Discrimination measures for survival outcome

V. Viallon [1] and A. Latouche [2]

[1]Université de Lyon 1, IFSTTAR
[2]Conservatoire national des arts et métiers

19 octobre 2011

Criteria for evaluating prognostic models

Discrimination

- Measures the ability to distinguish the individuals who developped the disease and those who did not
- The area under the receiver operating curve (AUC) is a standard tool for evaluating the discrimination of prognostic model

Calibration

- The calibration categorizes patients according to quantiles of risk (according to the model)
- Compares (average) predicted risk with the observed proportion of events in each quantile

Outline

- Extension of the AUC to survival outcomes
- A novel estimator of the time-dependent AUC based on the predictiveness curve
- A simulation study comparing the derived estimator to Heagerty and Zheng (Bcs, 2005), Chambless and Diao (SiM, 2006) proposals
- Illustration

AUC

- For a continuous (bio)marker X and a binary outcome D
- ROC plots sensitivity, $\mathrm{P}(X > c|D = 1)$, against 1 minus specificity, $1 - \mathrm{P}(X \leq c|D = 0)$, for all possible values c
- The AUC is then simply the area under ROC

AUC extensions

- Harrel's concordance index: the fraction of pairs of patients whose predicted survival times are correctly ordered among all pairs that can actually be ordered
- Gonen (Bka, 2005) derived an analytical expression of the c-index under the Cox model leading to an estimator that is not affected by censoring

Time–dependent ROC curves and AUC(t)

In prospective cohort study, a binary outcome can change over time e.g. a disease status $\Rightarrow$ legitimate to consider **time-dependent ROC curve**

- Heagerty et al. defined time-dependent sensitivity and specificity

- Leads to distinct definitions of the time-dependent ROC curves and time-dependent AUC, $AUC(t)$.

Heagerty and Zheng Taxonomy

Let T_i denotes the survival time for subject i

- *Cases* are said to be
  - *incident cases* where $T_i = t$, is used to define cases at time t
  - *cumulative cases* where $T_i \leq t$ is used.
- *Controls* are said to be
  - *static controls* when $T_i > t^\star$ for a fixed $t^\star$ is used to define them
  - *dynamic controls* when $T_i > t$ is used.

This talk focus on Cumulative/Dynamic:

Discriminating between subjects who die prior to a given time t' and those survive beyond t'

Some Notations for AUC

- Let T_i and C_i denote survival and censoring times for subject i
- We observe (Z_i, δ_i) where $Z_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$
- Denote $D_i(t)$ the time-dependent outcome status for subject i at time t

For any threshold c, the true positive and false positive rates are time-dependent functions defined as

- $\text{TPR}(c, t) = \text{P}(X > c | D(t) = 1)$
- $\text{FPR}(c, t) = \text{P}(X > c | D(t) = 0)$

The time-dependent ROC curve ROC(t) plots

- TPR(c, t) vs
- FPR(c, t) for any threshold c

so that

$$\text{AUC}(t_0) = \int_{-\infty}^{\infty} \text{TPR}(c, t_0) d\left[\text{FPR}(c, t_0)\right], \tag{1}$$

where $d\left[\text{FPR}(c, t_0)\right] = \partial c \times (\partial \text{FPR}(c, t_0)/\partial c)$.

Cumulative cases and Dynamic controls

The time-dependent outcome status $D_i(t) = 1\{T_i \leq t\}$

- Cumulative true positive rates are
  $\text{TPR}^{\mathbb{C}}(c, t) = \text{P}(X > c | T \leq t) = \text{P}(X > c | D_i(t) = 1)$

- Dynamic false positive rates are
  $\text{FPR}^{\mathbb{D}}(c, t) = \text{P}(X > c | T > t) = \text{P}(X > c | D_i(t) = 0)$

Estimators can not be directly derived from the above definitions as $D_i(t)$ is not fully observable with censoring

Work around for $AUC^{\mathbb{C},\mathbb{D}}$

Using Bayes's theorem

$$AUC^{\mathbb{C},\mathbb{D}}(t_0) = \int_{-\infty}^{\infty}\int_{c}^{\infty} \frac{F(t_0; X = x)[1 - F(t_0; X = c)]}{[1 - F(t_0)]F(t_0)} g(x)g(c)dxdc$$

with

- $F(t) = P(T \leq t)$ be the absolute risk
- $F(t; X = x) = P(T \leq t | X = x)$ be the conditional absolute risk
- g the density function of marker X

Predictiveness curve

- (Too) Many criteria are used for evaluating discrimination
- The proportion of explained variation
- The standardized total gain
- Risk reclassification measures (Pencina, SiM, 2006)
- All express as simple functions of the predictiveness curve (Gu and Pepe, International Journal of Biostatistics, 2009)
- Let $R(q) = P\left[D = 1 | X = G^{-1}(q)\right]$ be the risk associated to the qth quantile of marker X
- The predictiveness curve plots $R(q)$ versus q

A proposal for AUC C/D for binary outcome

- let $R(q) = P\left[D = 1 | X = G^{-1}(q)\right]$ denote the conditional absolute risk associated to the q-th quantile $(G^{-1}(q))$ of marker X.

