Time-dependent ROC analysis for censored event-time data: Review

Adina Najwa Kamarudin
Dr Ruwanthi Kolamunnage-Dona
Dr Trevor Cox
Outline

INTRODUCTION
- Time-dependent vs Standard ROC
- Sensitivity and Specificity
- ROC curve and AUC

LITERATURE REVIEW
- Results from MEDLINE and Internet
- Definition of Sensitivity and Specificity
- Estimation Methods

APPLICATION
- Simulation Study
- Mayo PBC data
Standard vs time-dependent ROC

Disease status fixed over time

\[ D=0 \]

\[ D(t)=0 \]

Disease status change over time

\[ D=0 \]

\[ D(t)=0 \]

\[ D(t)=1 \]

- Sensitivity is a probability that an individual having a disease will be correctly identified by a diagnostic test (positive).
- Specificity is a probability that an individual not having a disease will be correctly identified by a diagnostic test (negative).
- Time-dependent ROC takes account the time of the disease. Thus, it is more appropriate.
Literature Review: Results

MEDLINE search (300 papers)
Keywords: ROC, AUC, event-time, time-dependent censoring

Included (19 papers)

Excluded (281 papers)
5 papers are from the Internet

24 papers

12 papers; single baseline biomarker

Kaplan-Meier
NNE
IPCW/CIPC
W
Cox Regression
Weighted Mean Rank (WMR)

12 papers; longitudinal, competing risk

24 papers
Notation

- $T_i$: time-to-event (disease, death) for individual
- $x_i$: marker value for individual
- $D_i$: disease status for individual
- $D_i(t)$: disease status at time $t$ for individual

Standard Accuracy

<table>
<thead>
<tr>
<th>Standard Accuracy Measures</th>
<th>Time-dependent Accuracy Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Sensitivity}(c) = P(x_i &gt; c</td>
<td>D_i = 1)$</td>
</tr>
<tr>
<td>$\text{Specificity}(c) = P(x_i \leq c</td>
<td>D_i = 0)$</td>
</tr>
</tbody>
</table>

There are 3 different definitions of sensitivity and specificity.
Each individual can play the role of a control for an early time but then contributes as a case for later times, 
\[ \text{sensitivity}^I(c, t) = P(X_i > c | T_i = t) \]

An individual can play the role of a control for an early time but then play the role of case when

Each individual does not change disease status and is treated as either a case or a control. Cases are stratified according to the time at which the event occurs (incident) and controls are those individuals who are event free through a fixed follow-up period (0, )
Estimation Methods

Cumulative/Dynamic

- **KM**<sub>HLP</sub> (survivalR OC)
- **KM**<sub>CD</sub> (SAS macro)
- **NNE** (survivalR OC)
- **IPCW** (timeROC)
- **CIPCW** (timeROC)

Incident/Dynamic

- Cox regression
- Weighted Mean Rank (WMR)

Incident/Static

- Mixed-effect regression models
- Mixed-effect segmented regression process
Kaplan-Meier Methodology

- Widely used nonparametric estimator of survival function.
- Uses all the information in the data including the censored observations.

\[ \hat{S}(c,t) = \frac{1 - \hat{S}(t \mid X > c) (1 - F_X(c))}{1 - \hat{S}(t)} \]

\[ \hat{S}_p(c,t) = 0 = \frac{\hat{S}(t \mid X \leq c) F_X(c)}{\hat{S}(t)} \]

\[ \bar{S}e(c,t) = \sum_{k=1}^{m} I(X_{d(k)} > c) (S(t_{k-1}) - S(t_k)) \]

\[ \bar{S}_p(c,t) = \frac{F_X(c) \sum_{k=1}^{m} I(X_{d(k)} \leq c) (S(t_{k-1}) - S(t_k))}{\hat{S}(t_m)} \]
An improved methodology for $\text{KM}_{\text{HLP}}$ is suggested by Heagerty et al., (2000) using the NNE of the bivariate distribution of $(X,T)$.

Both sensitivity and specificity are monotone and bounded in $[0,1]$.

