Part [1.2] – Extensions: Competing Risks Endpoints and Non-Parametric AUC(t) Estimation

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

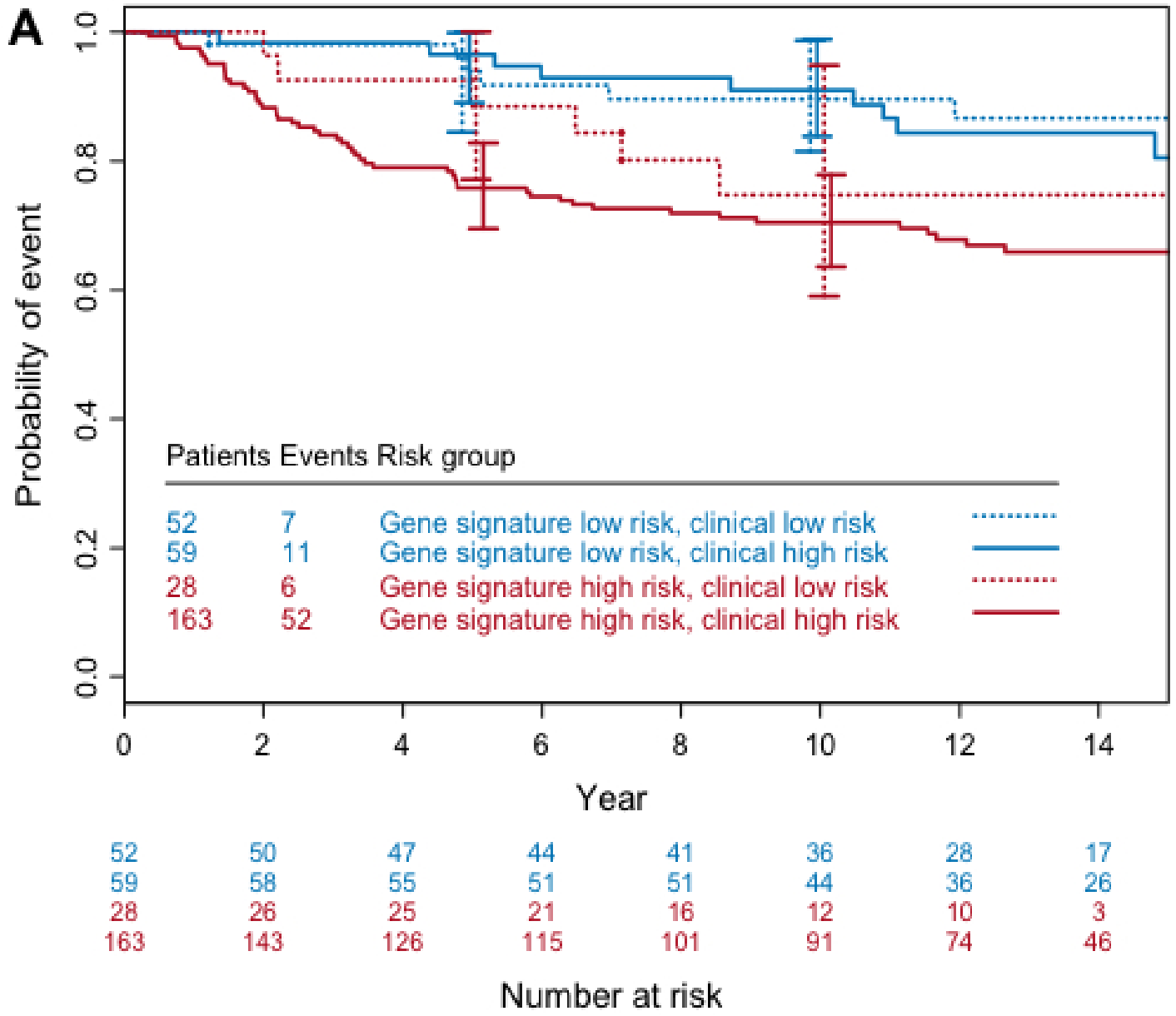
Session Outline

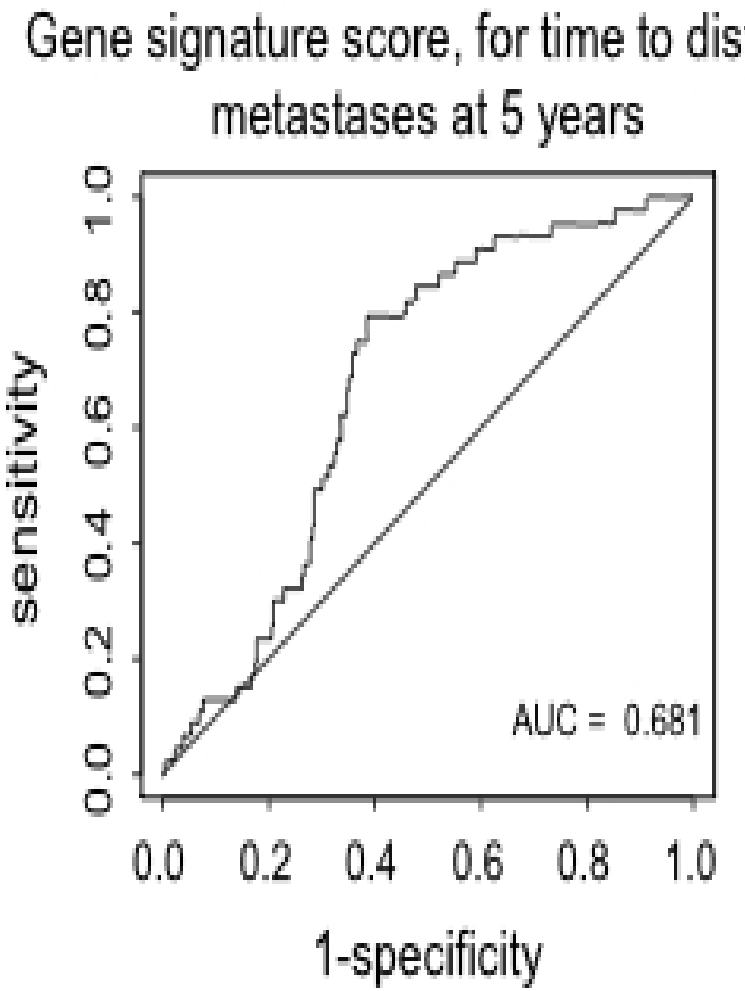
- Examples
  - ▷ Breast Cancer: 70 gene prediction / validation
  - ▷ HIV: markers of disease progression
- Competing Risks Data
- $TP^{\mathbb{C}}$ and cause-specific endpoints / Estimation (non-parametric)
- $TP^{\mathbb{I}}$ and cause-specific endpoints / Estimation (semi-parametric)
- Illustration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

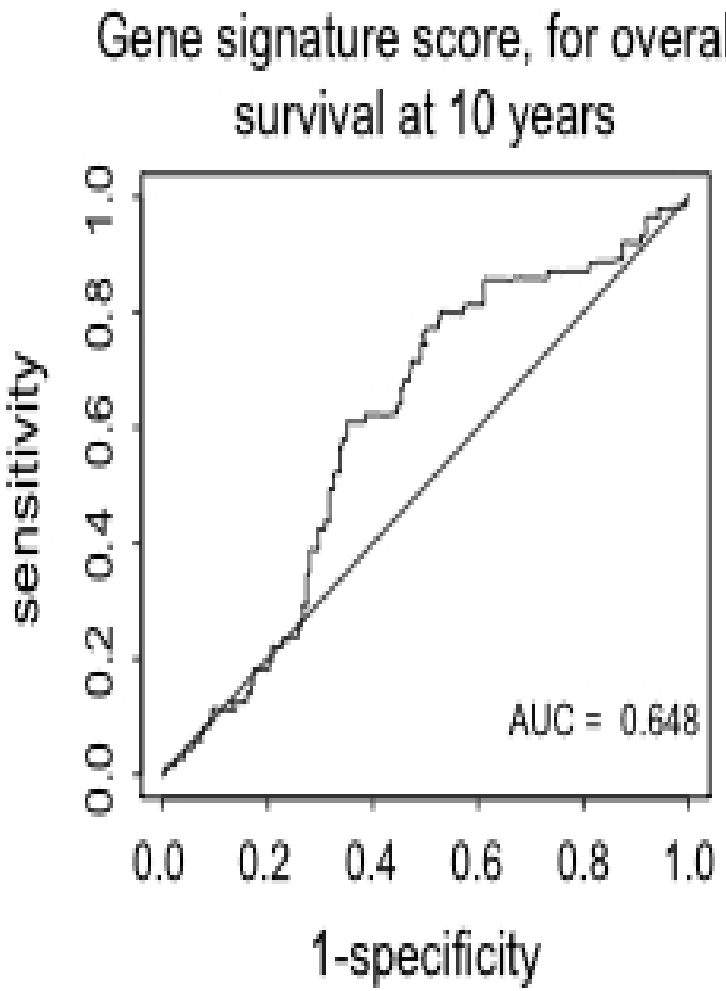
Example: BC and 70-gene Signature among Node-negative

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

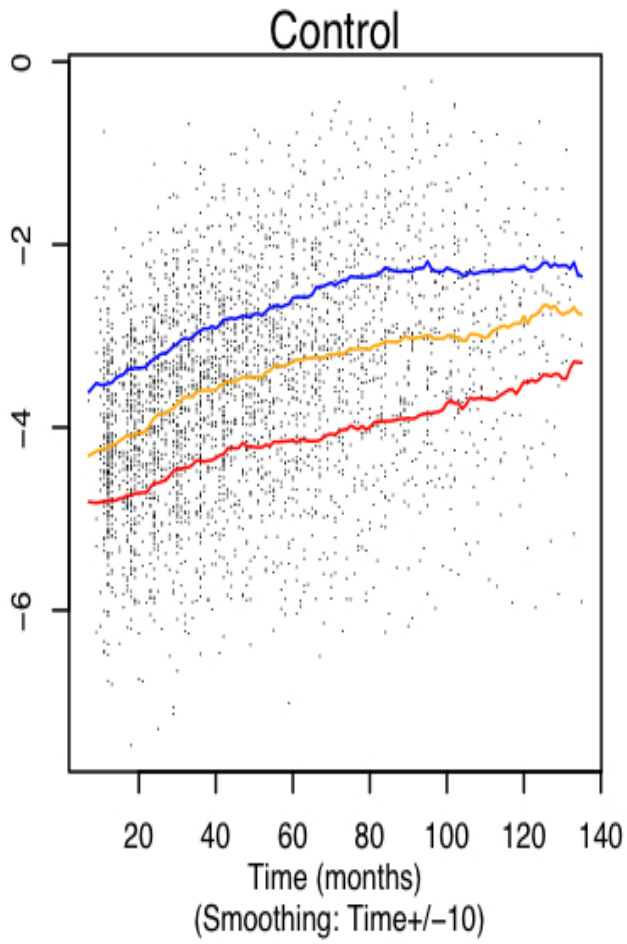

Gene signature score, for time to distant metastases at 5 years
sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.681

