Deep Learning of Semi-Competing Risk Data

via a Neural Expectation-Maximization Algorithm

Stephen Salerno and Yi Li

Department of Biostatistics, University of Michigan

ENAR 2023 Spring Meeting

March 22, 2023

- 5-year survival rate of *1 in 5* [Bade and Cruz, 2020]

- Prognosis depends on *individualized risk factors* [Ashworth et al., 2014]

- *Progression*, metastasis *prior to death* [Inamura and Ishikawa, 2010]

Approximately 1 in 5 cancer deaths
are attributed to lung cancer.

Source: World Health Organization, International Agency for Research on Cancer, Latest global
cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018.

Many studies report on *lung cancer outcomes*, however:

- Mortality is often studied without considering *competing events*

- Or *composite endpoints* such as *progression-free survival* are used [Jazić et al., 2016]

When progression and death *do not correlate well*, the effects of certain risk factors may differ across *states* of a patient's disease trajectory [Amir et al., 2012, Chakravarty and Sridhara, 2008] [1]

---

[1]https://www.aacr.org/blog/2016/01/04/advancing-lung-cancer-research/

Our data come from the ***Boston Lung Cancer Study***, one of the largest ***cancer epidemiology cohorts*** investigating lung cancer. Our ***scientific questions*** are:

1. What are the ***complex mechanisms*** governing the relationship between risk factors for lung cancer?

2. Do these relationships differ across ***states*** in a patient's ***clinical course***

3. Can we use this information to more ***accurately predict*** patient ***survival***?

- In *survival analysis*, the outcome is the time until a specific event, such as cancer *progression* or *death*, which may be *censored*

- Many survival processes involve *non-terminal* (e.g., progression) and *terminal* (e.g., death) events, which form *semi-competing* relationships [Fine et al., 2001]
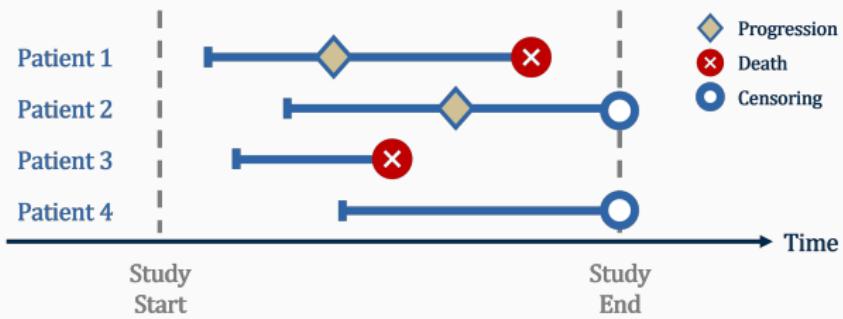

Figure: Schematic of four example patients with semi-competing risks

We base our approach on the ***illness-death model***, a compartment-type model for the ***hazards/transition rates*** between event states:
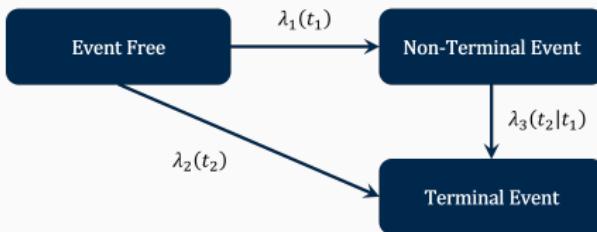

Figure: Illness-death model framework

$$\lambda_1 \left(t_1 \mid \gamma_i, x_i \right) = \gamma_i \times \lambda_{01} \left(t_1 \right) \times \exp \left\{ h_1(x_i) \right\}; \quad t_1 > 0$$

$$\lambda_2 \left(t_2 \mid \gamma_i, x_i \right) = \gamma_i \times \lambda_{02} \left(t_2 \right) \times \exp \left\{ h_2(x_i) \right\}; \quad t_2 > 0$$

$$\underbrace{\lambda_3 \left(t_2 \mid t_1, \gamma_i, x_i \right)}_{\text{Hazard Function}} = \underbrace{\gamma_i}_{\text{Frailty}} \times \underbrace{\lambda_{03} \left(t_2 - t_1 \right)}_{\text{Baseline Hazard}} \times \underbrace{\exp \left\{ h_3(x_i) \right\}}_{\text{Risk Function}}; \quad t_2 > t_1 > 0$$