- The predictiveness curve plots $R(q)$ versus q and describes the distribution of $\mathrm{P}(D = 1|X)$

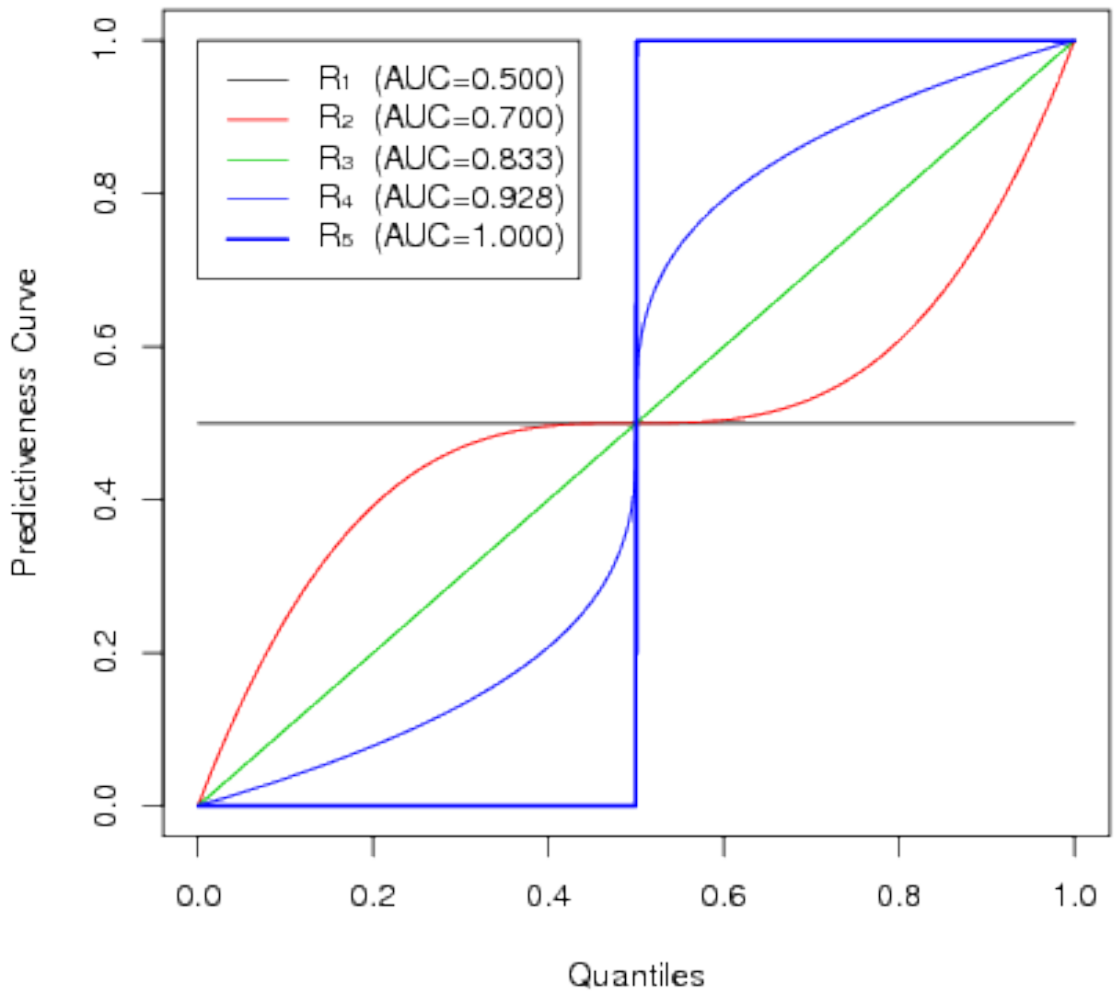
We established that

$$\mathrm{AUC} = \frac{\int_0^1 qR(q)dq - p^2/2}{p(1-p)}, \tag{2}$$

where $p = \mathrm{P}(D = 1) = \int_0^1 R(q)dq$.

Predictiveness curves and their corresponding AUC values

With $p = \mathrm{P}(D = 1) = \int_0^1 R(q)dq$=0.5



A proposal for AUC C/D for survival outcome

- Set $R(t;q) = \mathrm{P}(D(t) = 1 | X = G^{-1}(q)) = F(t | X = G^{-1}(q))$ the time-dependent predictiveness curve

- We established that

$$\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t) = \frac{\int_0^1 cR(t;c)dc - \frac{F(t)^2}{2}}{F(t)[1 - F(t)]}, \tag{3}$$

Proper estimation of $\mathbf{AUC}^{\mathbb{C},\mathbb{D}}(t)$ requires proper estimation of $R(t;c)$

A new estimator for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$

- Assume we are given an estimator $\widehat{F}_n(t_0; x)$ of the conditional absolute risk $F(t_0; x)$
- Recall that G and g denote the cumulative distribution function and the density function of X.
- Since $\int_0^1 qR(t_0; q)dq = \int_{-\infty}^{\infty} G(x)F(t_0; x)g(x)dx$,

the empirical counterpart of the quantity $\int_0^1 qR(t_0; q)dq$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{i}{n}\widehat{F}_n(t_0; X_{(i)}),$$

where $X_{(i)}$ denotes the i-th order statistic attached to the sample $X_1, ..., X_n$.