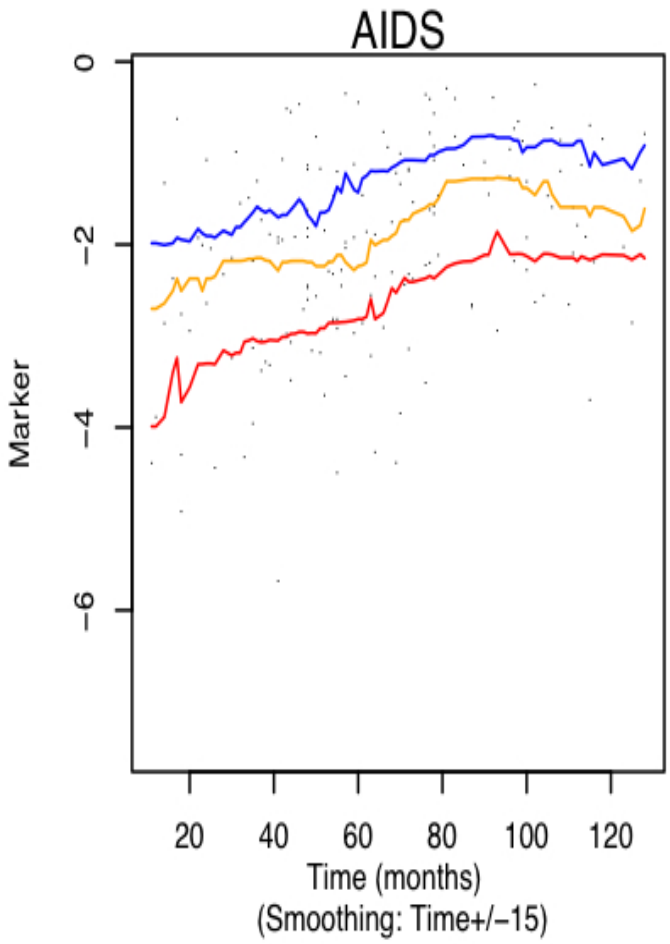

Gene signature score, for overall survival at 10 years
sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.648

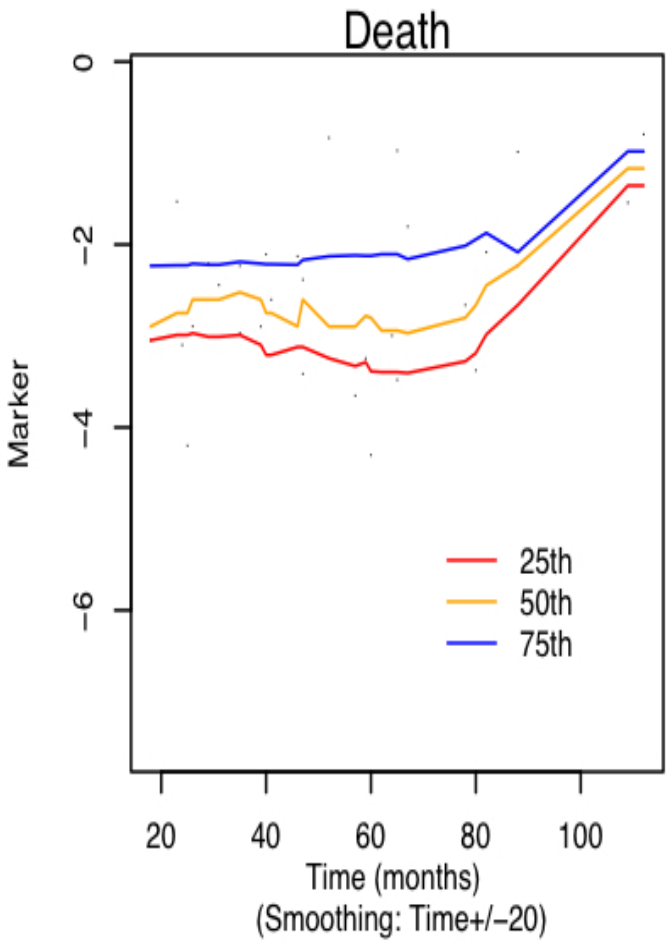
Example: Immune markers and disease progression

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 (2011)


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

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

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

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

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:

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

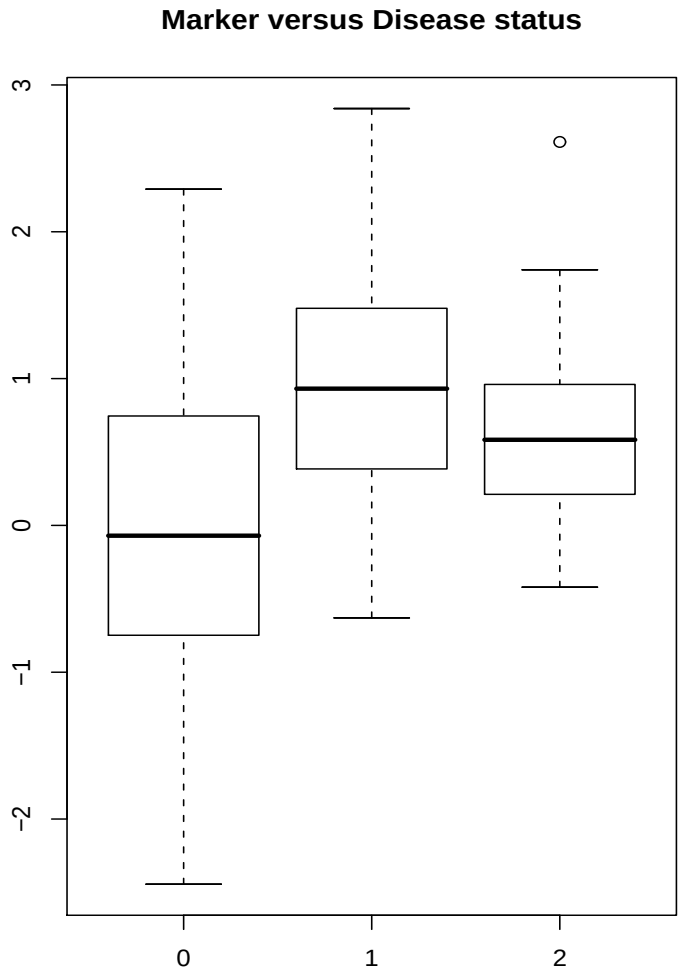
$$\begin{aligned} \text{numerator} \quad &= \quad \int_c^\infty P(T \leq t, \delta = d \mid M = m) \cdot P(M = m)\, dm \\ &= \quad \int_c^\infty C_d(t \mid M = m) \cdot P(M = m)\, dm \end{aligned}$$

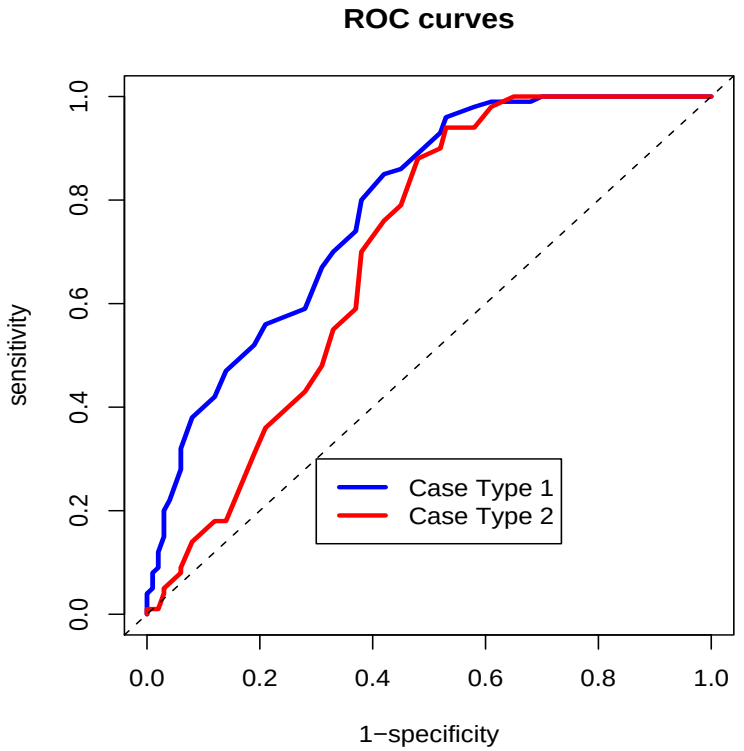
Estimation: Using "local" Cumulative Incidence

- 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.

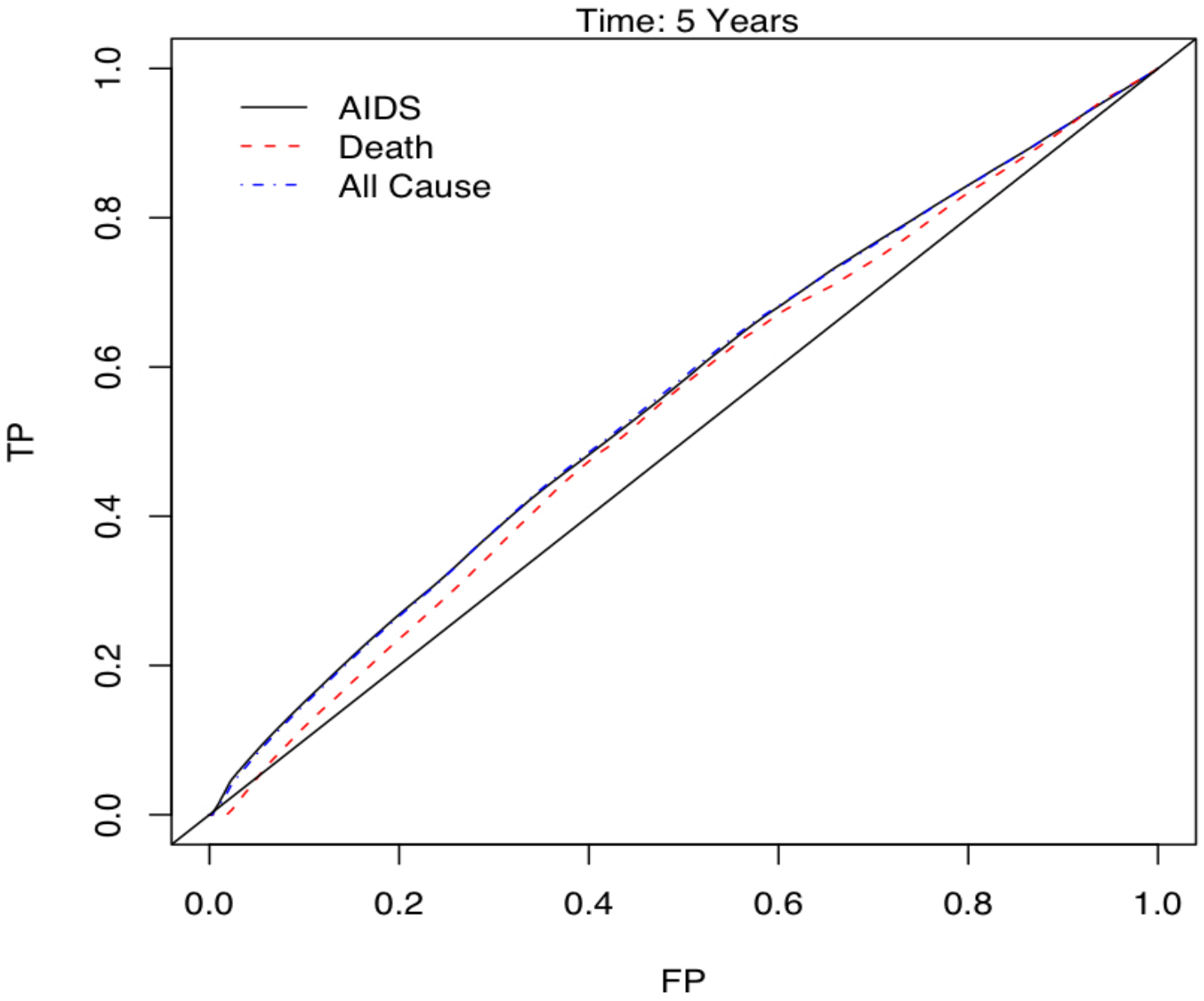

Marker versus Disease status
3
2
1
0
−1
−2
0
1
2


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 (2011)