The likelihood for the observed data $\mathcal{D}$:

$$L\left(\psi; \mathcal{D}\right) = \prod_{i=1}^{n} \int_{0}^{\infty} \frac{\theta^{-\frac{1}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)} \times \gamma_{i}^{\frac{1}{\theta}-1} \times e^{-\frac{\gamma_{i}}{\theta}} \times \gamma_{i}^{\delta_{i1}+\delta_{i2}} \times \left[\lambda_{01}\left(Y_{i1}\right) e^{h_1(\mathbf{x}_i)}\right]^{\delta_{i1}}$$

$$\times \left[\lambda_{02}\left(Y_{i2}\right) e^{h_2(\mathbf{x}_i)}\right]^{(1-\delta_{i1})\delta_{i2}} \times \left[\lambda_{03}\left(Y_{i2}-Y_{i1}\right) e^{h_3(\mathbf{x}_i)}\right]^{\delta_{i1}\delta_{i2}}$$

$$\times \exp\left\{-\gamma_i \left[\Lambda_{01}\left(Y_{i1}\right) e^{h_1(\mathbf{x}_i)} + \Lambda_{02}\left(Y_{i1}\right) e^{h_2(\mathbf{x}_i)} + \delta_{i1}\Lambda_{03}\left(Y_{i2}-Y_{i1}\right) e^{h_3(\mathbf{x}_i)}\right]\right\} d\gamma_i$$

- We want a ***non-parametric*** model for ***both*** the baseline hazards and covariate risk functions to achieve greater ***flexibility*** and ***accuracy***

- Direct maximization of the likelihood function is challenging

- We resort to the EM algorithm which provides a numerically stable approach for optimization

Treating frailties γ as the (unobserved) data, we form the ***complete data likelihood*** as follows:

$$L\left(\psi; \mathcal{D}, \boldsymbol{\gamma}\right) = \prod_{i=1}^{n} \frac{\theta^{-\frac{1}{\theta}}}{\Gamma\left(\frac{1}{\theta}\right)} \times \gamma_i^{\frac{1}{\theta}-1} \times e^{-\frac{\gamma_i}{\theta}} \times \gamma_i^{\delta_{i1}+\delta_{i2}} \times \left[\lambda_{01}\left(Y_{i1}\right) e^{h_1(\boldsymbol{x}_i)}\right]^{\delta_{i1}}$$

$$\times \left[\lambda_{02}\left(Y_{i2}\right) e^{h_2(\boldsymbol{x}_i)}\right]^{(1-\delta_{i1})\delta_{i2}} \times \left[\lambda_{03}\left(Y_{i2}-Y_{i1}\right) e^{h_3(\boldsymbol{x}_i)}\right]^{\delta_{i1}\delta_{i2}}$$

$$\times \exp\left\{-\gamma_i \left[\Lambda_{01}\left(Y_{i1}\right) e^{h_1(\boldsymbol{x}_i)} + \Lambda_{02}\left(Y_{i1}\right) e^{h_2(\boldsymbol{x}_i)} + \delta_{i1}\Lambda_{03}\left(Y_{i2}-Y_{i1}\right) e^{h_3(\boldsymbol{x}_i)}\right]\right\}$$

The **expected log-complete data likelihood**, or our 'Q' function, can be written as:

$$Q\left(\psi \mid \mathcal{D}, \psi^{(m)}\right) = Q_1 + Q_2 + Q_3 + Q_4,$$

where Q_1, Q_2, Q_3, and Q_4 are the **additive pieces** that are **separable** with respect to the model parameters:

$$Q_1 = \sum_{i=1}^{n} \delta_{i1} \mathbb{E}[\log(\gamma_i)|\mathcal{D}, \psi^{(m)}] + \delta_{i1}\{\log[\lambda_{01}(Y_{i1})] + h_1(\boldsymbol{x}_i)\} - \mathbb{E}[\gamma_i|\mathcal{D}, \psi^{(m)}]\Lambda_{01}(Y_{i1})\,e^{h_1(\boldsymbol{x}_i)}$$

$$Q_2 = \sum_{i=1}^{n} \delta_{i2} \mathbb{E}[\log(\gamma_i)|\mathcal{D}, \psi^{(m)}] + (1 - \delta_{i1})\,\delta_{i2}\{\log[\lambda_{02}(Y_{i2})] + h_2(\boldsymbol{x}_i)\} - \mathbb{E}[\gamma_i|\mathcal{D}, \psi^{(m)}]\Lambda_{02}(Y_{i1})\,e^{h_2(\boldsymbol{x}_i)}$$