A new estimator for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$

- The marginal absolute risk function F, can be directly estimated using Kaplan-Meier estimator $\widehat{F}_{n,(1)}(t_0)$.
- Observing that $F(t_0) = \int F(t_0; x)g(x)dx$, an alternative to $\widehat{F}_{n,(1)}(t)$ relying on the conditional risk estimate is

$$\widehat{F}_{n,(2)}(t_0) = \frac{1}{n}\sum_{i=1}^{n} \widehat{F}_n(t_0; X_i).$$

This yields two estimators for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t_0)$, namely, for $k = 1, 2$,

$$\mathrm{AUC}^{\mathbb{C},\mathbb{D}}_{n,(k)}(t_0) = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{i}{n}\widehat{F}_n(t_0; X_{(i)}) - \widehat{F}^2_{n,(k)}(t_0)/2}{\widehat{F}_{n,(k)}(t_0)\left[1 - \widehat{F}_{n,(k)}(t_0)\right]}. \qquad (4)$$

Experimental results (not shown) suggested better performances results obtained with $k = 2$.

Existing estimators for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$: HLP

Heagerty Lumley and Pepe (Bcs, 2000) developed a nonparametric estimator for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$ based on the nearest-neighbor bivariate distribution estimator of Akritas (1994).

- Rewriting sensitivity
  $\mathrm{P}(X > c|D(t) = 1) = F(t|X > c)\mathrm{P}(X > c)/F(t)$
- Rewriting specificity
  $\mathrm{P}(X \leq c|D(t) = 0) = S(t|X \leq c)\mathrm{P}(X \leq c)/\{1 - F(t)\}$

Naive plugin estimators of sensitivity and specifity for S may not be monotone in c.

Estimators for $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$: HLP

Proper estimates express sensitivity and specificity as functions of the bivariate survival function $S(c,t) = \mathrm{P}(X > c, T > t)$, that is

$$\mathrm{P}(X > c | D(t) = 1) = \frac{1 - G(c) - S(c,t)}{F(t)}$$

and

$$\mathrm{P}(X \leq c | D(t) = 0) = 1 - \frac{S(c,t)}{1 - F(t)}$$

An use Equation (1) with simple numerical integration: `survivalROC` package

Existing estimators for $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$: Chambless-Diao

- They suggested a recursive calculation over the ordered times of events for $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$.

- Given two random individuals i and j, $\text{AUC}^{\mathbb{C},\mathbb{D}}(t) = \text{P}(X_i > X_j | D_i(t) = 1, D_j(t) = 0)$, with

$$D_i(t) = 1\{T_i \leq t\}$$

.

- Applying Bayes' theorem leads to

$$\text{AUC}^{\mathbb{C},\mathbb{D}}(t) = \frac{\text{P}(X_i > X_j, D_i(t) = 1, D_j(t) = 0)}{\text{P}(D_i(t) = 1)\text{P}(D_j(t) = 0)}$$

We refer to this method as CD1: `SAS`

Existing estimators for $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$: Chambless-Diao

From the Work Around Equation above, the authors observe that

$$\text{AUC}^{\mathbb{C},\mathbb{D}}(t_0) = \frac{\text{E}\left[\{1 - S(t;U)\}S(t;V)I(V < U)\right]}{\text{E}\{1 - S(t;X)\}\text{E}\{S(t;X)\}},$$

where U and V are independent observations of X.

- They Suggest to estimate the conditional survival functions under a Cox model
- The bivariate expectation is estimated as the mean over all (U, V) pairs of distinct observations.

We refer to this method as CD2: `SAS` and `R`

Simulation Study

- Compare our estimators of $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$ with those proposed in the literature
- Assess the effect of a misspecified model – when estimating the conditional absolute risk– on the $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$ estimation.
-

$$\begin{aligned}
\lambda_1(t|X) &= \frac{\exp(\beta X)}{1+t} \\
\lambda_2(t|X) &= t\exp\left(\frac{\beta X t^2}{2}\right) \\
\lambda_3(t|X) &= \beta_0 t + \frac{\beta}{t+1}X,
\end{aligned}$$

evaluation times: the first quartile t_{q1}, the median t_{q2} and third quartile t_{q3} of the survival time distribution.

Simulations: Censoring schemes

- We applied an "administrative censoring" occurring at the time corresponding to the 80% percentile of the survival time distribution.
- (i) no additional censoring,
- (ii) $C_i \sim \mathcal{E}(\tau_1)$
- (iii) $C_i \sim \mathcal{E}(\tau_2)$,

where rates τ_1 and τ_2 of the exponential distribution $\mathcal{E}(\cdot)$ were respectively chosen so that censoring rate attained 25% and 75% respectively.

Mean Bias

Table: Results of the simulation study. Comparisons between several estimators of $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$. Averaged bias (multiplied by 100) obtained from 100 runs are reported.

					100× **Bias**			
Eval. Time	CD2 Cox	VL Cox	CD2 Aalen	VL Aalen	HLP NNE	CD1	CD2 KM	VL KM
	Standard Cox model							
	Censoring scheme 1							
t_{q1}	-0.302	-0.168	-0.495	-0.361	-1.033	0.131	-1.185	-1.052
t_{q2}	-0.284	-0.082	0.107	0.310	-1.377	-0.239	-1.463	-1.262
t_{q3}	-0.301	0.103	1.083	1.485	-1.457	-0.598	-1.822	-1.413
	Censoring scheme 2							
t_{q1}	-0.016	0.117	-0.422	-0.288	-1.191	0.031	-1.244	-1.111
t_{q2}	-0.031	0.170	0.220	0.423	-1.304	-0.159	-1.316	-1.115
t_{q3}	0.009	0.415	1.728	2.132	-0.853	-0.280	-1.185	-0.774

Mean Bias

Table: Results of the simulation study. Comparisons between several estimators of $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$. Averaged bias (multiplied by 100) obtained from 100 runs are reported.

