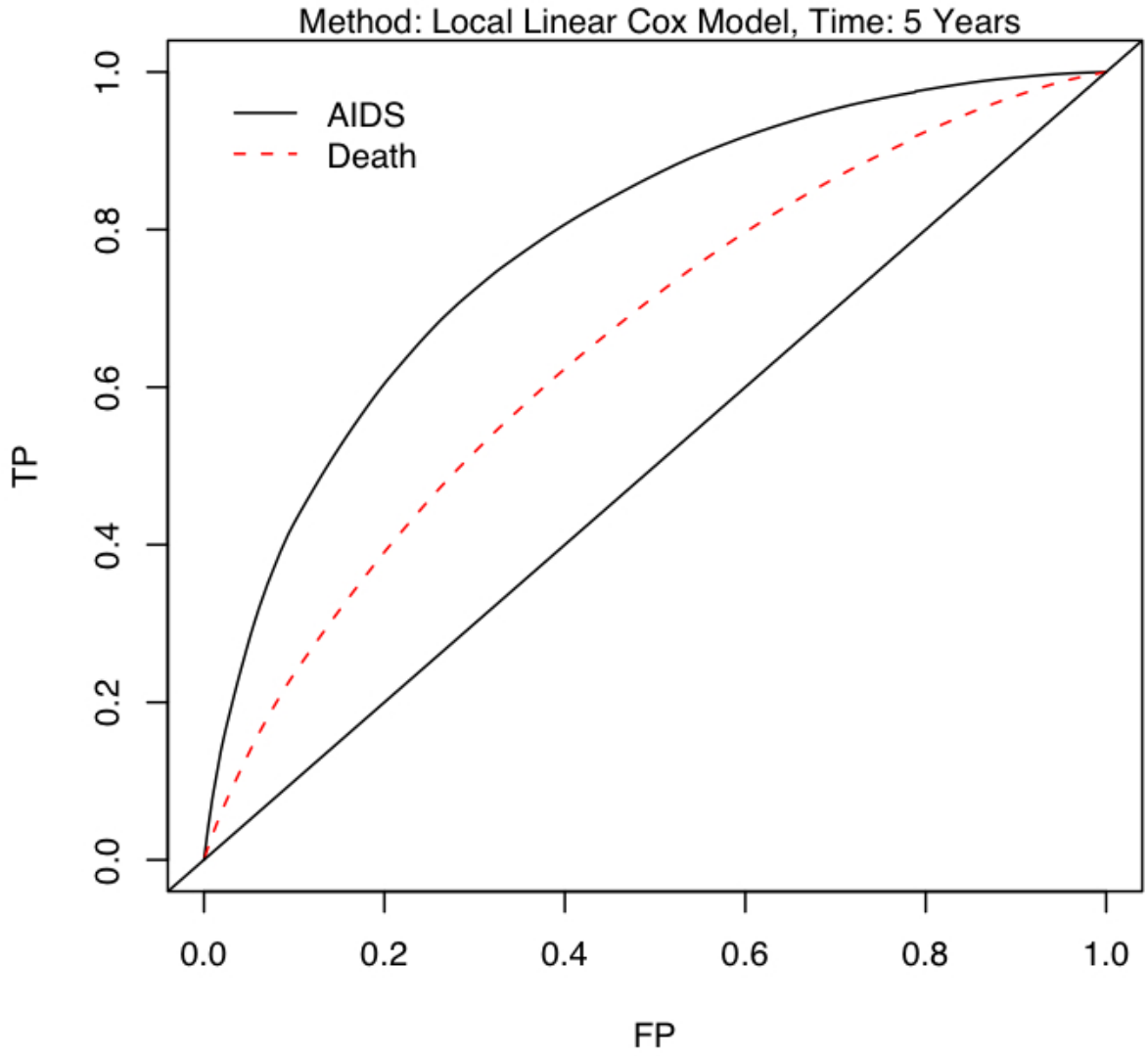
- **Software**: CRAN package for R called `risksetROC` – we have extended this to implement the competing risks calculations, and to handle time-dependent covariates. (P. Saha)