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

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

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

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

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

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

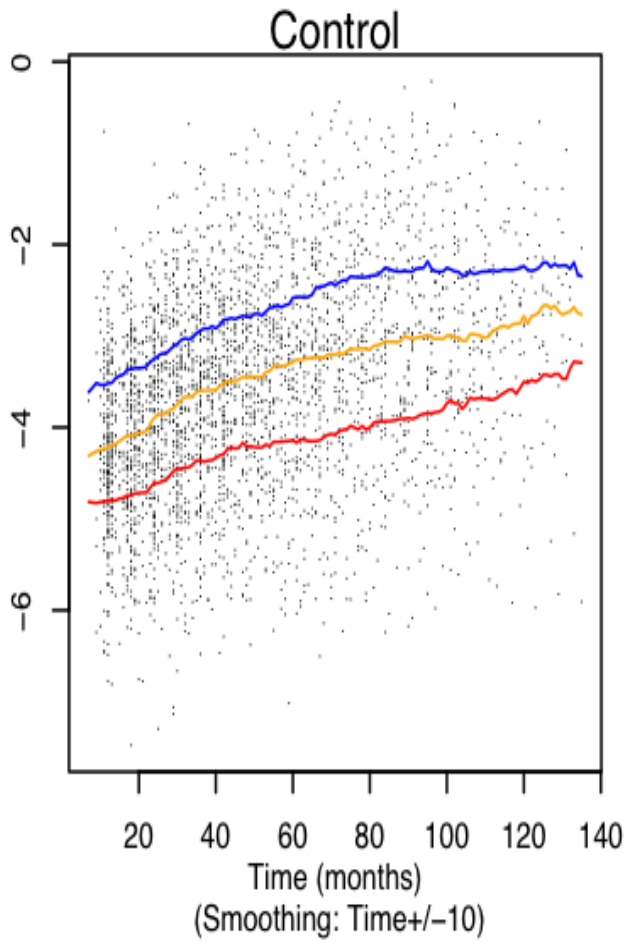
$$\text{Estimate} \quad \Longleftarrow \quad \text{Smooth model} \quad + \quad \text{Empirical}$$

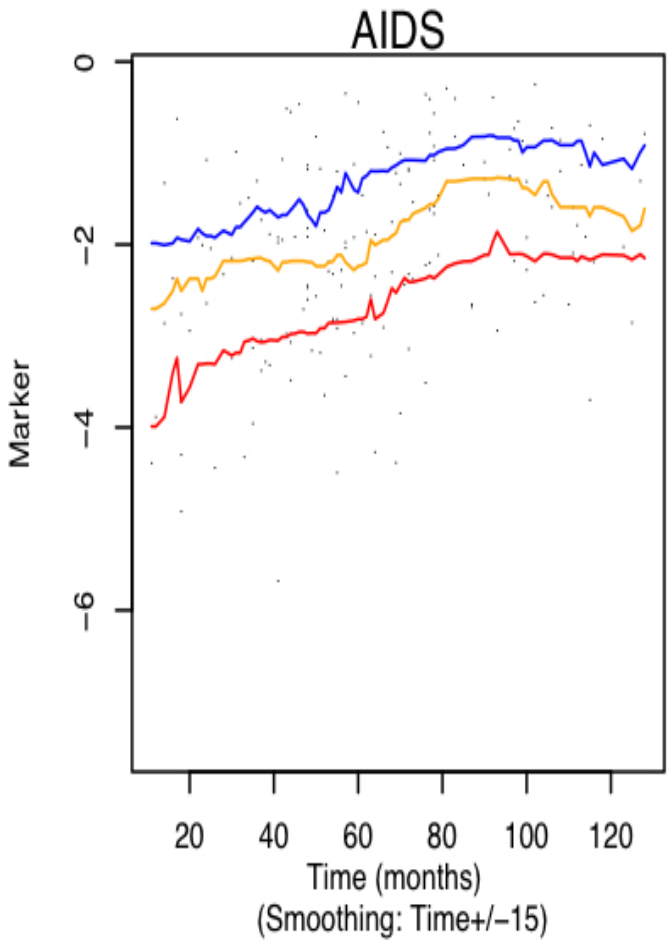
Note: direct (easy) generalization of the HZ(2005) methods.

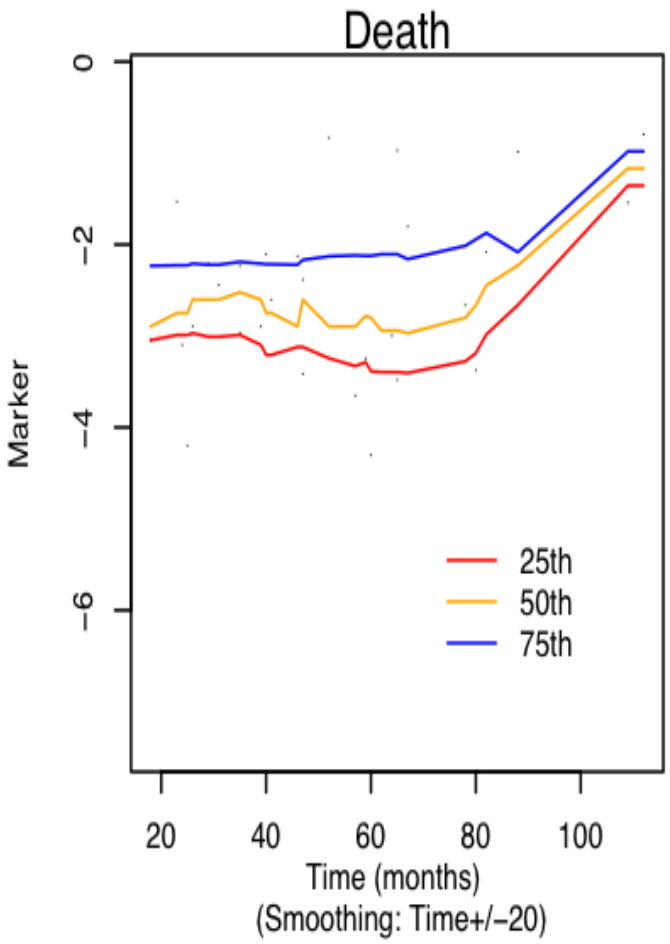
Software / Illustration

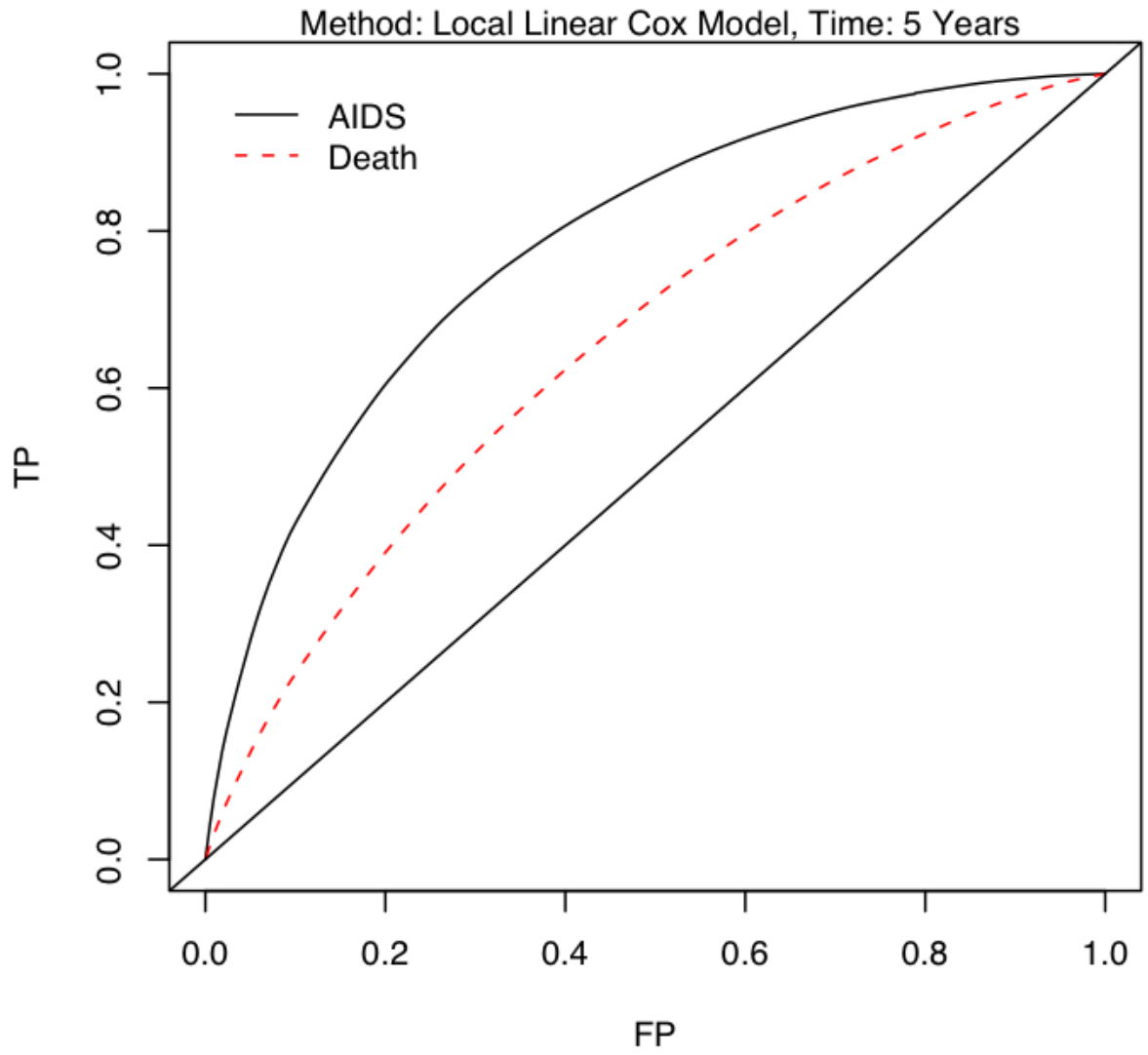
- **Software**: CRAN package for R called `risksetROC` – we have extended this to implement the competing risks calculations, and to handle time-dependent covariates. (P. Saha)