$$Q_3 = \sum_{i=1}^{n} \delta_{i1}\,\delta_{i2}\{\log[\lambda_{03}(Y_{i2})] + h_3(\boldsymbol{x}_i)\} - \mathbb{E}[\gamma_i|\mathcal{D}, \psi^{(m)}]\delta_{i1}(\Lambda_{03}(Y_{i2} - Y_{i1}))\,e^{h_3(\boldsymbol{x}_i)}$$

$$Q_4 = \sum_{i=1}^{n} -\frac{1}{\theta}\log(\theta) + \left(\frac{1}{\theta} - 1\right)\mathbb{E}[\log(\gamma_i)|\mathcal{D}, \psi^{(m)}] - \frac{1}{\theta}\mathbb{E}[\gamma_i|\mathcal{D}, \psi^{(m)}] - \log\Gamma\left(\frac{1}{\theta}\right)$$
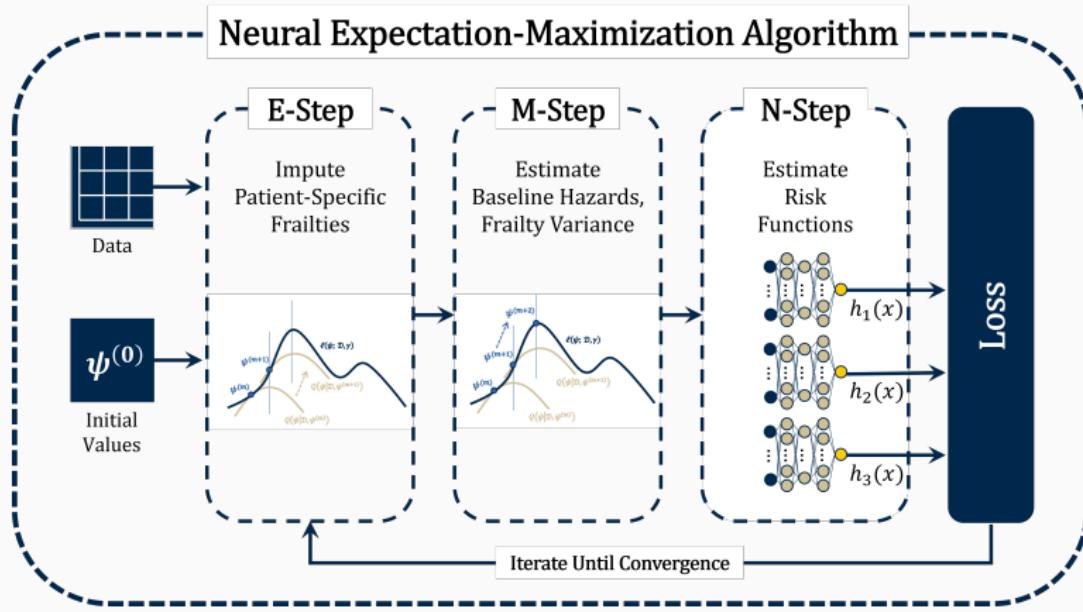
Figure: Overview of our proposed neural expectation-maximization algorithm

We extend the *EM alorithm* to a hybrid *multi-task deep learning* approach for semi-competing risk prediction. Our proposed *neural EM algorithm* consists of

· **E-Step**: *Frailties* estimated given data, current baseline hazard values, and current risk function estimates

· **M-Step**: Estimate fully *non-parametric cumulative baseline hazard* by non-decreasing *step functions* (i.e., jumps at unique failure times); estimate frailty *variance*

· **N-Step**: Maximize the expectation of the complete data likelihood given the data and the estimates of frailty and baseline hazard with respect to deep neural network parameters for each *risk function*, $h_g(x_i)$; $g = 1, 2, 3$

· Sub-network predictions are based on an *L*-fold ***composite function***

$$F_L(\cdot) = f_L \circ f_{L-1} \circ \cdots \circ f_1(\cdot) \text{ where } (g \circ f)(\cdot) = g(f(\cdot))$$

$$f_l(x) = \sigma_l\left(\mathbf{W}_l x + \mathrm{b}_l\right) \in \mathbb{R}^{k_{l+1}}$$

· where σ_l is an activation function, $\mathbf{W}_l$ are weights, and b_l are biases

· For ***identifiability***, we require $h_g(\mathbf{0}) = 0$, where $\mathbf{0}$ is a *p*-vector of 0's