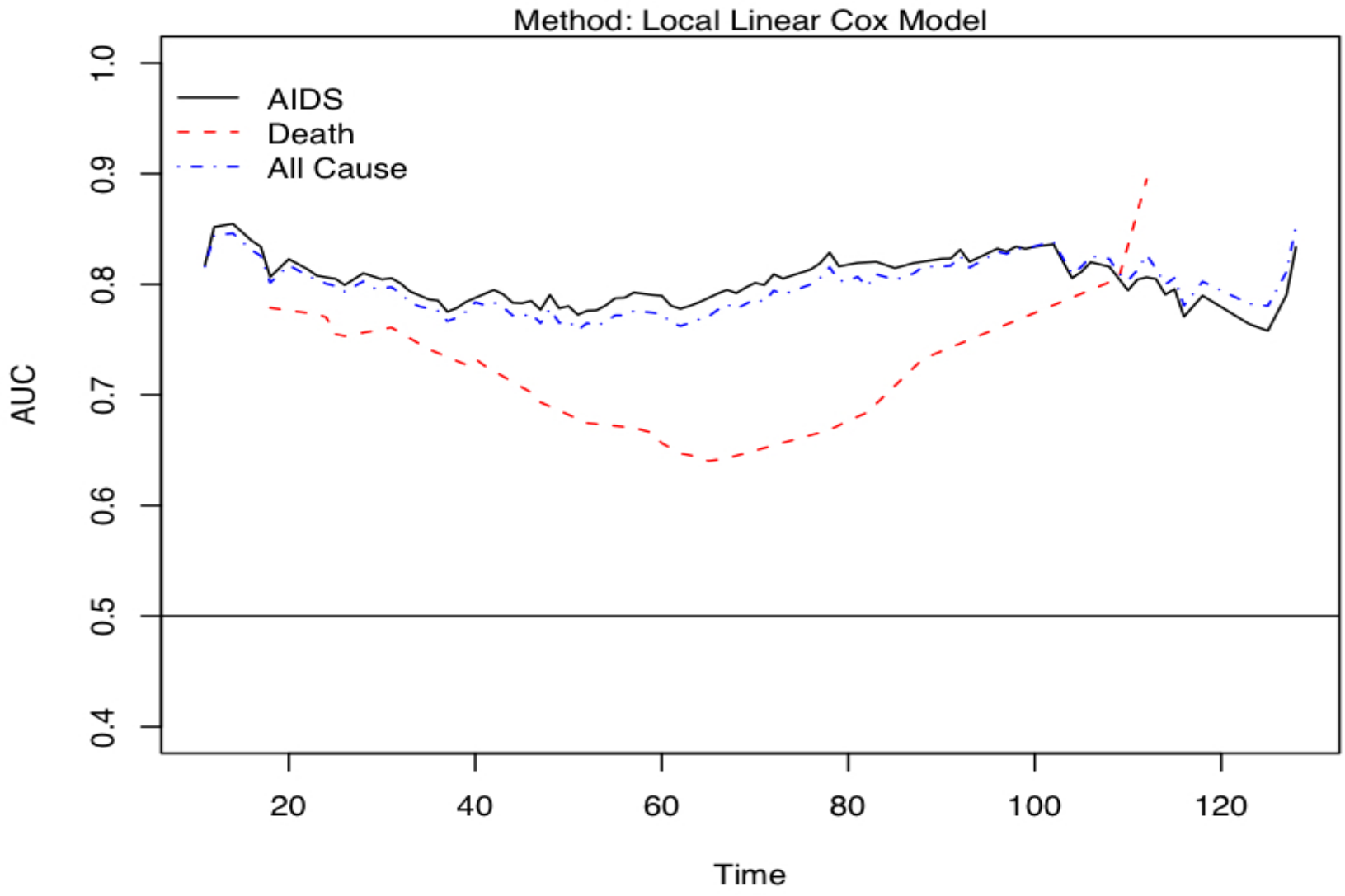
- **MACS Data**
  - Longitudinal values of CD4 and CD8
  - Linear combination based on Cox regression
  - Case Type 1 = AIDS (n=176)
  - Case Type 2 = death before AIDS (n=34)
  - ROC curve, and AUC versus time


Control
0
−2
−4
−6
20
40
60
80
100
120
140
Time (months)
(Smoothing: Time+/−10)


AIDS
Marker
0
−2
−4
−6
20
40
60
80
100
120
Time (months)
(Smoothing: Time+/−15)


Death
Marker
0
−2
−4
−6
20
40
60
80
100
Time (months)
(Smoothing: Time+/−20)
25th
50th
75th


Method: Local Linear Cox Model, Time: 5 Years
AIDS
Death
TP
FP
0.0
0.2
0.4
0.6
0.8
1.0


Method: Local Linear Cox Model
AIDS
Death
All Cause
AUC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
20
40
60
80
100
120
Time

Summary

- Extension of time-dependent ROC methods to competing risks data.
- Cumulative Cases – uses non-parametric methods based on local cumulative incidence calculations.
- Incident Cases – uses semi-parametric methods that parallel those outlined in Heagerty and Zheng (2005).
- Time-dependent markers.

