### Cox Proportional Hazard Model

**How it works:** Semi-parametric model for survival analysis, learned by optimizing a partial likelihood function.

**Assumptions:**
- Hazard functions for different learning tasks, which calculate the rate of time before an event occurs, are proportional over time. Thus, patients in the same 30-day readmission window have the same baseline function, but with different attributes, to calculate time before readmission.
- No pre-determined distribution for failure time, which is when a patient experiences a readmission

**Formulation:**

$$ l(B) = \frac{\partial}{\partial d} \sum_{i=1}^{n} \log \left( \sum_{k=1}^{K} \exp \left( B \cdot z_{ij} \right) \right) $$

**Advantages:**
- Allows sharper patient-by-patient predictions and better model interpretability than the current RSF model, through ability to see individualized patient readmission risk projections
- Focuses primarily on how the covariates affect the base hazard function.

**Limitations:**
- Assumes that the survival curves of different patient groups are naturally similarly shaped, which is not a practical assumption

**Results:** Concordance index of 0.70, after 10-fold cross validation.

### Multi-task Cox Proportional Hazard Model

**How it works:** Semi-parametric model, learned by grouping patients into tasks by their diagnoses and finding a shared representation of significant features, to decrease the prediction error of each task.

**Creating tasks:** Patients grouped into task groups based on their admission diagnosis

<table>
<thead>
<tr>
<th>Task</th>
<th>Diagnosis Code</th>
<th>Patient No.</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1</td>
<td>Age, income...</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>4</td>
<td>Age, income...</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>Age, income...</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>6</td>
<td>Age, income...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N</td>
<td>Z</td>
<td>12</td>
<td>Age, income...</td>
</tr>
</tbody>
</table>

**Assumptions:**
- Covariates are generally significant or insignificant across all tasks
- Patient groupings by diagnoses are representative of inherent differences between the different tasks

**Formulation:**

$$ l(B) = \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{B_{ik}}{S_{ik}} - d_{ik} \log \left( \sum_{k=1}^{K} \exp \left( B_{ik} \cdot z_{ij} \right) \right) + \lambda \|B\|_{2,1} $$

**Advantages:**
- Captures dependencies in outcomes at various timepoints by evaluating shared attributes across different tasks
- Creates a separate baseline hazard function for each task-grouping of diagnoses, to allow for more personalized predictions

**Limitations:**
- $L_2,1$ norm seeks features that are either completely “relevant” or “irrelevant” across the different tasks.
- Questions related to how tasks are grouped, including number of observations within each group and the number of overall groups especially within the dataset used to train the multi-task model

**Results:** Concordance index of 0.52.

### Conclusions and Future Works

**Conclusion:** The Cox model outperforms the multi-task Cox model based on concordance index, because:
- The sparse nature of the training dataset limited some of the task groupings to a small number of observations which in turn decreased the robustness of the model for certain tasks.
- Grouping tasks by diagnosis code might not be most suitable for multi-task modeling due to the nature of the patient population and how the diagnosis code feature is distributed.

**Future Works:** The multi-task model is an innovative approach to predicting hospital readmissions, but can be improved by:
- Larger, more robust dataset that is better suited for multi-task learning
- Greater choice and experimentation of tasks by which patients are grouped: utilizing age groups as an alternative to diagnosis codes for task groupings