· ***Hyperparameters*** (e.g., hidden layers, dropout fraction, learning rate) are ***optimized*** over a grid of values based on ***predictive performance***

No metrics tailored to semi-competing risks exist to assess *predictive accuracy*. We propose a bivariate extension to the **Brier Score** [Brier et al., 1950]

$$BBS_c(t) = \frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \leq t,\ \delta_{i1} = 1,\ Y_{i1} \leq Y_{i2}\}}{\hat{G}_i(Y_{i1})}$$

$$+ \frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \leq t,\ Y_{i2} \leq t,\ \delta_{i1} = 0,\ \delta_{i2} = 1,\ Y_{i1} \leq Y_{i2}\}}{\hat{G}_i(Y_{i2})}$$

$$+ \frac{[1 - \pi_i(t)]^2 \cdot \mathbb{I}\{Y_{i1} > t,\ Y_{i2} > t\}}{\hat{G}_i(t)}$$

· $\pi_i(t)$ is an **estimate** of $S_i(t) = \Pr(T_{i1} > t,\ T_{i2} > t)$

· $\hat{G}_i(t)$ is an **estimate** of $G_i(t) = \Pr(C_i > t) > 0$

· In **expectation**, we have $\mathbb{E}[BBS_c(t)]$ equal to the **MSE** of $\pi_i(t)$, plus a **constant piece** with respect to $\pi_i(t)$

We generated 500 **independent datasets**, $x_i \sim N_2(0, I_2)$, $\beta_g = [1, 1]^T$, $g = 1, 2, 3$, $\phi_{11} = \phi_{21} = 2$, $\phi_{31} = 0.75$, $\phi_{12} = \phi_{22} = 2.25$, and $\phi_{32} = 2$, and **we varied**:

- **Sample Sizes** (n): 1,000 and 10,000

- **Frailty Variances** (θ): 0.5 and 2

- **Censoring Rates**: 0%, 25%, and 50%

- **Log-Risk Functions**:

  - **Linear**: $h_g(\boldsymbol{X}_i) = \boldsymbol{X}_i' \beta_g$; $\beta_g = \mathbf{1}_p = (1, 1, \ldots, 1)$; $g = 1, 2, 3$
  - **Non-Linear**: $h_g(\boldsymbol{X}_i) = \sum_{j=1}^{p} x_{ij}^3 \beta_{gj}$; $\beta_{gj} = 1$; $g = 1, 2, 3$; $j = 1, \ldots, p$
  - **Non-Monotonic**: $h_g(\boldsymbol{X}_i) = \log(|\boldsymbol{X}_i' \beta_g| + 1)$; $\beta_g = \mathbf{1}_p = (1, 1, \ldots, 1)$; $g = 1, 2, 3$

Performance was assessed via the **bivariate Brier score** integrated up to $t = 1$ year and the **MISE for the log-risk surfaces**, separately

---

Table: Average (SD) mean integrated squared errors (MISE) for the simulated log-risk surfaces, $h_g(\boldsymbol{X}_i)$, for each state transition hazard (50% censoring)

	Simulation Settings		Parametric Approach			Neural EM Algorithm		
n	θ	Risk	$h_1(x_i)$	$h_2(x_i)$	$h_3(x_i)$	$h_1(x_i)$	$h_2(x_i)$	$h_3(x_i)$
1,000	0.5	Linear	0.02 (0.02)	0.03 (0.02)	0.05 (0.03)	0.10 (0.07)	0.10 (0.09)	0.18 (0.17)
10,000	0.5	Linear	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.10 (0.07)	0.11 (0.08)	0.17 (0.16)
1,000	2.0	Linear	0.03 (0.03)	0.03 (0.02)	0.03 (0.05)	0.22 (0.13)	0.19 (0.13)	0.22 (0.17)
10,000	2.0	Linear	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)	0.14 (0.09)	0.14 (0.10)	0.16 (0.14)
1,000	0.5	Non-Linear	0.16 (0.11)	0.20 (0.12)	0.43 (0.42)	0.11 (0.05)	0.09 (0.06)	0.08 (0.02)
10,000	0.5	Non-Linear	0.19 (0.03)	0.22 (0.05)	0.22 (0.06)	0.08 (0.03)	0.08 (0.03)	0.11 (0.05)
1,000	2.0	Non-Linear	0.30 (0.27)	0.31 (0.20)	0.37 (0.30)	0.14 (0.11)	0.12 (0.06)	0.13 (0.06)
10,000	2.0	Non-Linear	0.22 (0.05)	0.20 (0.04)	0.19 (0.06)	0.16 (0.08)	0.14 (0.07)	0.14 (0.08)
1,000	0.5	Non-Monotonic	2.04 (0.51)	2.00 (0.66)	2.57 (1.01)	0.11 (0.12)	0.13 (0.13)	0.18 (0.14)
10,000	0.5	Non-Monotonic	2.04 (0.20)	2.05 (0.19)	2.33 (0.25)	0.06 (0.03)	0.09 (0.09)	0.14 (0.09)
1,000	2.0	Non-Monotonic	2.13 (0.69)	2.00 (0.68)	2.43 (0.88)	0.18 (0.10)	0.18 (0.09)	0.16 (0.11)
10,000	2.0	Non-Monotonic	1.94 (0.22)	1.95 (0.24)	2.25 (0.30)	0.10 (0.05)	0.11 (0.08)	0.15 (0.10)