					100× **Bias**			
Eval. Time	CD2 Cox	VL Cox	CD2 Aalen	VL Aalen	HLP NNE	CD1	CD2 KM	VL KM
	Time-varying Cox model							
	Censoring scheme 1							
t_{q1}	6.775	6.906	2.748	2.882	-1.783	0.163	-0.864	-0.731
t_{q2}	-2.303	-2.107	6.002	6.199	-2.333	0.274	-0.756	-0.556
t_{q3}	-9.046	-8.629	7.012	7.377	-1.419	-0.047	-0.721	-0.317
	Censoring scheme 2							
t_{q1}	5.796	5.927	2.395	2.528	-2.457	-0.229	-1.329	-1.196
t_{q2}	-3.200	-3.004	5.670	5.867	-2.828	0.071	-1.080	-0.881
t_{q3}	-9.948	-9.535	7.176	7.536	-1.343	0.492	-0.456	-0.057

Mean Bias

Table: Results of the simulation study. Comparisons between several estimators of $\text{AUC}^{\mathbb{C},\mathbb{D}}(t)$. Averaged bias (multiplied by 100) obtained from 100 runs are reported.

					100× **Bias**			
Eval. Time	CD2 Cox	VL Cox	CD2 Aalen	VL Aalen	HLP NNE	CD1	CD2 KM	VL KM
	Aalen additive model							
	Censoring scheme 1							
t_{q1}	-7.807	-7.674	0.470	0.603	-1.432	0.496	-0.686	-0.554
t_{q2}	-5.157	-4.955	0.047	0.248	-1.861	-0.015	-0.980	-0.779
t_{q3}	-2.186	-1.778	0.221	0.621	-1.324	0.294	-0.500	-0.099
	Censoring scheme 2							
t_{q1}	-7.757	-7.624	-0.337	-0.204	-2.247	-0.416	-1.553	-1.420
t_{q2}	-5.099	-4.898	-0.269	-0.070	-1.638	-0.199	-0.917	-0.718
t_{q3}	-2.109	-1.703	-0.420	-0.022	-1.791	-1.342	-0.994	-0.593

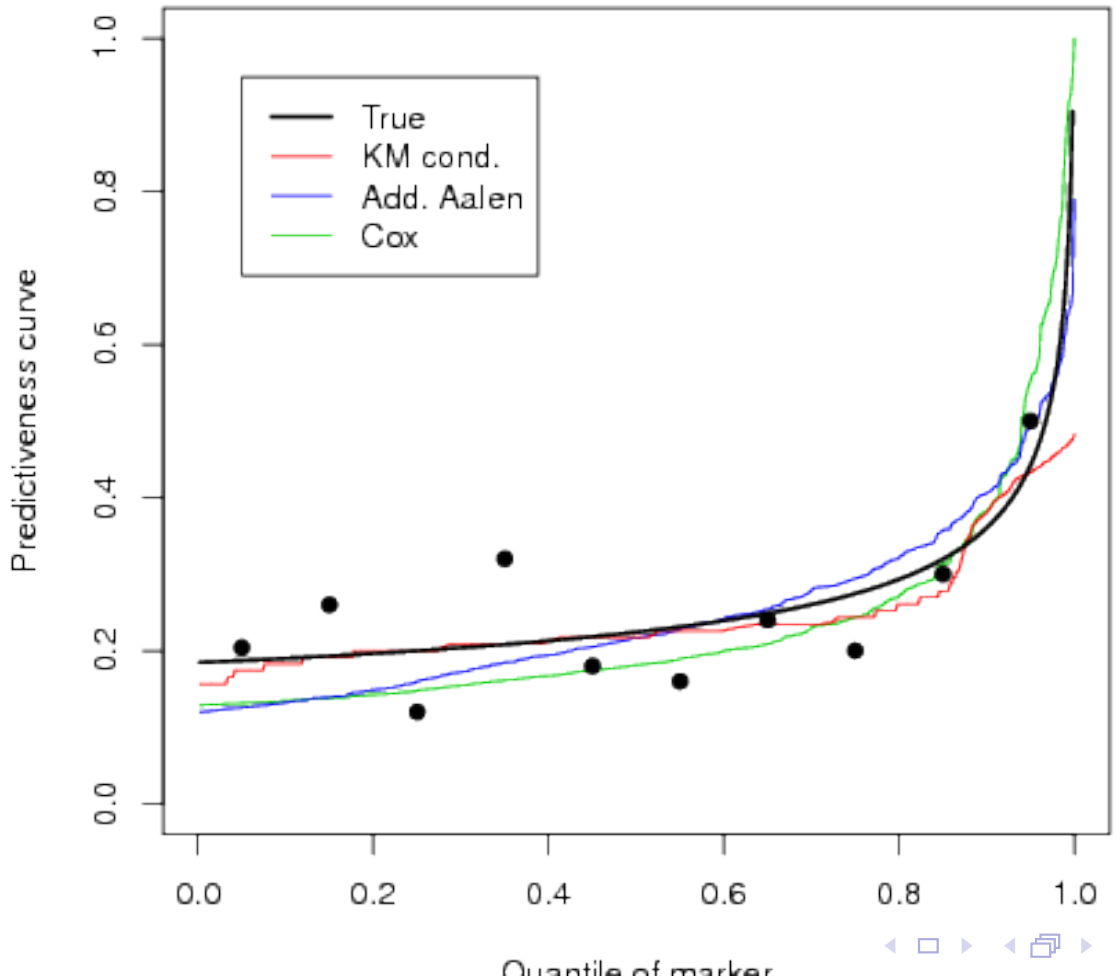
Assessing the accuracy of $AUC^{\mathbb{C},\mathbb{D}}$ estimates using predictiveness curves