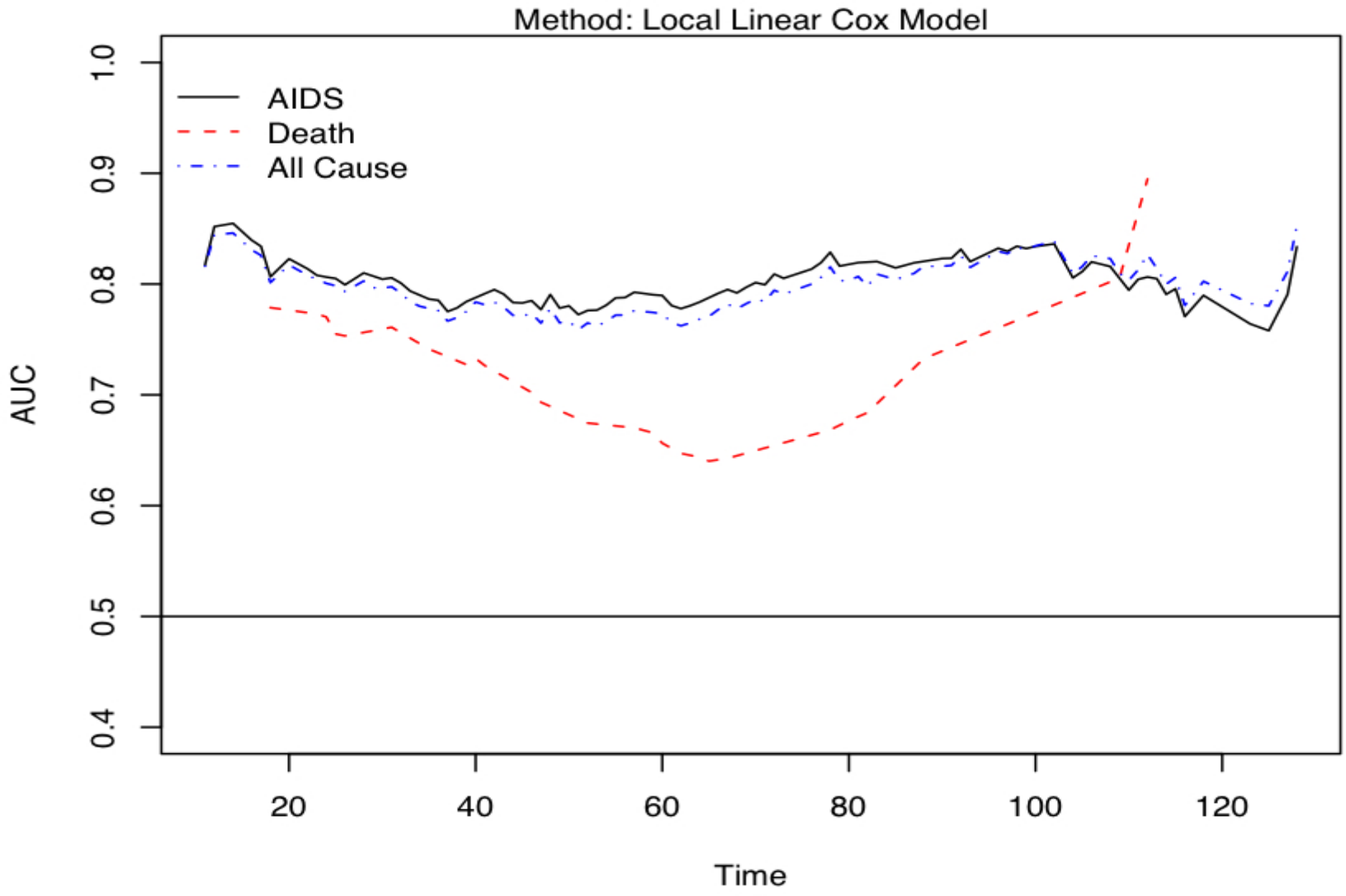
- **MACS Data**
  - Longitudinal values of CD4 and CD8
  - Linear combination based on Cox regression
  - Case Type 1 = AIDS (n=176)
  - Case Type 2 = death before AIDS (n=34)
  - ROC curve, and AUC versus time


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.

Motivation: Treatment Prioritization

- Organ transplantation seeks to prioritize limited donor organs by identifying those subjects who are at risk of death without intervention (and who would do well if transplanted).
  - ▷ Lung Allocation Score (see Gries et al. 2010)
  - ▷ MELD Score (Model for Endstage Liver Disease)
- The scientific goal is one where **over time** a good model/marker would identify those subjects at risk of death (from among those still at-risk).
- **Q**: Where do diseased subjects who die **rank** among those in the risk set?

Development of the Allocation System for Deceased Donor Liver Transplantation

John M. Coombes, MD and James F. Trotter, MD

As the number of pre- and post-transplant solid organ recipients continues to grow, it becomes important for all physicians to have an understanding of the process of organ procurement and allocation. In the United States, the current system for allocation and transplantation of human solid organs has been heavily influenced by the experience in deceased donor liver transplantation (DDLT). This review highlights the significant changes that have occurred over the past 10 years in DDLT, with specific attention to the impact of the Model for Endstage Liver Disease (MELD) score on organ allocation and pre- and post-transplant survival.

DDLT is managed by the United Network for Organ Sharing (UNOS) which oversees organ procurement and allocation across geographically defined Organ Procurement Organizations (OPOs). For many years, deceased donor livers were allocated to waiting list patients based on subjective parameters of disease severity and accrued waiting time. In addition, organs have traditionally been retained within the OPO where they are procured contributing to geographic disparities in disease severity at the time of transplantation among deceased donor recipients.

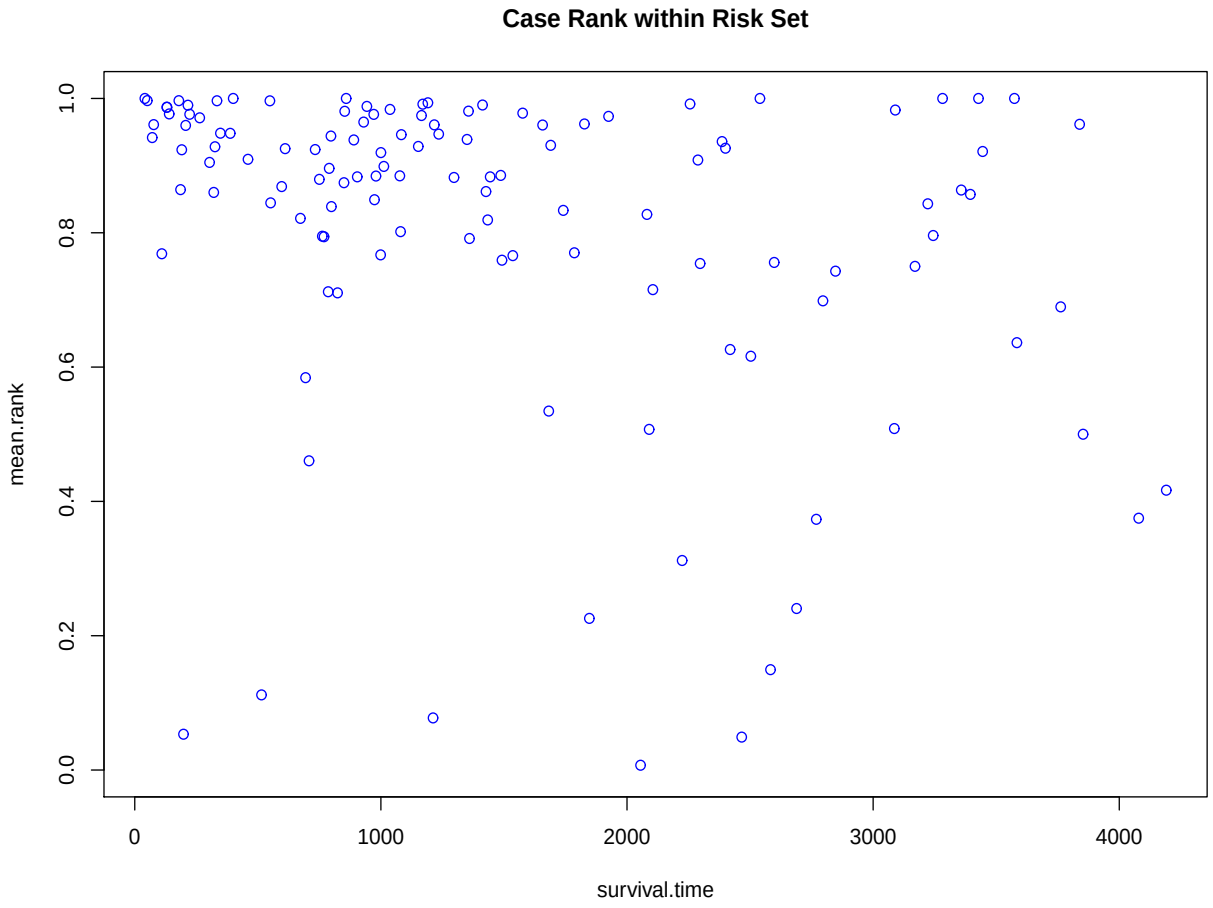
In response to a perceived unfairness in organ allocation, Congress issued its "Final Rule" in 1998. The Rule called for a more objective ranking of waiting list patients and more parity in disease severity among transplant recipients across OPOs. To date, little progress has been made in eliminating geographic inequities. Patients in the smallest OPOs continue to receive liver transplants at a lower level of disease severity. However, strides have been made to standardize assessments of disease severity and better prioritize waiting list patients. The MELD score has emerged as an excellent predictor of short-term mortality in patients with advanced liver disease, and patients listed for liver transplantation are now ranked based on their respective MELD scores. This has improved organ access to the most severely ill patients without compromising waiting list mortality or post-transplant survival.

Weighted Mean Rank: Motivation

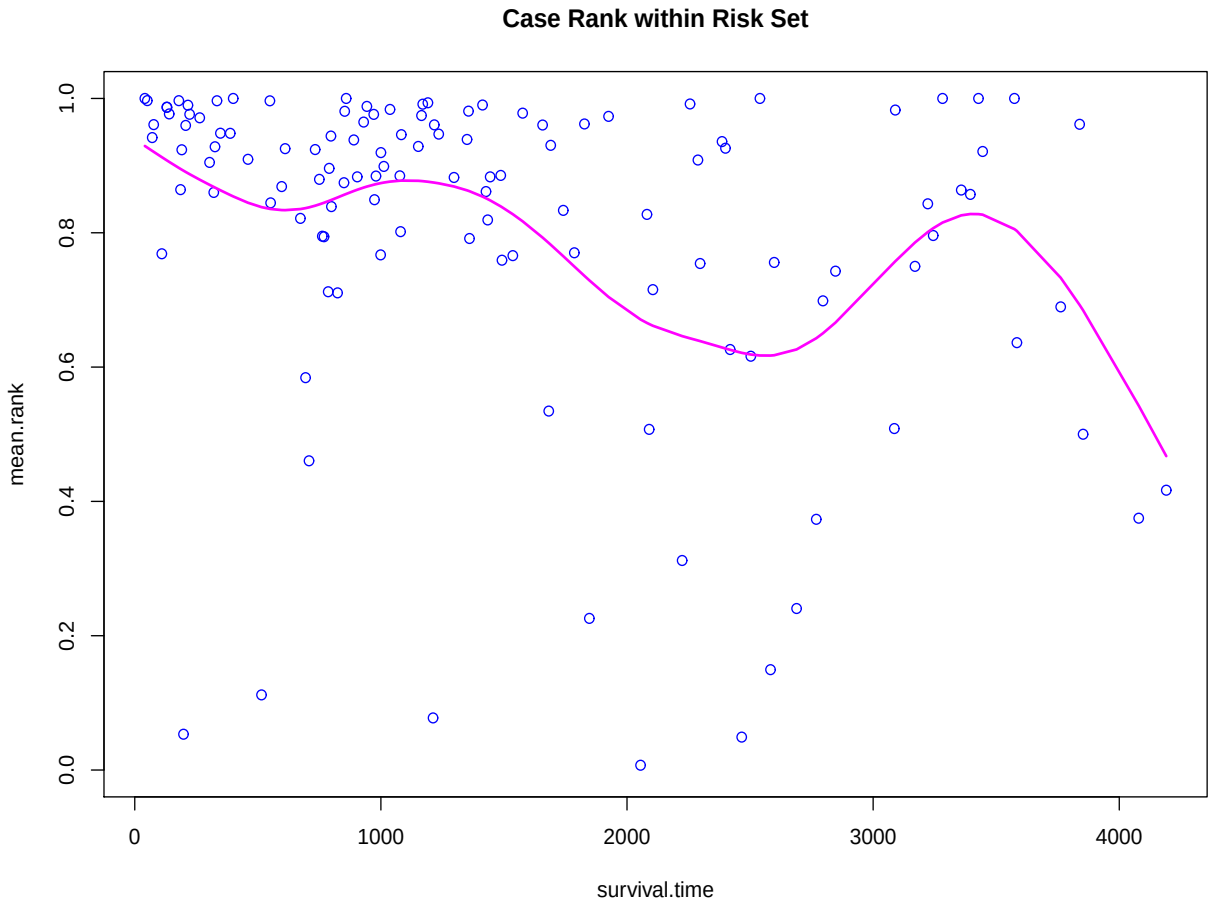
- **Descriptive**:
  - ▷ **Q**: Where does the CASE rank among members of risk set?
  - ▷ **Q**: If we considered the top 10% of CONTROL marker values within a risk set then what is the probability that the CASE is within the top 10%?
- **Connection**:
  - ▷