Thanks!

Nonparametric Estimation of AUC(t)

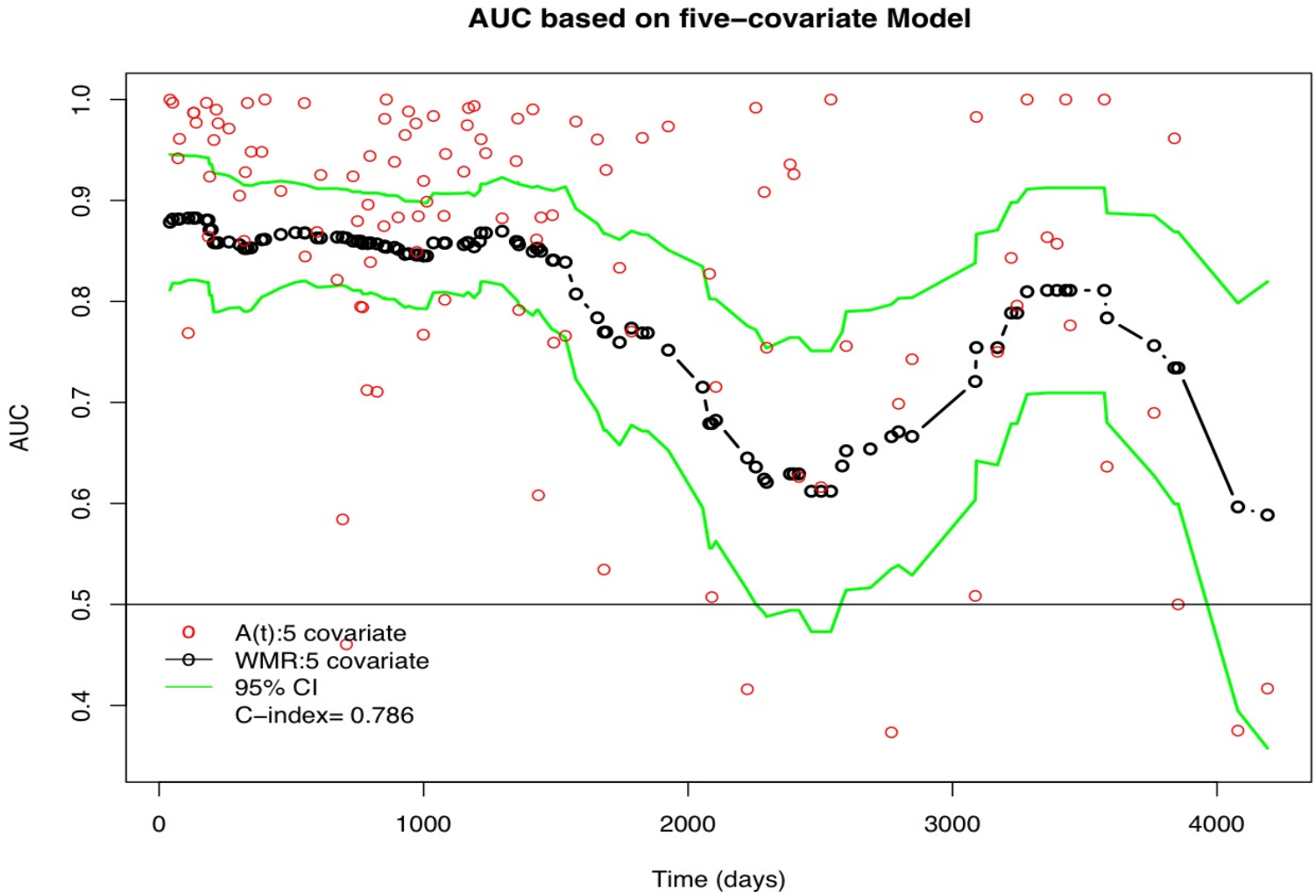
Risk set ranks can be used to nonparametrically estimate $AUC(t)$.