Effect of a misspecified model – when estimating the conditional absolute risk– on the $AUC^{\mathbb{C},\mathbb{D}}(t)$ estimation

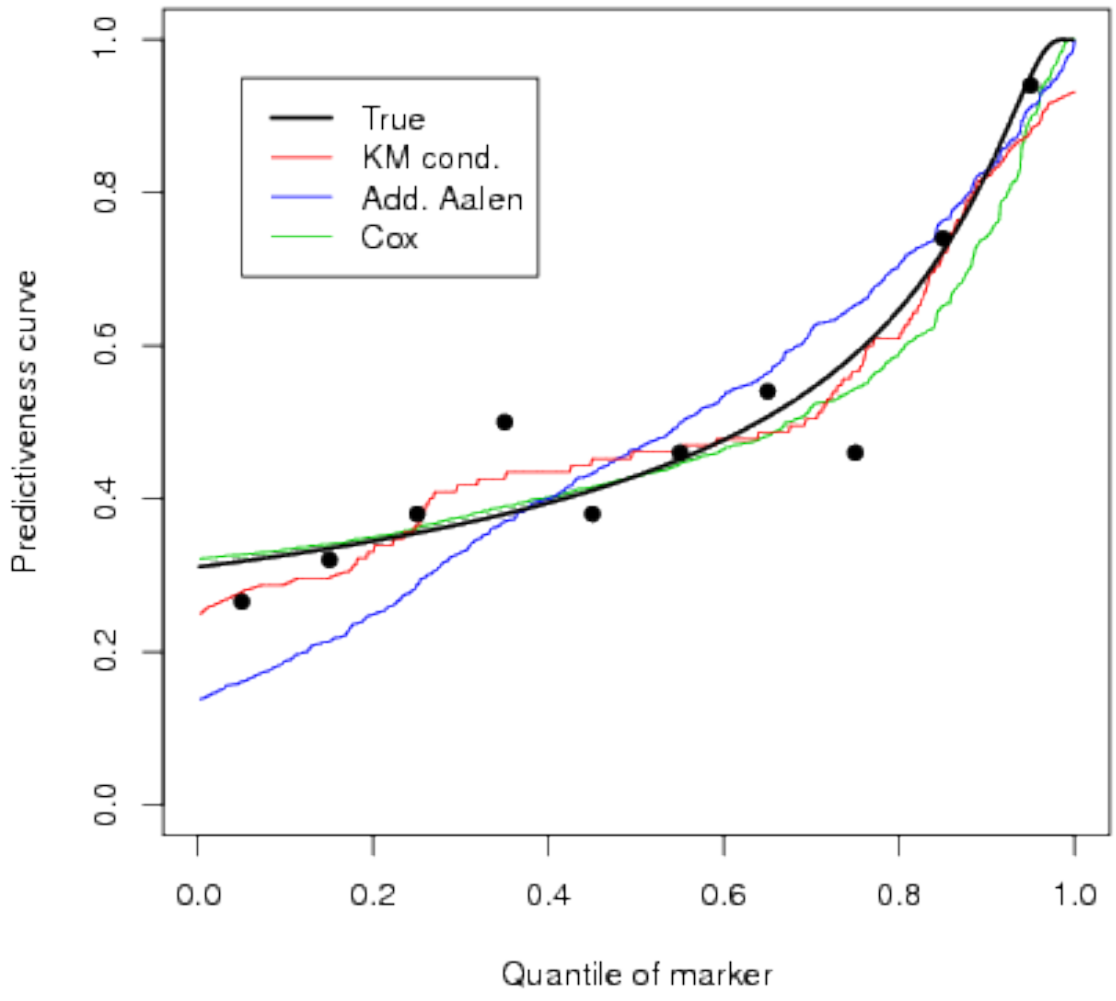
- Accurate estimates of $R(t_0; q)$ should yield accurate estimates for $AUC^{\mathbb{C},\mathbb{D}}(t_0)$.
- Two evaluation times were considered: the first quartile t_{q1} and the median t_{q2} of the survival time distribution.
- Black bullets represent KM estimators of the unconditional absolute risk for each decile of predicted risk

PC Cox time-varying effect; 1st quartile

PC is underestimated on the quantiles interval $[0, 0.85]$ and slightly overestimated on the interval [0.85,1]



PC Cox time-varying effect ; median



$AUC(t)$: Time Varying Cox model

$\text{AUC}^{\mathbb{C},\mathbb{D}}(t_1)$ is largely overestimated with Cox at first quartile