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

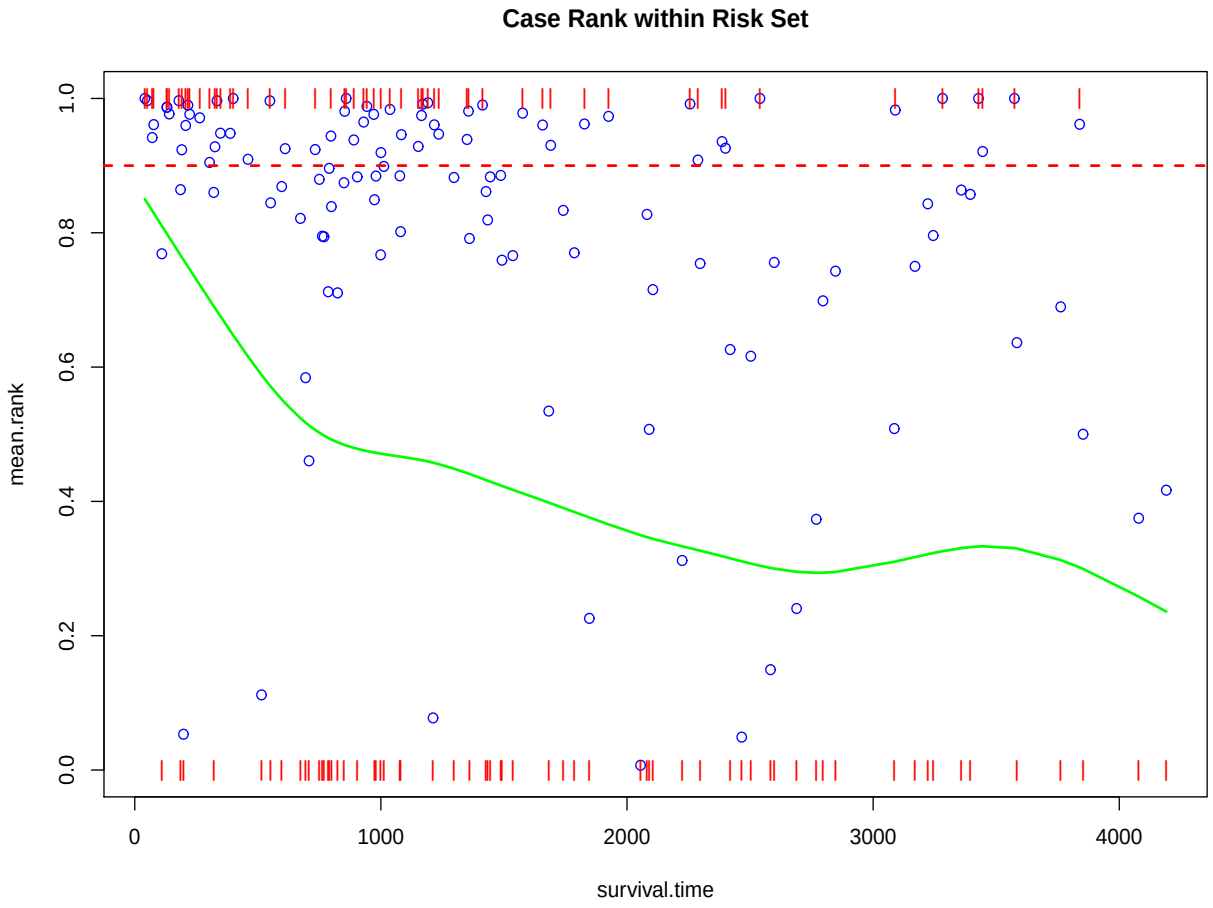
Example: PBC and Model(5) Score



Example: PBC and Model(5) Score



Example: PBC and Model(5) Score



Case Rank within Risk Set

- **Define**:
  - Rank **CASE** relative to **CONTROLS**
  - Controls: $\mathcal{R}^0(t) = \{k \ : \ N_k(t) = 0 \ \}$
  - Let $n_t^0 = |\mathcal{R}^0(t)|$

$$M^*(t) \quad = \quad M_j \ \text{ for } \ dN_j(t) = 1$$

$$A(t) \quad = \quad \frac{1}{n_t^0} \sum_{k \in \mathcal{R}^0(t)} 1[M^*(t) > M_k]$$

$$A(t) \quad = \quad \frac{[\text{risk set rank of } M^*(t) \ ] - 1}{n_t^0}$$

Nonparametric Estimation of AUC(t)

- Multiple cases in a risk set leads to:
  - Let $\mathcal{R}^1(t)$ denote j such that $dN_j(t) = 1$
  - $n_t^1 = |\mathcal{R}^1(t)|$

$$A(t) = \frac{1}{n_t^1} \sum_{j \in \mathcal{R}^1(t)} \frac{1}{n_t^0} \sum_{k \in \mathcal{R}^0(t)} 1[M_j > M_k]$$

- Note that $A(t)$ is a random variable where:

$$E[A(t)] = P[M_j > M_k \mid dN_j(t) = 1, N_k(t) = 0] = AUC^{\mathbb{I}/\mathbb{D}}(t)$$

Nonparametric Estimation of AUC(t)

- **Estimation**: – Given that $A(t)$ is a random variable smoothing, or **local averages** can be used to estimate $AUC(t)$:
  - ▷ Define a neighborhood of time t based on a sample-size dependent bandwidth h_n.

$$\text{e.g.} \quad \mathcal{N}_t(h_n) = [t - h_n, t + h_n]$$

  - ▷ Compute a local average:

$$\widehat{AUC}(t) = \frac{1}{|\mathcal{N}_t(h_n)|} \sum_{t_j \in \mathcal{N}_t(h_n)} A(t_j)$$

Nonparametric Estimation of AUC(t)

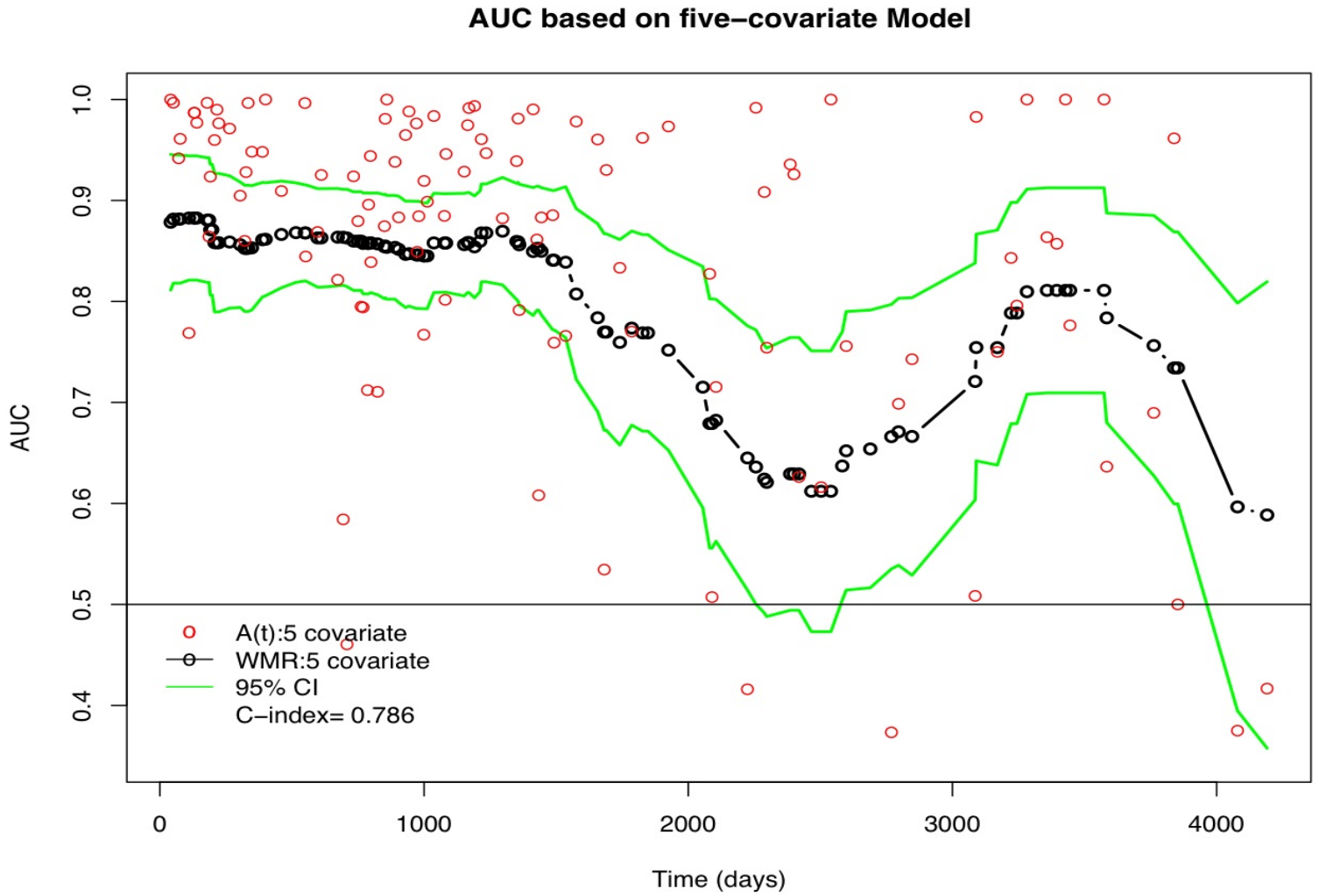
- More generally use a kernel function to obtain a weighted average
  - $K_{h_n}(x)$ is a kernel function with bandwidth h_n
- Define the **Weighted Mean Rank** (WMR) Estimator:

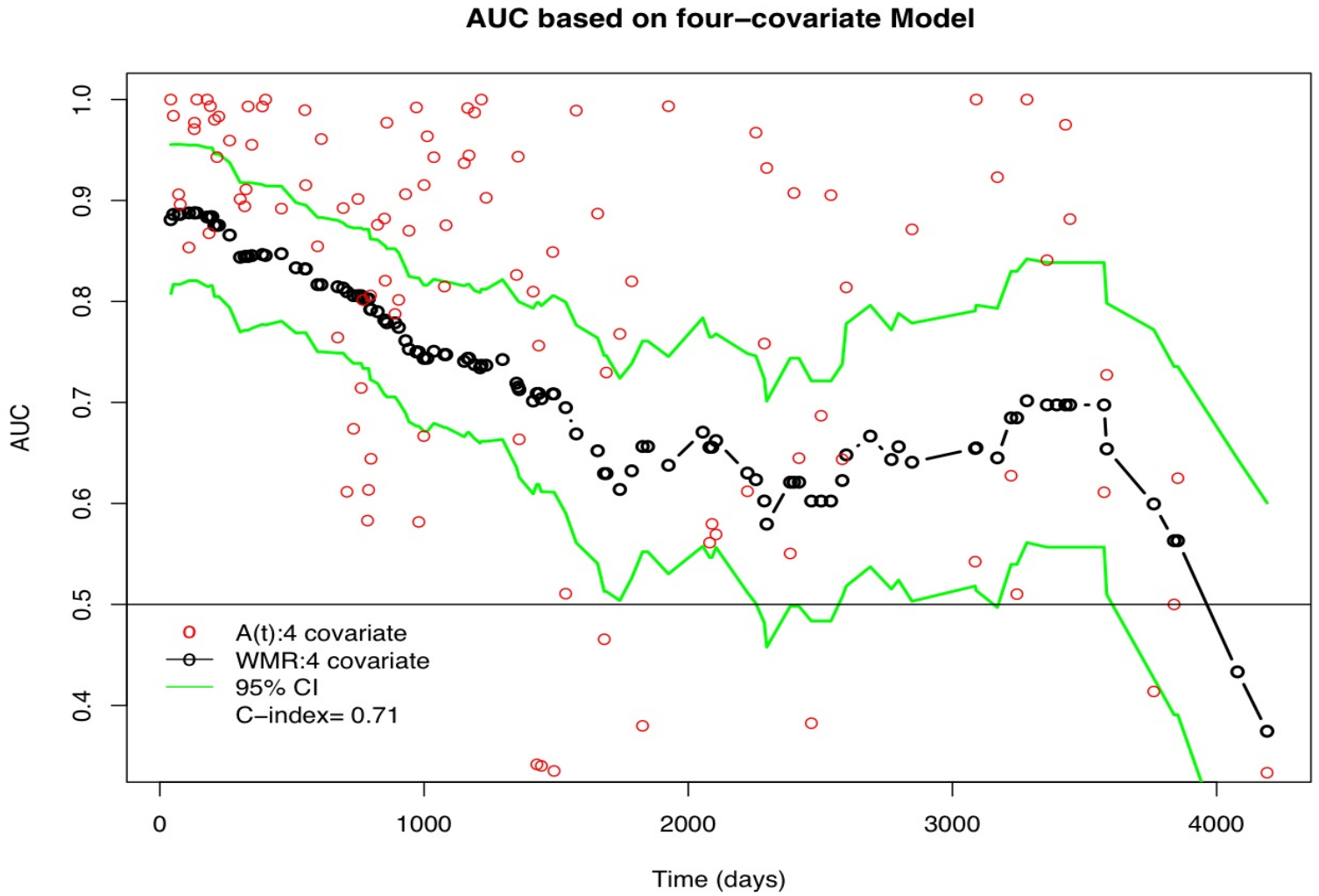
$$\widehat{AUC}(t) = \sum_j K^*_{h_n}(t_j - t) \cdot A(t_j)$$

  - Where K^* is normalized version of kernel function such that $\sum_j K^*_{h_n}(t_j - t) = 1$.

Asymptotic Properties of WMR Estimator

- We can show that the estimator is consistent and asymptotically normal (CAN).
- Theory uses a **U-statistic** central limit theorem.
- Variance estimation is obtained based on analytical expressions that are straight-forward to compute.
- Data-driven bandwidth – we have implemented a jackknike cross-validation method to estimate the integrated mean squared error (IMSE) and can choose a bandwidth to minimize this criterion.


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

Comparison of Markers

- A key use of ROC and AUC methods is to compare the prognostic potential of different markers.
- Data used for comparison is paired: (M_{i1}, M_{i2})
- Using WMR methods we simply compute locally weighted averages of the difference:

$$\begin{aligned}
\widehat{AUC}_1(t) &= \sum_j K^*_{h_n}(t_j - t) \cdot A_1(t_j) \\
\widehat{AUC}_2(t) &= \sum_j K^*_{h_n}(t_j - t) \cdot A_2(t_j) \\
\widehat{D}_{12}(t) &= \sum_j K^*_{h_n}(t_j - t) \cdot [A_1(t_j) - A_2(t_j)]
\end{aligned}$$

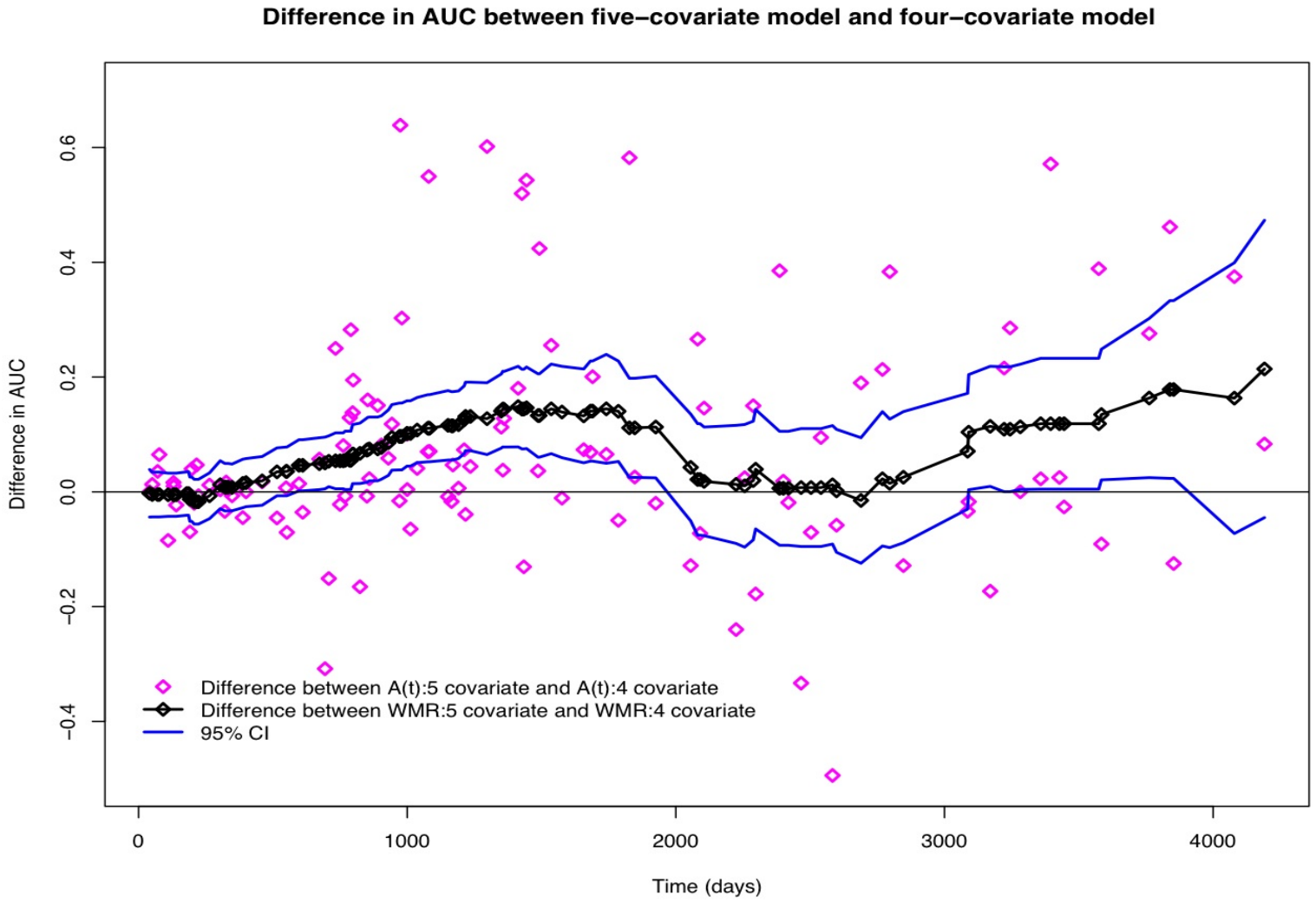

Difference in AUC between five−covariate model and four−covariate model
Difference in AUC
0.6
0.4
0.2
0.0
−0.2
−0.4
0
1000
2000
3000
4000
Time (days)
Difference between A(t):5 covariate and A(t):4 covariate
Difference between WMR:5 covariate and WMR:4 covariate
95% CI

Table 3: Simulation results for comparison of semi-parametric approach of Heagerty and Zheng (2005) (HZ2005) and non-parametric approach WMR. We assumed $(\log T, M)$ follows a mixture of two multivariate normal distributions: $\left(\log T^{(I)}, M^{(I)}\right) \sim N_2(-1.5, -1.5, 1, 1, 0)$ if $Z = 1$ and $\left(\log T^{(N)}, M^{(N)}\right) \sim N_2(0, 0, 1, 1, -0.8)$ if $Z = 0$ where $Z \sim$ Bernoulli (0.2). We show the estimated HZ2005, WMR, MCSDs and the SD estimated using the proposed variance estimator (EstSD) and the coverage (nominal: 95.0) for WMR.

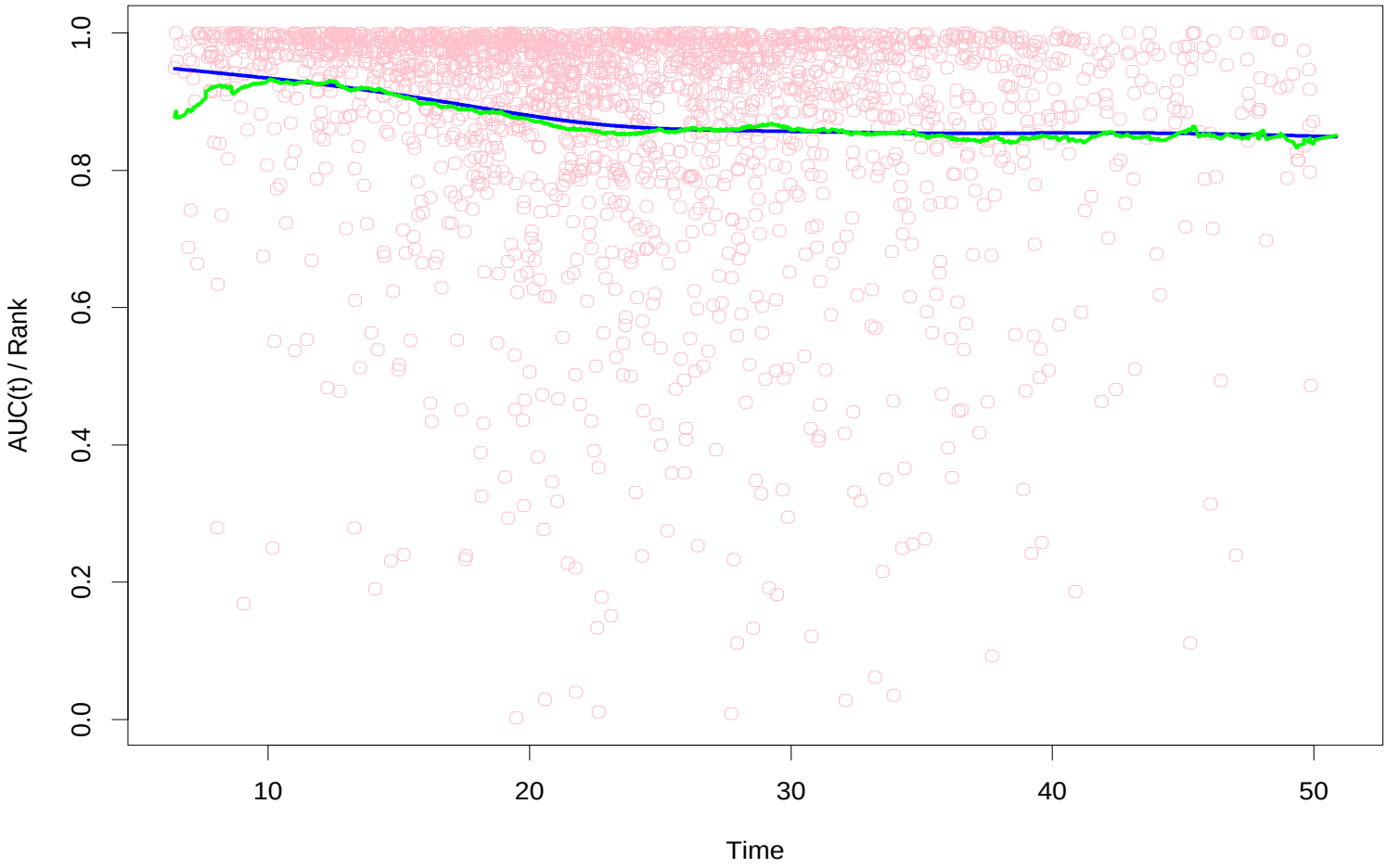
Log time	AUC(t)	HZ2005	MCSD	WMR	MCSD	EstSD	Coverage
-2.5	0.378	0.173	0.069	0.376	0.109	0.097	87.7
-2.0	0.481	0.346	0.087	0.477	0.083	0.075	91.1
-1.5	0.591	0.551	0.071	0.595	0.062	0.057	92.0
-1.0	0.673	0.660	0.048	0.674	0.048	0.043	91.6
-0.5	0.709	0.689	0.032	0.708	0.035	0.034	93.7
0.0	0.709	0.684	0.023	0.710	0.030	0.030	94.8
0.5	0.691	0.666	0.022	0.692	0.034	0.033	92.5
1.0	0.669	0.646	0.022	0.669	0.042	0.041	93.9

