Introduction

Cancer patients and doctors need to make decisions about treatments and end-of-life care.

Good prognostic models help manage the uncertainty and make better decisions.

Use statistics such as median survival time, 5 year survival rate.

Now: estimated from a large heterogenous population, based on only:

- Site + Stage of cancer
- (Perhaps) 1 or 2 covariates (e.g., age, sex, race)

- Should use other information already in electronic health records:
  - prescriptions
  - blood test results
  - performance assessment by physicians
- e.g. the Alberta Cancer Registry
- Our GOAL: Build accurate patient-specific prognostic predictors with electronic health records

Basic Survival Analysis

Survival function:

\[ S(t) = P(T \geq t) \]

is the proportion of patients surviving longer than \( t \) months

Hazard function:

\[ h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \]

is the instantaneous rate of failure/death at time \( t \), given the patient survives longer than \( t \)

Censored Observations: Some patients leave in the middle of a study or remain alive at the end of a study ⇒ only a lower bound on their survival time.

Survival Regression Models

- One of the most important regression model in survival analysis is the Cox Proportional Hazards (PH) Model [1]
- Given covariates \( X \), Cox regression assumes the hazard function have the form

\[ h(t \mid x) = h_0(t) \exp(\beta^T x) \]

where \( h_0(t) \) is a base hazard independent of \( x \).

- The hazard ratio is constant (independent of \( t \)) for two patients \( x_1 \) and \( x_2 \):

\[ h(t \mid x_1)/h(t \mid x_2) = \exp(\beta^T (x_1 - x_2)) \]

- Objective depends on rank of survival time, cannot handle covariates with time-varying effects
- Other survival regression models include Aalen linear hazards model [2], parametric survival regression based on Weibull, log-normal distributions

Survival Analysis VS Survival Prediction:

Survival Analysis

Goal: Evaluate prognostic factors and treatment effectiveness

Focus: Evaluation on populations

Uses: Testing new drugs & medical devices

Survival Prediction

Goal: Predict accurate patient-specific survival time

Focus: Evaluation on individuals

Uses: Treatment planning & patient management

Data - Cancer Patient Composition and Patient Attributes

<table>
<thead>
<tr>
<th>Site/sex</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bronchus &amp; Lung</td>
<td>44</td>
<td>32</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Colorectal</td>
<td>157</td>
<td>233</td>
<td>545</td>
<td>827</td>
</tr>
<tr>
<td>Head and Neck</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>Esophagus</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Stomach</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>128</td>
</tr>
<tr>
<td>Other Digestive</td>
<td>0</td>
<td>1</td>
<td>77</td>
<td>0</td>
</tr>
<tr>
<td>Misc</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>123</td>
</tr>
</tbody>
</table>

General wellbeing

- taste funny, constipation, pain
- dental problem, dry mouth, vomit, diarrrhea, performance status
- granulocytic, LDH, serum, HGB, albumin

Blood test

- liver function test, WBC count
- albumin, creatinine, albumin

Predicting Personalized Survival Distributions

Consider a simpler problem: will a subject with covariate \( x \) survive \( \geq t \) months?

Model this with a simple logistic regression model:

\[ P_s(T \geq t \mid x) = \frac{1}{1 + \exp(\theta^T x + b)} \]

Multiple thresholds \( t_1, t_2, \ldots \) to capture survival function at different times

\[ P_s(T \geq t_j \mid x) = \frac{1}{1 + \exp(\theta^T x + b)} \]

Predict a sequence of dependent bits

- Multiple-task logistic regression for survival prediction:

\[ \exp(\sum_{j=1}^N \theta_j^T x + b) \]

- Add multi-task regularizer to prevent overfitting:

\[ \min \left\{ \frac{1}{2} \sum_{j=1}^N \|	heta_j^T x + b\|^2 \mid \sum_{j=1}^N \|	heta_j^T x + b\|^2 \right\} \]

- Similar to CRF training objective, but
  - no transition features
  - no sharing of node potentials
  - Partition function \( Z(\hat{\theta}, \hat{x}) \) involves only a linear number of terms
- Model also related to local regression approaches [3]
- Advantage of the MTLR approach:
  - No PH assumption - effects of covariates can change with time
  - Handles censoring: by integrating out hidden variables in a survival sequence
  - More accurate survival probability predictions (below)

Experimental Results

- Classification Accuracy and Survival Rate Prediction (5CV):

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>1-month</th>
<th>2-month</th>
<th>3-month</th>
<th>4-month</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTLR</td>
<td>0.85(0.7)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
</tr>
<tr>
<td>Cox</td>
<td>0.78(0.7)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
</tr>
<tr>
<td>Aalen</td>
<td>0.74(0.7)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
<td>0.74(1.3)</td>
</tr>
</tbody>
</table>

- Optimizing clinically relevant loss functions (prediction \( p \), true survival \( t \)):
  - Absolute error on survival time (AE): \( |p - t| \)
  - Absolute error on log survival time (AE-log): \( \log |p - t| \)
  - Relative absolute error (RAE): \( \min \left\{ |p - t| \mid p \neq t \right\} \)

- CSVs is censored support vector regression [4]
- Similar results on other large survival datasets, SUPPORT2 and RHC

References