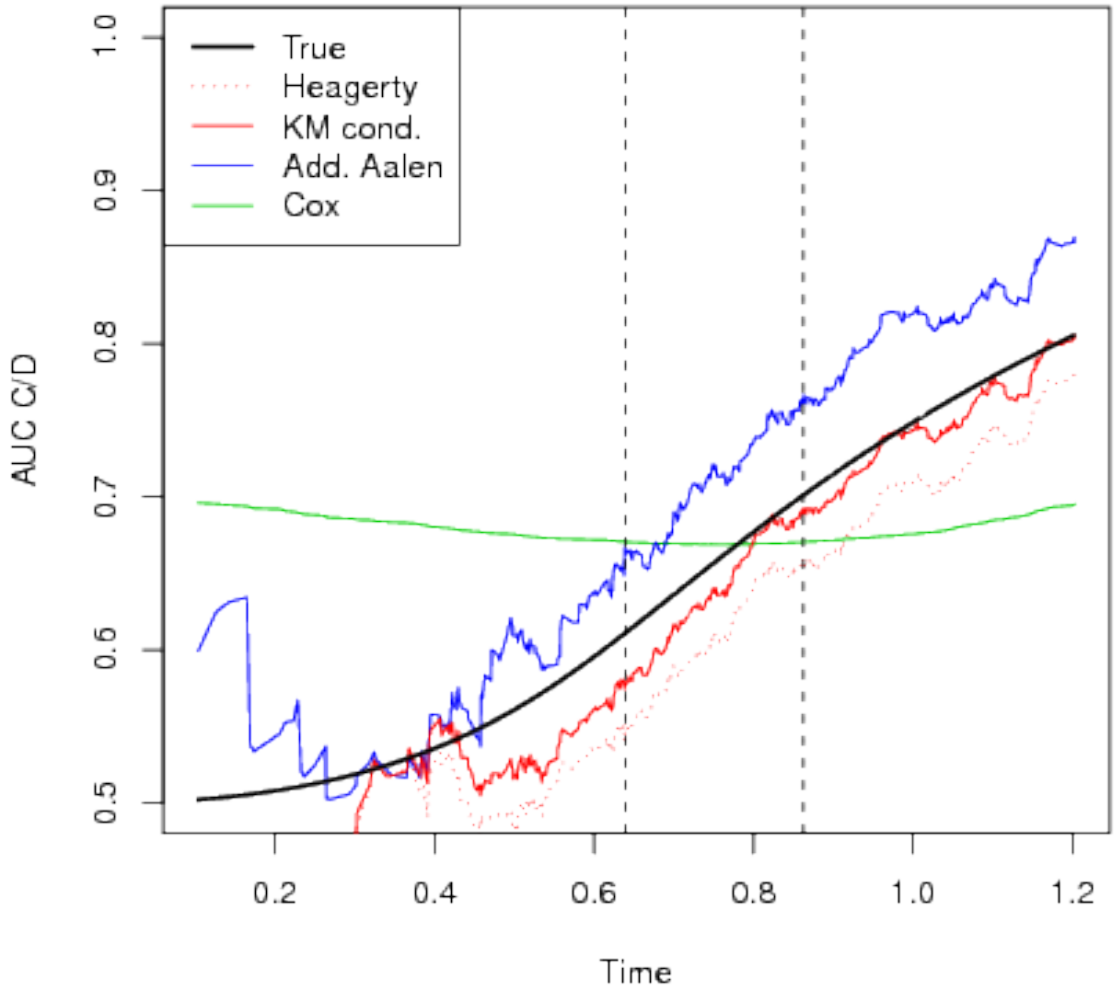
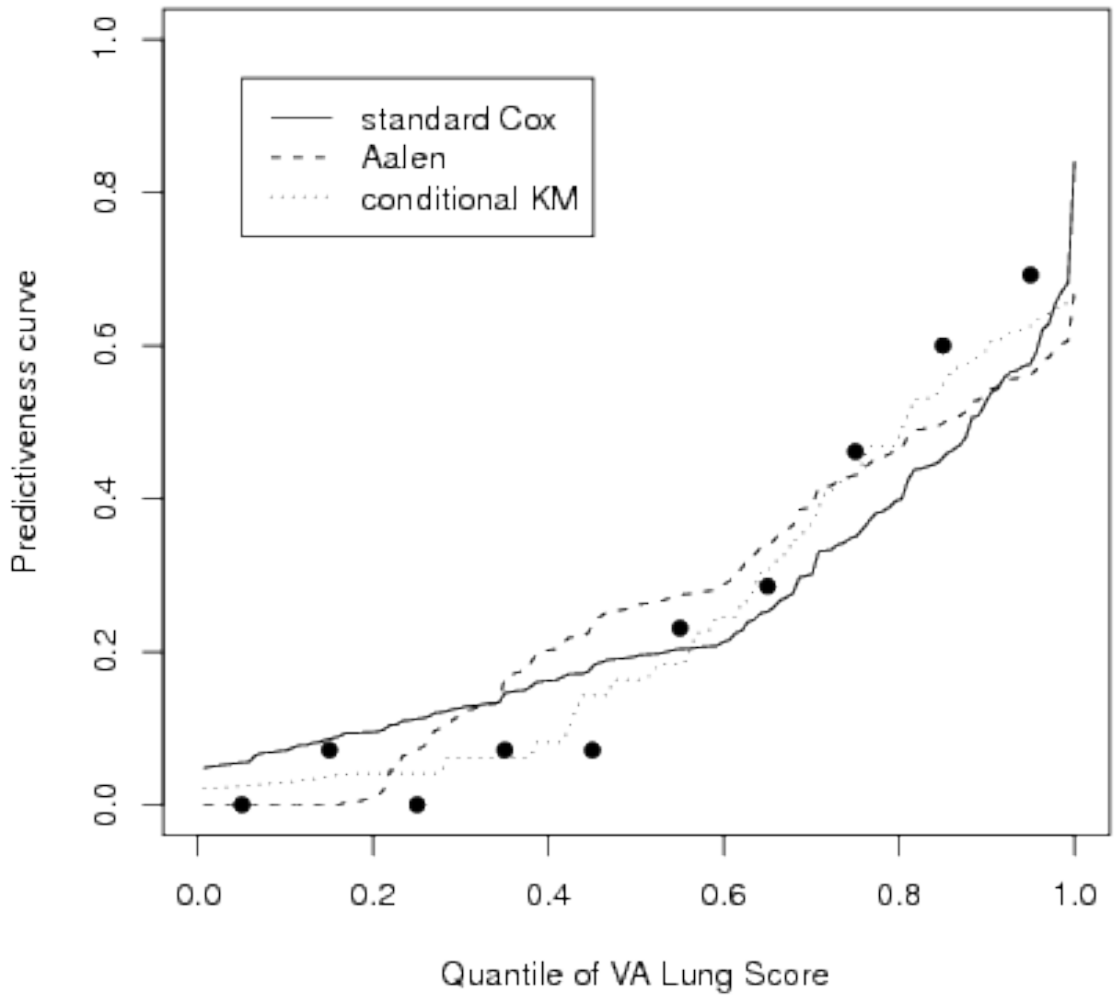