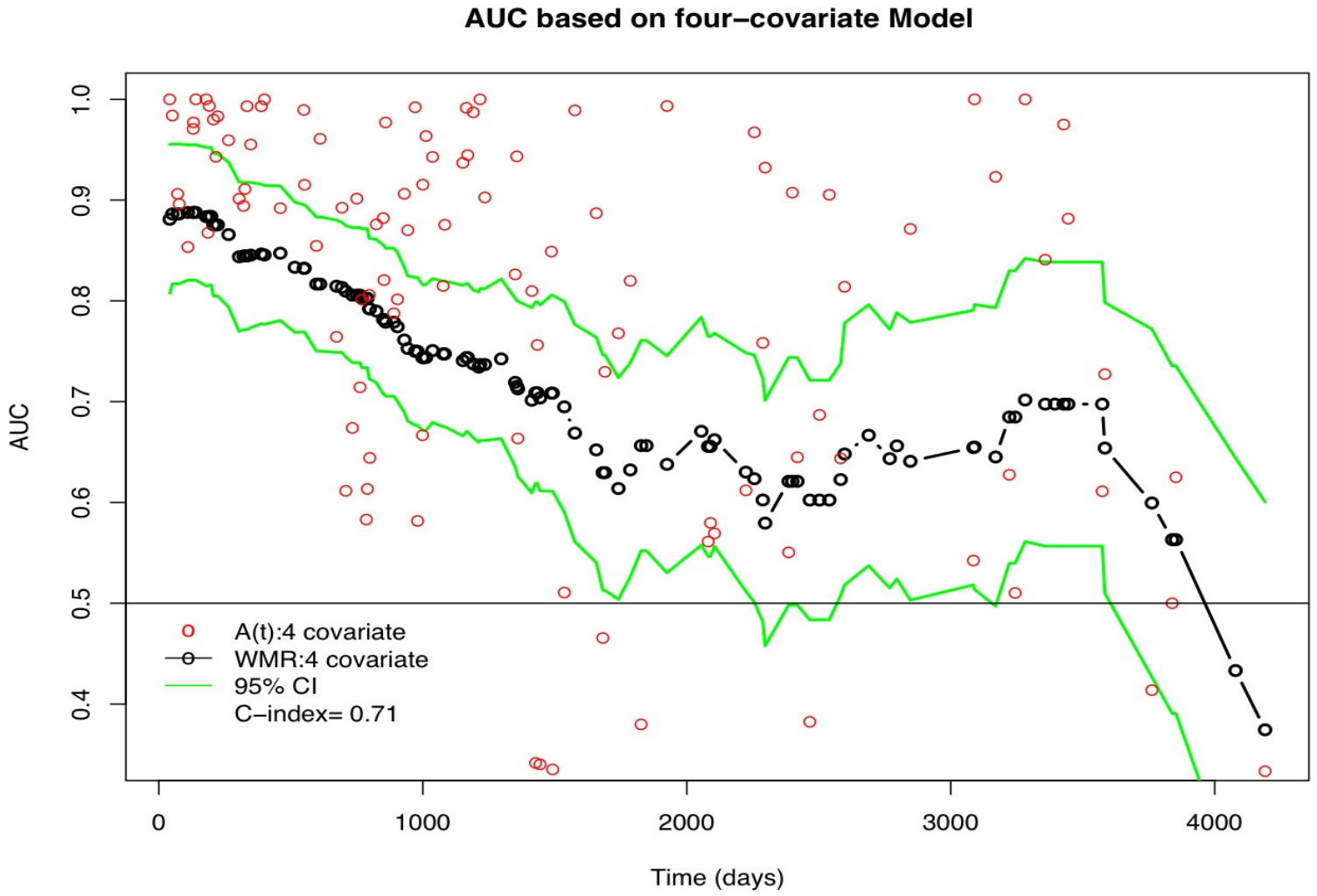
$$AUC(t) = P[M_j > M_k \mid dN_j(t) = 1, N_k(t) = 0]$$

