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 developed 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
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
For a continuous (bio)marker $X$ and a binary outcome $D$

ROC plots sensitivity, $P(X > c | D = 1)$, against 1 minus specificity, $1 - P(X \leq c | D = 0)$, 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
In prospective cohort study, a binary outcome can change over time e.g. a disease status ⇒ 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, $AUC(t)$. 
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 $T_i > t^*$ for a fixed $t^*$ 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 $t'$
Let $T_i$ and $C_i$ denote survival and censoring times for subject $i$.

We observe $(Z_i, \delta_i)$ 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:

- $\text{TPR}(c, t) = P(X > c | D(t) = 1)$
- $\text{FPR}(c, t) = P(X > c | D(t) = 0)$
The time-dependent ROC curve $\text{ROC}(t)$ plots

- $\text{TPR}(c, t)$ vs

- $\text{FPR}(c, t)$ for any threshold $c$

so that

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

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

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

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

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

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

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

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 $R(q) = P[D = 1|X = G^{-1}(q)]$ be the risk associated to the $q$th quantile of marker $X$
- The predictiveness curve plots $R(q)$ versus $q$
let \( R(q) = P[D = 1|X = G^{-1}(q)] \) 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 \( P(D = 1|X) \). We established that

\[
AUC = \frac{\int_0^1 qR(q) dq - p^2/2}{p(1-p)},
\]

where \( p = P(D = 1) = \int_0^1 R(q) dq \).
Predictiveness curves and their corresponding AUC values

With \( p = P(D = 1) = \int_0^1 R(q) dq = 0.5 \)
A proposal for AUC C/D for survival outcome

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

- We established that

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

Proper estimation of $AUC_{C,D}^{C,D}(t)$ requires proper estimation of $R(t; c)$
A new estimator for AUC\(_{C,D}(t)\)

- Assume we are given an estimator \( \hat{F}_n(t_0; x) \) 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

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{i}{n} \hat{F}_n(t_0; X(i)),
\]

where \( X(i) \) denotes the \( i \)-th order statistic attached to the sample \( X_1, \ldots, X_n \).
A new estimator for $\text{AUC}^{C,D}(t)$

- The marginal absolute risk function $F$, can be directly estimated using Kaplan-Meier estimator $\hat{F}_{n,(1)}(t_0)$.

- Observing that $F(t_0) = \int F(t_0; x)g(x)dx$, an alternative to $\hat{F}_{n,(1)}(t)$ relying on the conditional risk estimate is

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

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

$$\text{AUC}_{n,(k)}^{C,D}(t_0) = \frac{\frac{1}{n} \sum_{i=1}^{n} \frac{i}{n} \hat{F}_n(t_0; X_{(i)}) - \hat{F}_{n,(k)}^2(t_0)/2}{\hat{F}_{n,(k)}(t_0) [1 - \hat{F}_{n,(k)}(t_0)]}. \quad (4)$$

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

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

- Rewriting sensitivity
  \[ P(X > c | D(t) = 1) = F(t | X > c)P(X > c)/F(t) \]

- Rewriting specificity
  \[ P(X \leq c | D(t) = 0) = S(t | X \leq c)P(X \leq c)/(1 - F(t)) \]

Naive plugin estimators of sensitivity and specificity for $S$ may not be monotone in $c$. 
Proper estimates express sensitivity and specificity as functions of the bivariate survival function $S(c, t) = P(X > c, T > t)$, that is

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

and

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

An use Equation (1) with simple numerical integration: survivalROC package
They suggested a recursive calculation over the ordered times of events for $\text{AUC}^{C,D}(t)$. 

Given two random individuals $i$ and $j$, 

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

with 

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

Applying Bayes’ theorem leads to 

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

We refer to this method as CD1: SAS
From the Work Around Equation above, the authors observe that

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

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 $\text{AUC}^{C,D}(t)$ with those proposed in the literature.

- Assess the effect of a misspecified model – when estimating the conditional absolute risk – on the $\text{AUC}^{C,D}(t)$ estimation.

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

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 $\tau_1$ and $\tau_2$ 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 \( \text{AUC}^{C,D}(t) \). Averaged bias (multiplied by 100) obtained from 100 runs are reported.