This study subset includes ***7,460 patients*** with non-small cell lung cancer, diagnosed between June 1983 and October 2021 [Christiani, 2017]

We investigate ***time to disease progression and death***, where progression might be censored by death or the study endpoint

Table: Semi-competing event rates among $n = 7,460$ patients in our analytic sample.

Progression Observed / Death Observed	Yes	No
Yes	143 (1.9%)	295 (4.0%)
No	2,720 (36.5%)	4,302 (57.7%)

Figure: Average estimated cumulative baseline hazard functions and 95% bootstrap confidence intervals for each state transition based on 50 bootstrap samples of our data.
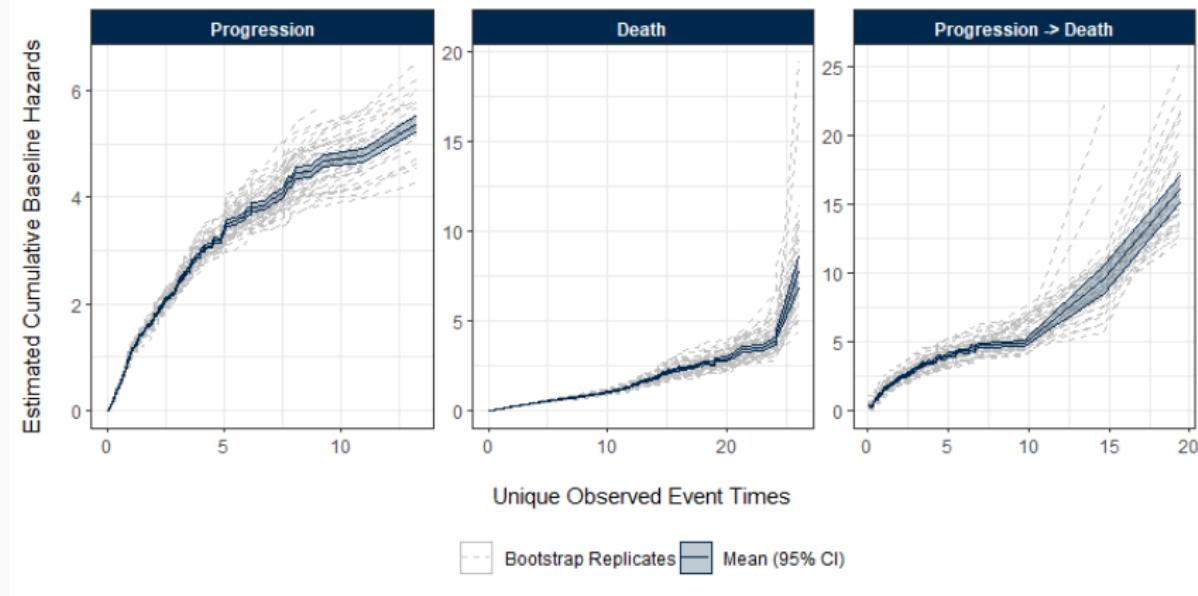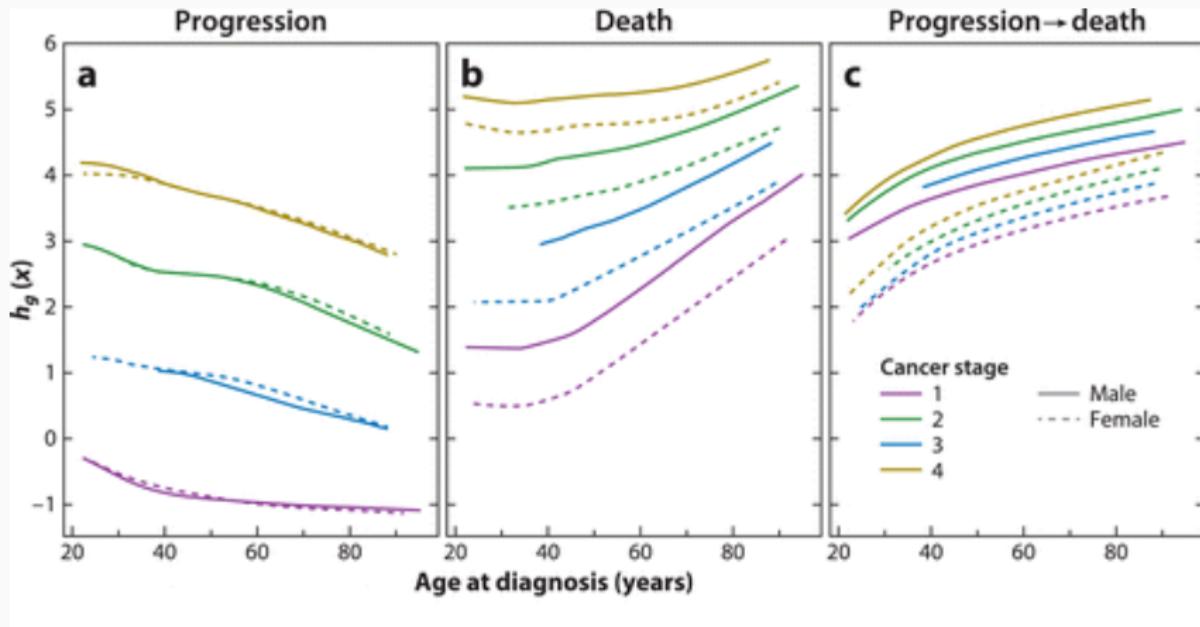
Figure: Example log-risk functions of age at diagnosis on each state transition, stratified by cancer stage (line color) and sex (solid versus dashed lines).