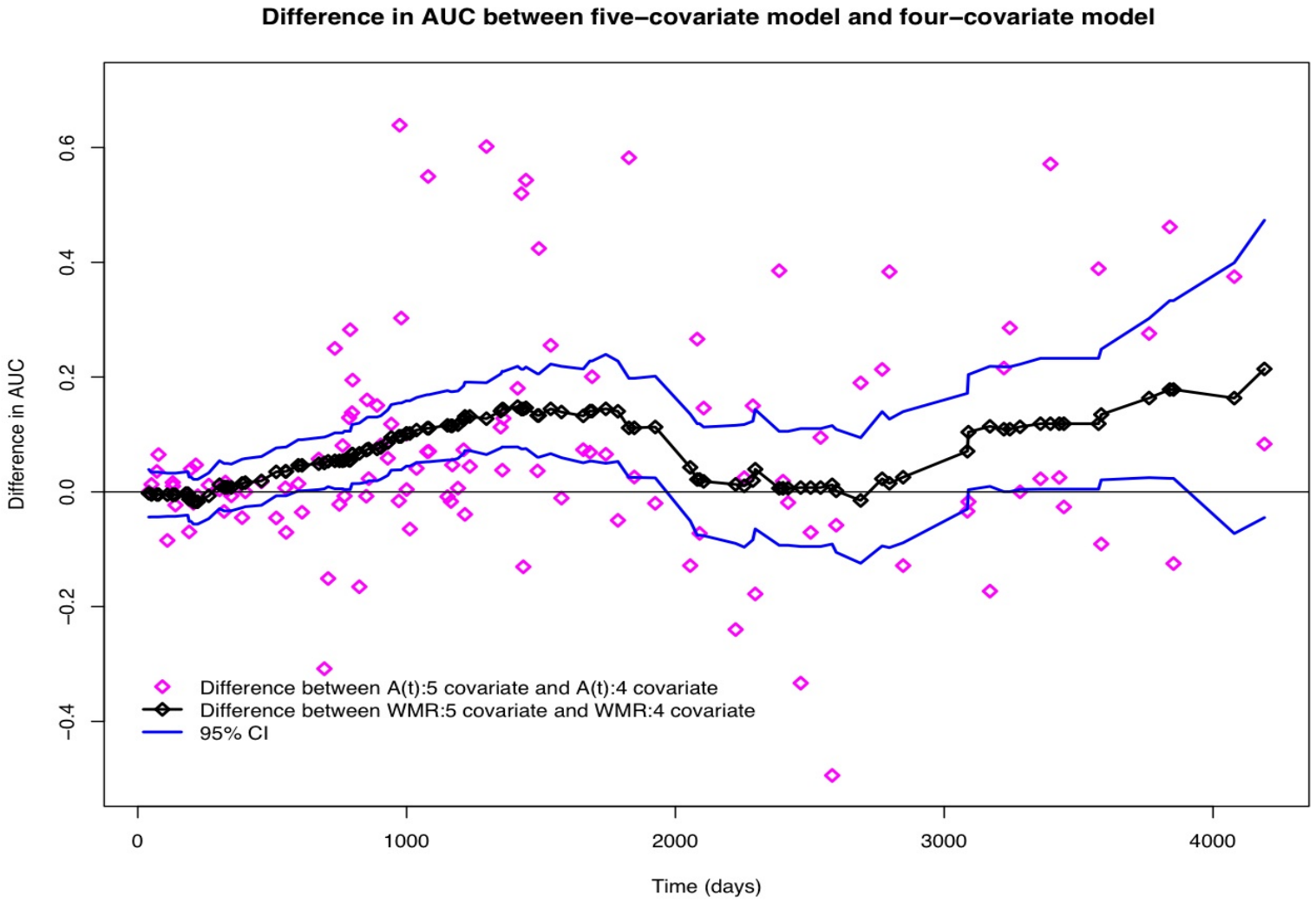
$M^*(t) = M_j$ for $dN_j(t) = 1$ $\qquad$ $\mathcal{C}(t) = \{k \ : \ N_k(t) = 0 \}$