IPCW & CIPCW

- IPCW corrects standard estimate of the sensitivity by adding weights to the observations in the subsample of uncensored individual before time $t$ by their probability of being uncensored.
- The specificity is remain the same.
- CIPCW is proposed by weighting the observations of uncensored individuals at time $t$ by the conditional probability of being uncensored given the marker, instead of weighting by the marginal probability of being uncensored.
Simulation Study

- The data (Marker, log(time)) are simulated using the bivariate normal distribution for n=200 individuals.
- Several scenarios were generated by changing the correlation, \( r \) between marker and log(time), to -0.5, -0.9.
- Several scenarios were generated by changing the percentage of censoring, \( c \), to 40%.
- 500 replications.
Simulation Study: Results (AUC(SD, %Bias))

<table>
<thead>
<tr>
<th>Log(T)</th>
<th>True</th>
<th>Standard</th>
<th>KM$_{HLP}$</th>
<th>NNE</th>
<th>IPCW</th>
<th>CIPCW</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.5</td>
<td>0.735</td>
<td>0.791(0.073, 8)</td>
<td>0.790(0.076, 8)</td>
<td>0.775(0.073, 5)</td>
<td>0.789(0.074, 7)</td>
<td>0.794(0.067, 8)</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.691</td>
<td>0.758(0.053,10)</td>
<td>0.756(0.056,10)</td>
<td>0.743(0.053,8)</td>
<td>0.755(0.054,9)</td>
<td>0.762(0.049,10)</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.653</td>
<td>0.738(0.047,13)</td>
<td>0.735(0.049,12)</td>
<td>0.735(0.049,13)</td>
<td>0.734(0.047,12)</td>
<td>0.743(0.051,13)</td>
</tr>
<tr>
<td>0</td>
<td>0.623</td>
<td>0.733(0.044,17)</td>
<td>0.728(0.046,17)</td>
<td>0.728(0.046,17)</td>
<td>0.728(0.045,17)</td>
<td>0.735(0.047,18)</td>
</tr>
<tr>
<td>-1.5</td>
<td>0.935</td>
<td>0.969(0.018,4)</td>
<td>0.973(0.035,4)</td>
<td>0.948(0.017,1)</td>
<td>0.969(0.018,4)</td>
<td>0.969(0.018,4)</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.896</td>
<td>0.955(0.035,7)</td>
<td>0.958(0.035,7)</td>
<td>0.942(0.019,5)</td>
<td>0.954(0.019,7)</td>
<td>0.955(0.019,7)</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.852</td>
<td>0.946(0.018,11)</td>
<td>0.946(0.032,11)</td>
<td>0.935(0.018,10)</td>
<td>0.943(0.018,11)</td>
<td>0.945(0.020,11)</td>
</tr>
<tr>
<td>0</td>
<td>0.809</td>
<td>0.944(0.019,1)</td>
<td>0.942(0.031,1)</td>
<td>0.930(0.019,1)</td>
<td>0.940(0.020,1)</td>
<td>0.942(0.022,1)</td>
</tr>
</tbody>
</table>
Mayo PBC dataset

• The data are from Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver disease from 1974 to 1984 which contained 312 randomized patients.
• The marker is a prognostic score containing five covariates at baseline: log (bilirubin), albumin, log (prothrombin time), edema and age.
• Clinical question is how well the prognostic score can distinguish between individual who dead and individual who censored.
### Mayo PBC dataset: Results

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC(SD)</th>
<th>Year 1</th>
<th>Year 5</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.846(0.023)</td>
<td>0.885(0.021)</td>
<td>0.883(0.030)</td>
</tr>
<tr>
<td>Standard</td>
<td></td>
<td>0.918(0.043)</td>
<td>0.918(0.022)</td>
<td>0.868(0.032)</td>
</tr>
<tr>
<td>KM_HLP</td>
<td></td>
<td>0.886(0.051)</td>
<td>0.888(0.025)</td>
<td>0.879(0.034)</td>
</tr>
<tr>
<td>NNE</td>
<td></td>
<td>0.918(0.043)</td>
<td>0.915(0.021)</td>
<td>0.858(0.036)</td>
</tr>
<tr>
<td>IPCW</td>
<td></td>
<td>0.918(0.045)</td>
<td>0.917(0.021)</td>
<td>0.860(0.034)</td>
</tr>
<tr>
<td>CIPCW</td>
<td></td>
<td>0.886(0.051)</td>
<td>0.888(0.025)</td>
<td>0.879(0.034)</td>
</tr>
</tbody>
</table>

![ROC curves for Year 1, Year 5, and Year 10](image)
Conclusion

• As the time increases, the AUC is decreases indicating that performance of the biomarker is decreasing with time.
• The simulation studies have shown that the percentage of bias for standard ROC is slightly higher compared to most of the time-dependent AUC.
• As the correlation between the marker and the time is increases, the AUC values increase.
Future Work

• Continue with the estimation methods from other definition of sensitivity and specificity (Incident/Dynamic and Incident/Static).
• Look into those estimation methods which using the longitudinal biomarker and consider the competing risk.
• Investigate other possible error contributes to ROC analysis performance such as measurement error.
References


Thank you for your attention!