Illustration: VA Lung

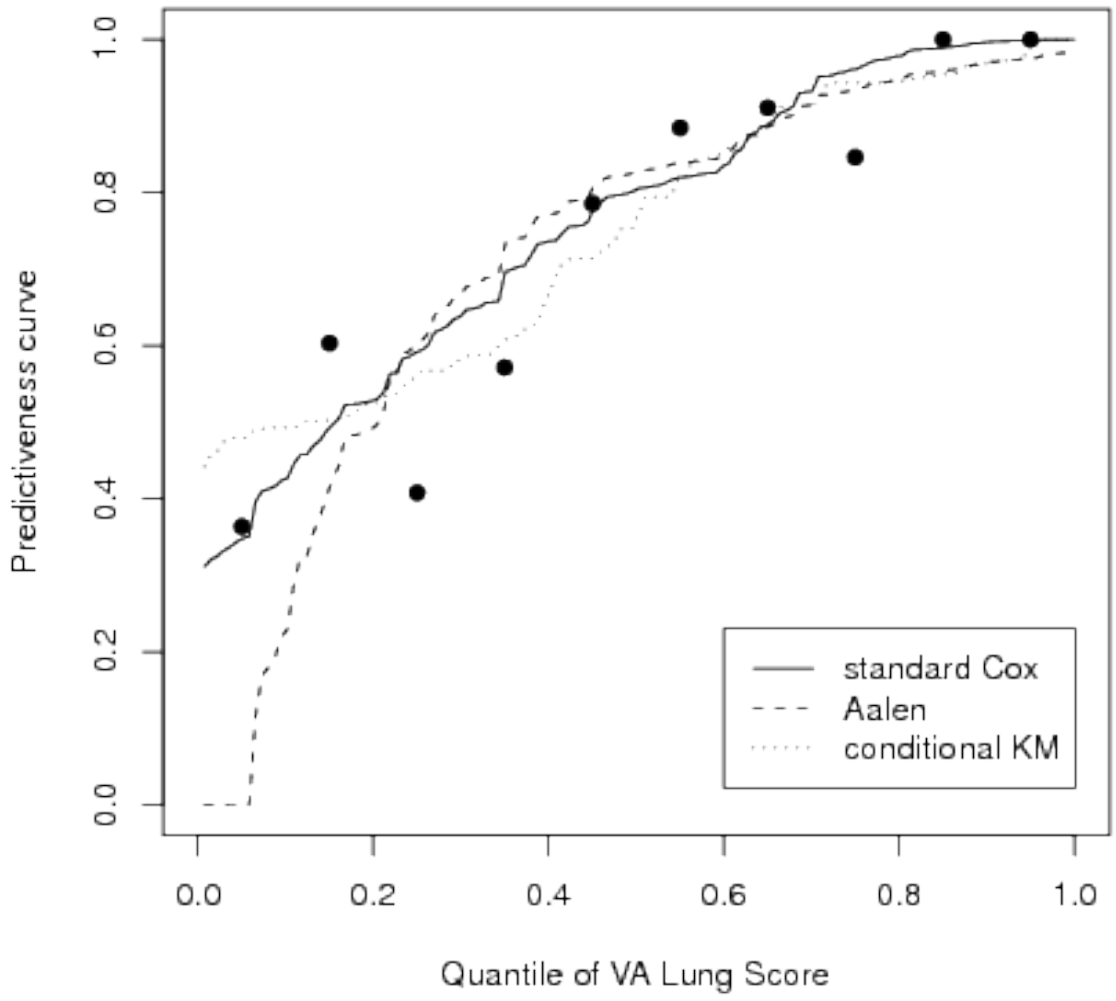
- Overall, 137 males with inoperable cancer were randomized to a standard or a test chemotherapy.
- Death was considered as the endpoint, and more than 93% of the participants died during the study.
- Predictors of mortality include type of treatment, age, histological type of tumor and the Karnofsky score (which is a performance status measure).
- We considered a 500-day follow-up and a Cox model was used to build a risk score out of these baseline covariates.
- Our objective: estimate the $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$ attached to this score.