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

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

<table>
<thead>
<tr>
<th>Eval. Time</th>
<th>CD2 Cox</th>
<th>VL Cox</th>
<th>CD2 Aalen</th>
<th>VL Aalen</th>
<th>HLP NNE</th>
<th>CD1 KM</th>
<th>CD2 KM</th>
<th>VL KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{q1}$</td>
<td>6.775</td>
<td>6.906</td>
<td>2.748</td>
<td>2.882</td>
<td>-1.783</td>
<td>0.163</td>
<td>-0.864</td>
<td>-0.731</td>
</tr>
<tr>
<td>$t_{q2}$</td>
<td>-2.303</td>
<td>-2.107</td>
<td>6.002</td>
<td>6.199</td>
<td>-2.333</td>
<td>0.274</td>
<td>-0.756</td>
<td>-0.556</td>
</tr>
<tr>
<td>$t_{q3}$</td>
<td>-9.046</td>
<td>-8.629</td>
<td>7.012</td>
<td>7.377</td>
<td>-1.419</td>
<td>-0.047</td>
<td>-0.721</td>
<td>-0.317</td>
</tr>
</tbody>
</table>

**Time-varying Cox model**

*Censoring scheme 1*

| $t_{q1}$ | 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 |

*Censoring scheme 2*
### Mean Bias

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

<table>
<thead>
<tr>
<th>Eval. Time</th>
<th>CD2 Cox</th>
<th>VL Cox</th>
<th>CD2 Aalen</th>
<th>VL Aalen</th>
<th>HLP NNE</th>
<th>CD1 KM</th>
<th>CD2 KM</th>
<th>VL KM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aalen additive model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Censoring scheme 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_q1$</td>
<td>-7.807</td>
<td>-7.674</td>
<td>0.470</td>
<td>0.603</td>
<td>-1.432</td>
<td>0.496</td>
<td>-0.686</td>
<td>-0.554</td>
</tr>
<tr>
<td>$t_q2$</td>
<td>-5.157</td>
<td>-4.955</td>
<td>0.047</td>
<td>0.248</td>
<td>-1.861</td>
<td>-0.015</td>
<td>-0.980</td>
<td>-0.779</td>
</tr>
<tr>
<td>$t_q3$</td>
<td>-2.186</td>
<td>-1.778</td>
<td>0.221</td>
<td>0.621</td>
<td>-1.324</td>
<td>0.294</td>
<td>-0.500</td>
<td>-0.099</td>
</tr>
<tr>
<td><strong>Censoring scheme 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_q1$</td>
<td>-7.757</td>
<td>-7.624</td>
<td>-0.337</td>
<td>-0.204</td>
<td>-2.247</td>
<td>-0.416</td>
<td>-1.553</td>
<td>-1.420</td>
</tr>
<tr>
<td>$t_q2$</td>
<td>-5.099</td>
<td>-4.898</td>
<td>-0.269</td>
<td>-0.070</td>
<td>-1.638</td>
<td>-0.199</td>
<td>-0.917</td>
<td>-0.718</td>
</tr>
<tr>
<td>$t_q3$</td>
<td>-2.109</td>
<td>-1.703</td>
<td>-0.420</td>
<td>-0.022</td>
<td>-1.791</td>
<td>-1.342</td>
<td>-0.994</td>
<td>-0.593</td>
</tr>
</tbody>
</table>
Assessing the accuracy of $\text{AUC}^{C,D}$ estimates using predictiveness curves

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

- Accurate estimates of $R(t_0; q)$ should yield accurate estimates for $\text{AUC}^{C,D}(t_0)$.

- 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**

\[ AUC^{C, D}(t_1) \] is largely overestimated with Cox at first quartile
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 $\text{AUC}_{C,D}^{C,D}(t)$ attached to this score.

we computed estimates of $\text{AUC}_{C,D}^{C,D}(t)$ with HLP and ours
Predictiveness Curve VA Lung 1st Quartile
Predictiveness Curve VA Lung 3rd Quartile
AUC^{C,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


- See also `survAUC` implements various estimators