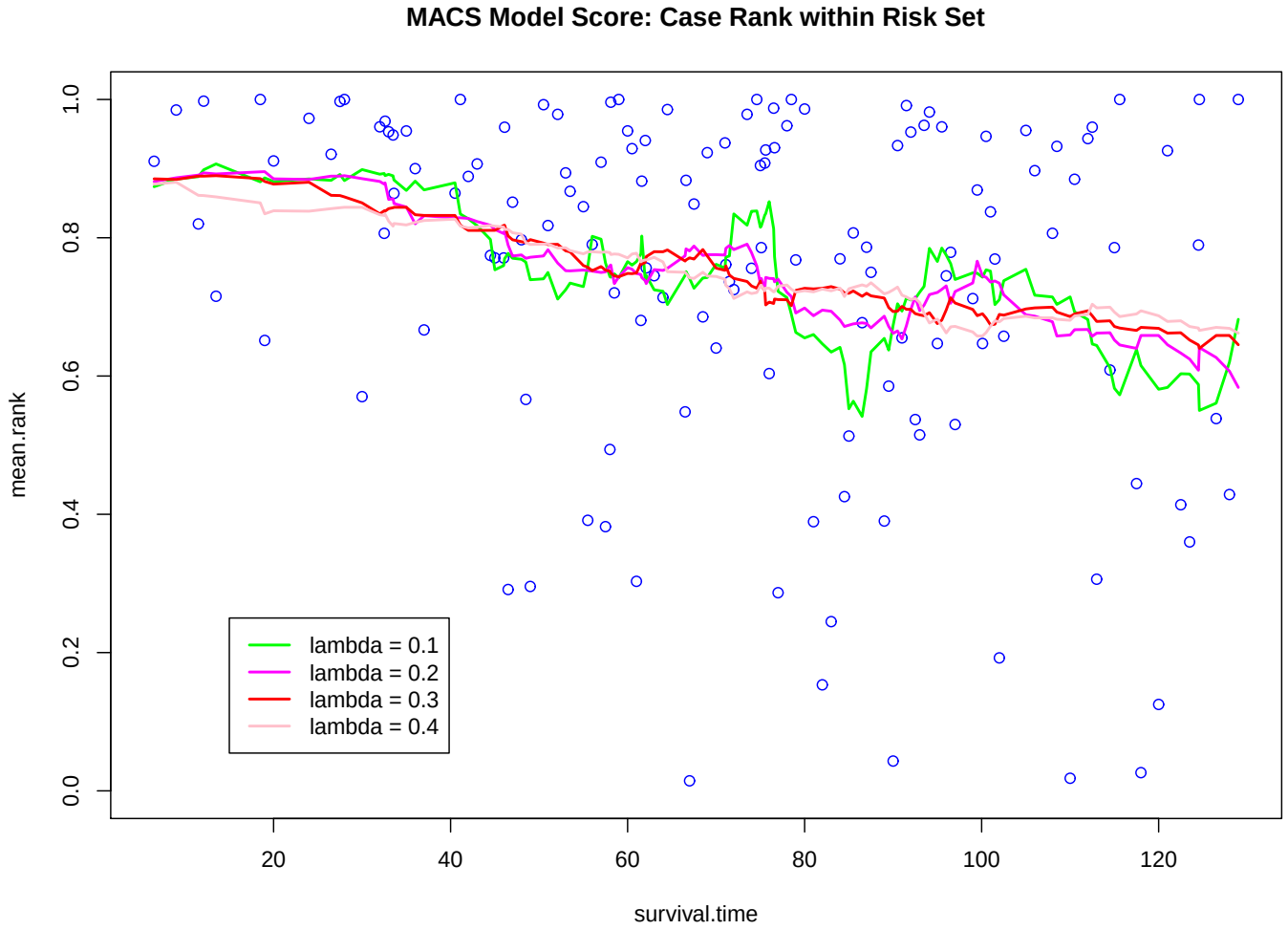
Additional Comments

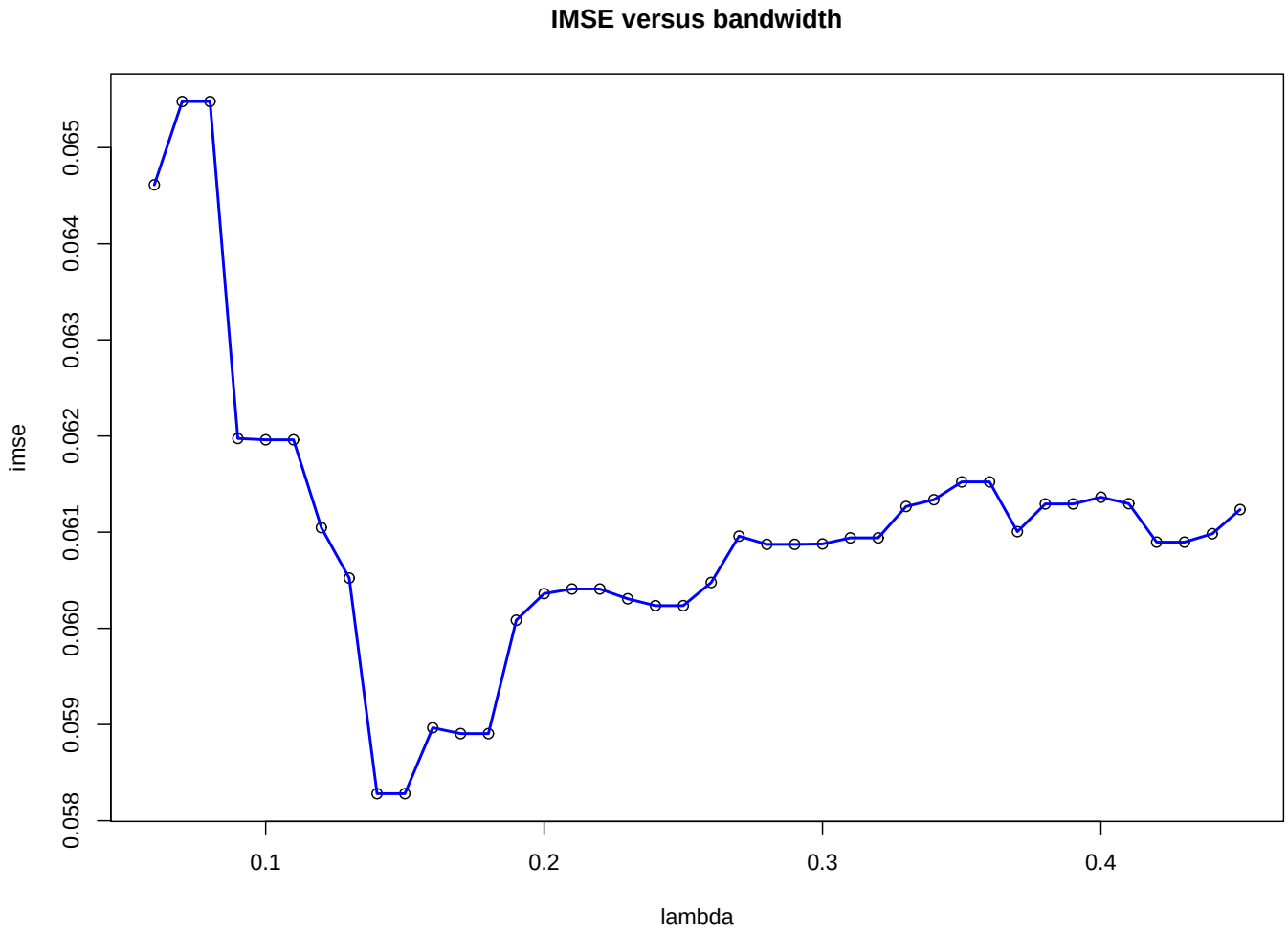
- For a baseline marker the **C-index** can also be estimated as the global weighted average:

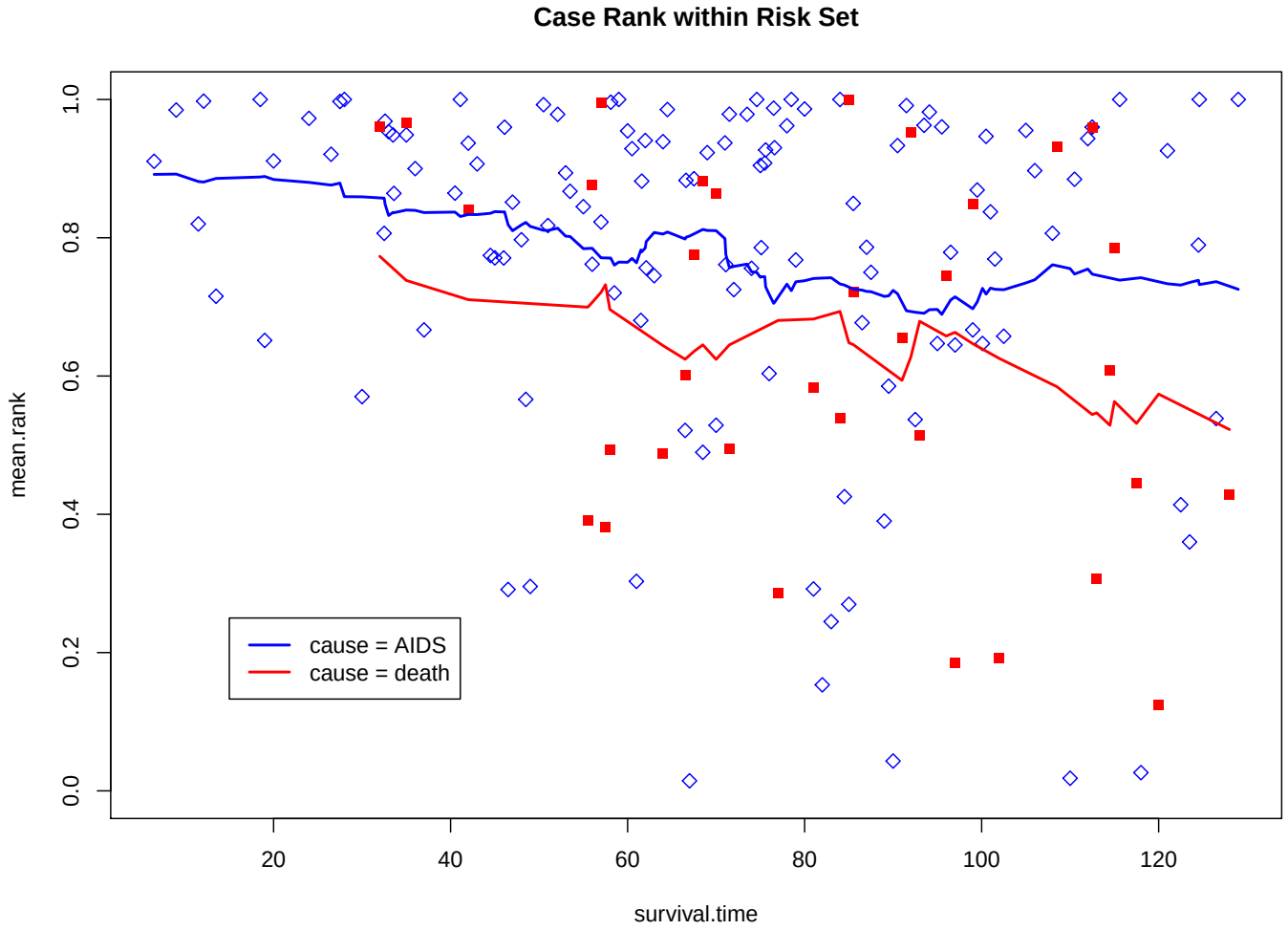
$$\widehat{C} = \int A(t) \cdot 2\widehat{f}(t)\widehat{S}(t)\, dt$$