- we computed estimates of $\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$ with HLP and ours

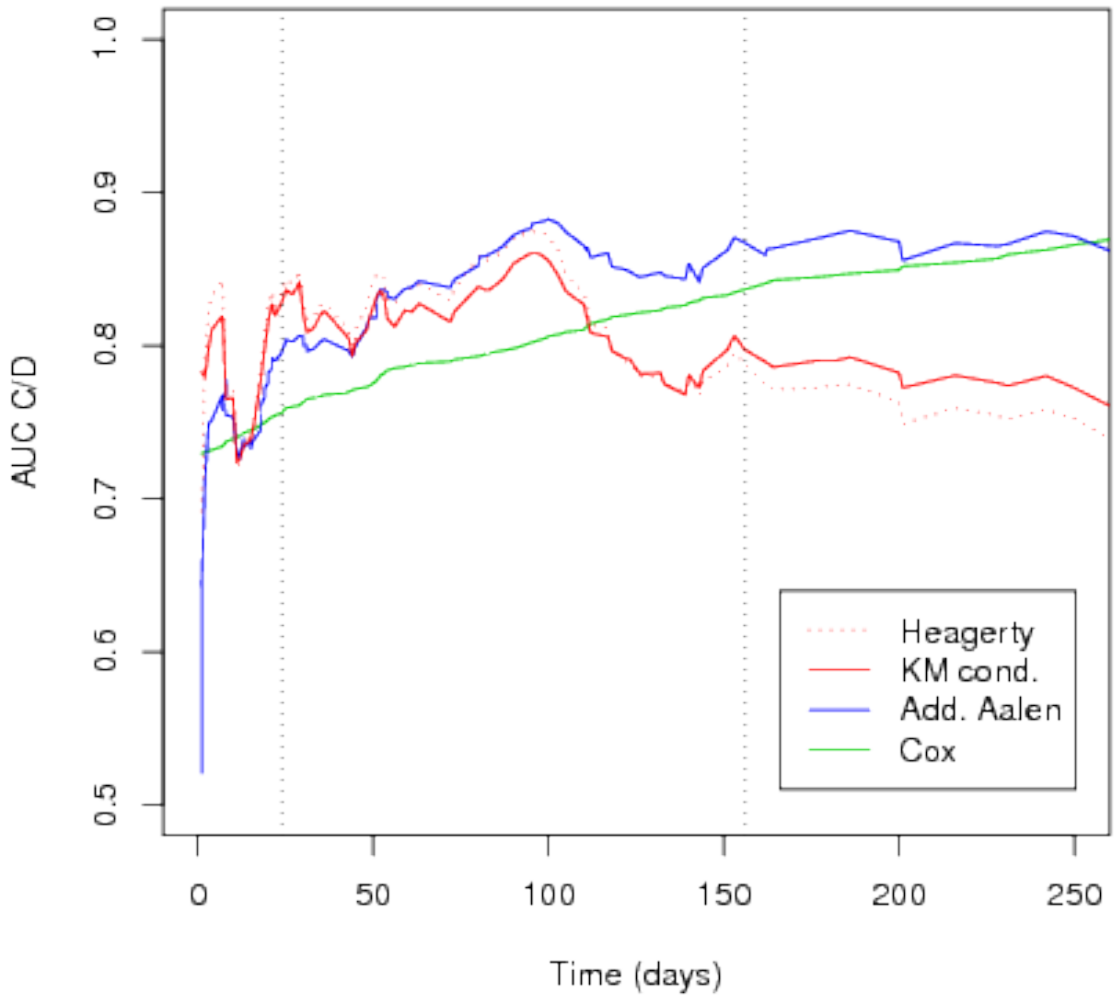
Predictiveness Curve VA Lung 1st Quartile



Predictiveness Curve VA Lung 3rd Quartile



$\mathrm{AUC}^{\mathbb{C},\mathbb{D}}(t)$ VA Lung



Conclusion

- Our approach relies on the additional estimation of the cumulative distribution of X which might increase variability.
- The nonparametric estimator of Chambless-Diao was observed to slightly outperform its three nonparametric competitors (including our approach) in most of our empirical examples
- Except for high censoring rates and late evaluation times; where our approach appeared to perform the best
- Conditional risk function, through the predictiveness curve, is the key when assessing discrimination of prognostic tools

Readings

- Pepe, M.S. *et al*. Integrating the Predictiveness of a Marker with Its Performance as a Classifier. American Journal of Epidemiology. 2007
- Heagerty, P.J. and Zheng, Y. Survival Model Predictive Accuracy and ROC Curve. Biometrics. 2005; 61, 92-105.
- Heagerty, Lumley and Pepe.Time–dependent ROC curves for censored Survival Data and a Diagnostic Marker. Biometrics; 2000 56, 337–344
- Viallon, V and Latouche, A. ; Discrimination Measures for survival outcome: connection between the AUC and the predictiveness curve. Biometrical Journal. 2011; 53(2):217-36
- See also `survAUC` implements various estimators