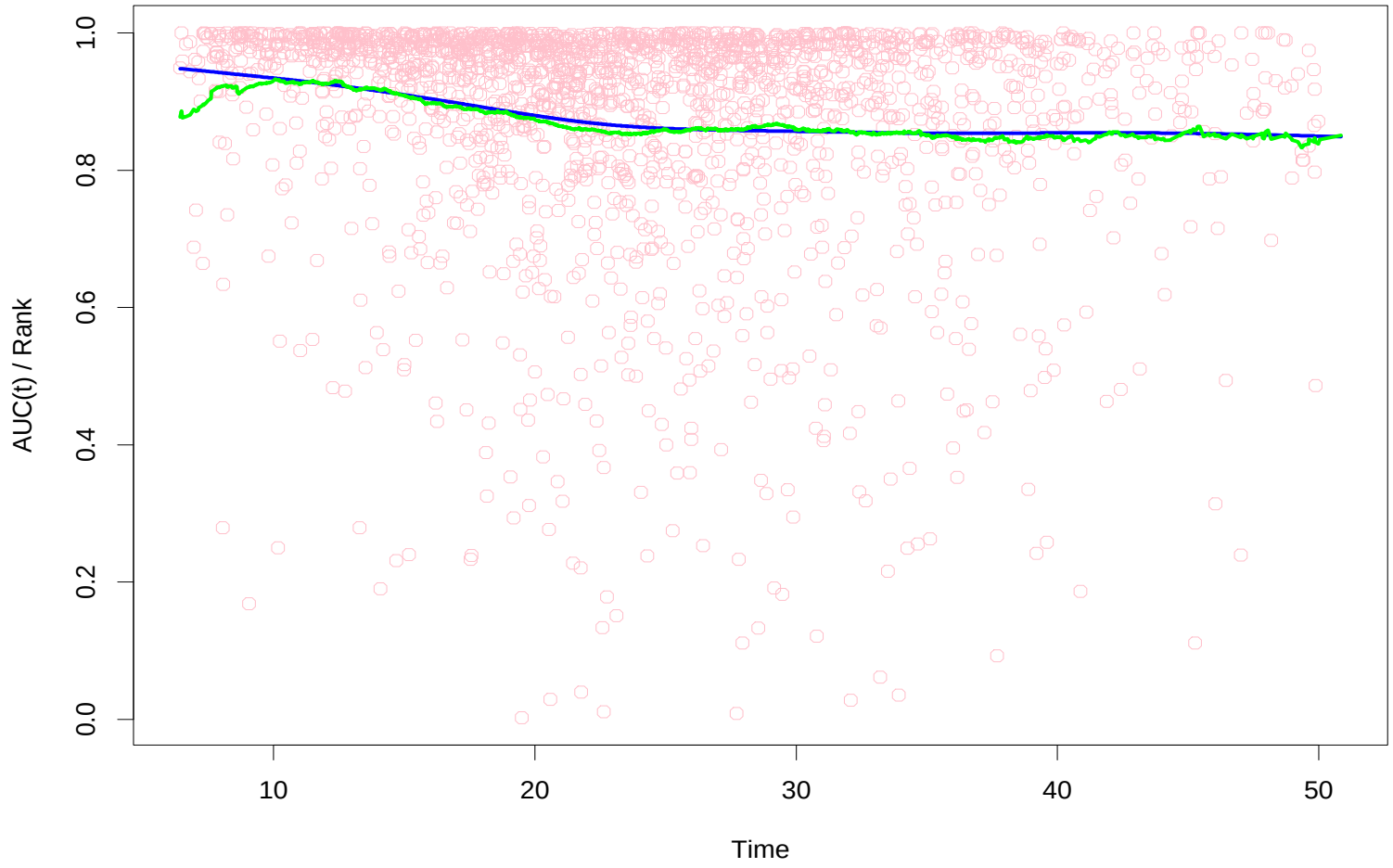
$$\begin{aligned} AUC(t) &= E\{ \ 1[M^*(t) > M_k \mid k \in \mathcal{C}(t)] \ \} \\ \widehat{AUC}(t) &= \frac{1}{n_t} \sum_{k \in \mathcal{C}(t)} 1[M^*(t) > M_k] \\ \widehat{AUC}(t) &= \frac{[\text{risk set rank of } M^*(t) \ ] - 1}{n_t} \end{aligned}$$

Saha & Heagerty (submitted)


AUC based on five−covariate Model
AUC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
1000
2000
3000
4000
Time (days)
A(t):5 covariate
WMR:5 covariate
95% CI
C−index= 0.786


AUC based on four−covariate Model
AUC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
1000
2000
3000
4000
Time (days)
A(t):4 covariate
WMR:4 covariate
95% CI
C−index= 0.71


Difference in AUC between five−covariate model and four−covariate model
Difference in AUC
0.6
0.4
0.2
0.0
−0.2
−0.4
0
1000
2000
3000
4000
Time (days)
Difference between A(t):5 covariate and A(t):4 covariate
Difference between WMR:5 covariate and WMR:4 covariate
95% CI


AUC Based on Risk Set Rank
AUC(t) / Rank
0.0
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
Time