- We can also directly apply the WMR estimator to time-dependent covariates, $M(t)$, since the method is based on risk-sets and the case rank within the riskset.
- **Time-dependent Covariate Example**:
  - ▷ Cystic Fibrosis Data
  - ▷ FEV1, height, and weight are time-dependent
  - ▷ Compare semi-parametric estimate of HZ(2005) to WMR


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


MACS Model Score: Case Rank within Risk Set
mean.rank
survival.time
0.0
0.2
0.4
0.6
0.8
1.0
20
40
60
80
100
120
lambda = 0.1
lambda = 0.2
lambda = 0.3
lambda = 0.4


IMSE versus bandwidth
imse
0.058
0.059
0.060
0.061
0.062
0.063
0.064
0.065
0.1
0.2
0.3
0.4
lambda


Case Rank within Risk Set
mean.rank
survival.time
0.0
0.2
0.4
0.6
0.8
1.0
20
40
60
80
100
120
cause = AIDS
cause = death

Some Current Extensions

- Recall that one of our motivating questions asked:
  - ▷ **Q**: How often does the CASE marker rank in the top 10% of the risk set (or among controls)?
- This concept is directly connected to sensitivity:

$$TP^{\mathbb{I}/\mathbb{D}}(p,t) = P[M_j > c^p \mid T_j = t]$$

$$E\{1[A(t) > (1-p)]\} = \textbf{CASE}(\text{t}) > (\text{1-p})\% \text{ of } \textbf{CONTROLS}(\text{t})$$

$$= TP^{\mathbb{I}/\mathbb{D}}(p,t)$$

Some Current Extensions

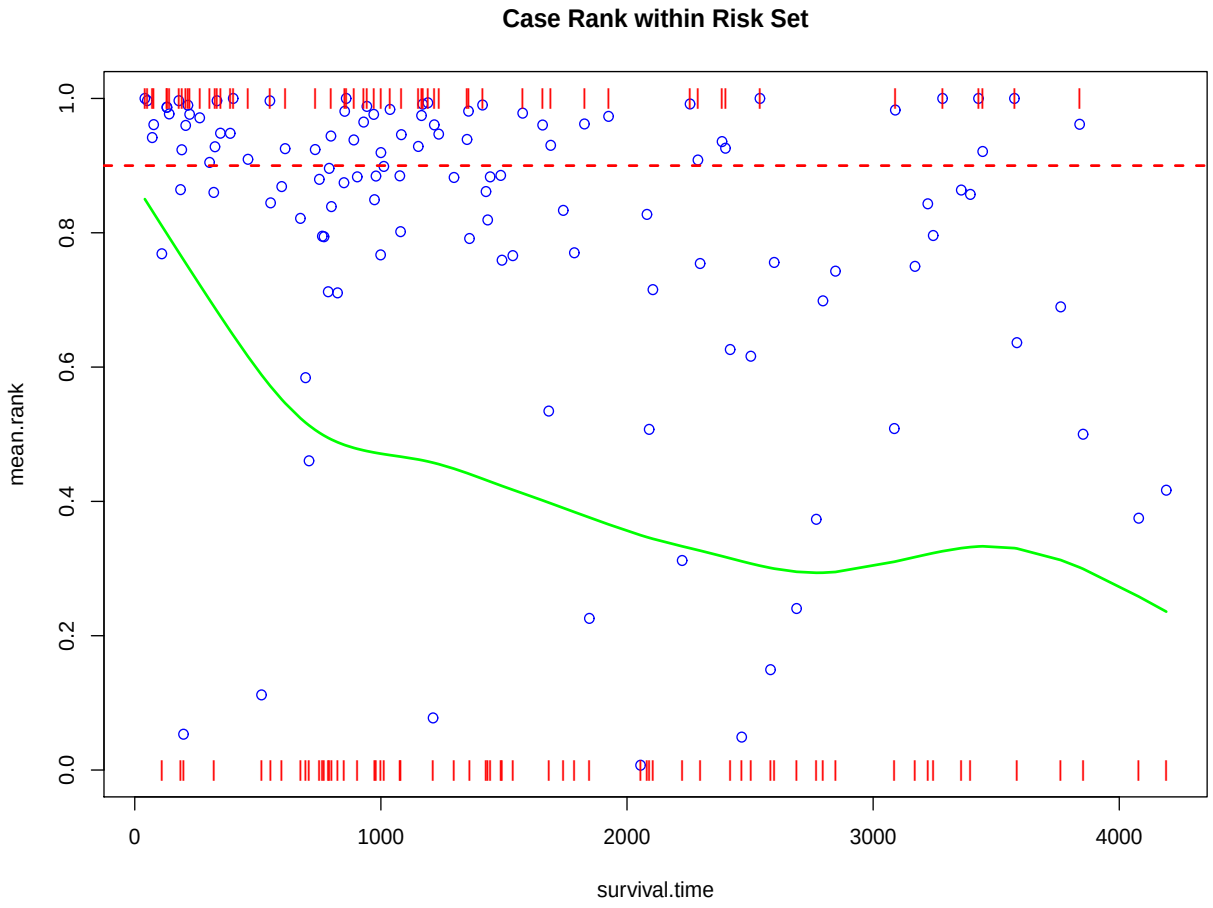
- **Non-Parametric** $\widehat{TP}^{\mathbb{I}/\mathbb{D}}$
  - ▷ Select a value of false-positive rate: p
  - ▷ Derive the indicators:

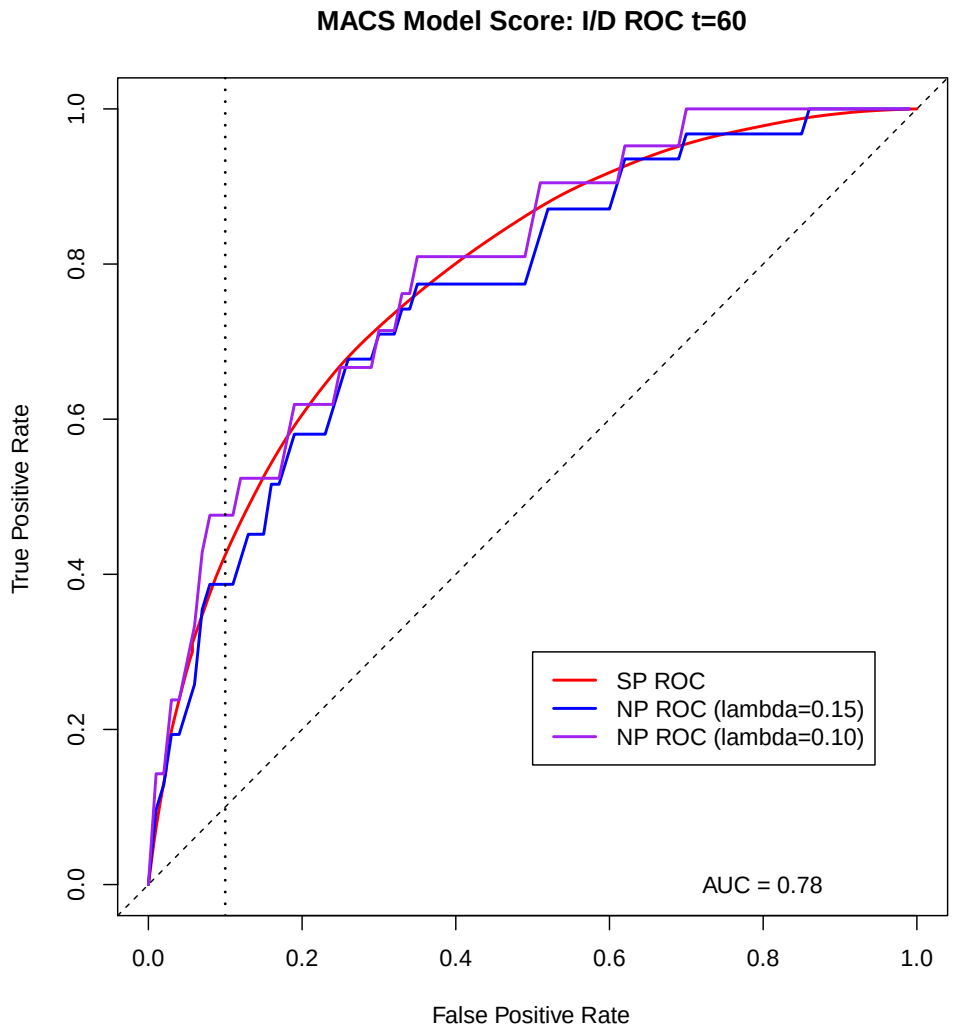
$$H(t,p) = 1[\ A(t)\ >\ (1-p)\ ]$$

  - ▷ Locally weighted averages to obtain smooth curve in time:

$$\widehat{TP}^{\mathbb{I}/\mathbb{D}}_{h_n}(t,p) = \sum_j K^*_{h_n}(t-t_j)\cdot H(t,p)$$

Example: PBC and Model(5) Score




MACS Model Score: I/D ROC t=60
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
SP ROC
NP ROC (lambda=0.15)
NP ROC (lambda=0.10)
AUC = 0.78

Summary

- The **Case rank** is a descriptive summary that is clinically meaningful.
- Using the case-rank provide a basis for non-parametric estimation of time-dependent accuracy summaries.
- WMR provides non-parametric estimation with analytical expressions for standard errors.
- Methods extend to allow **time-dependent markers**.
- Methods extend to estimation of time-dependent sensitivity.

Summary

Accuracy summary			Estimation
$ROC_t^{\mathbb{I}/\mathbb{D}}(p)$	:	vary (M,t)	SP, NP
$AUC(t)$	:	vary (t)	SP, NP
$\overline{ROC}(p)$	:	vary (M)	SP, NP
C	:	global	SP, NP

Thanks!