- Baseline hazards highest in **_sojourn time_** between progression and death

- Non-linear relationship between **_age_** and hazards, which **_differs by transition_**

- Stage at diagnosis appears to have a strong effect on the **_risk of progression or death_** from diagnosis, and to a lesser extent death following progression

- **_Males_** have a **_higher risk_** of mortality than females

- **_5-year iBBS_** for our method was 0.32 vs. 0.68 from a traditional model, suggesting that a linear model is not be predictive

We estimate the frailty variance, , to be 3.15 (bootstrapped 95

- Assuming a ***linear relationship*** between risk factors may be an oversimplification in settings of complex cancers

- Deep learning allows for ***non-parametric*** estimation of risk functions

- Neural networks circumvent the ***curse of dimensionality*** by projecting data into a lower relevant representational space through weighting [Bauer and Kohler, 2019, Poggio et al., 2017]

· We proposed a ***neural expectation-maximization algorithm*** to predict time-to-event outcomes arising from semi-competing risks

· In simulation, our results show ***high accuracy*** in estimating the relationship predictors and the transition hazards in increasingly complex settings

· Our approach had a much greater ***predictive accuracy*** than traditional semi-competing regression approaches when applied to the ***BLCS***

· We detected potential ***non-linear effects*** and ***interactions*** between commonly-studied risk factors such as age, sex, and stage at diagnosis

Questions?

salernos@umich.edu

https://arxiv.org/abs/2212.12028

- My advisor: Yi Li
- My committee
- My lab mates
- David Christiani and the Christiani lab, Xinan Wang, and Jui Kothari
- My co-workers at KECC
- STATCOM

Appendix

Let T_{i1}, T_{i2}, and C_i denote the time to ***progression***, ***death***, and ***censoring*** for the ith individual in our study. In practice, we can only ***observe***:

$$\mathcal{D} = \{(Y_{i1},\ \delta_{i1},\ Y_{i2},\ \delta_{i2},\ x_i);\ i = 1, \ldots, n\}$$

Observed Data Definitions

$Y_{i2} = \min(T_{i2}, C_i)$

$\delta_{i2} = I(T_{i2} \leq C_i)$

$Y_{i1} = \min(T_{i1}, Y_{i2})$

$\delta_{i1} = I(T_{i1} \leq Y_{i2})$

$x_i = $ Covariates

1. $\mathcal{D} = (T_1, 1, T_2, 1)$
2. $\mathcal{D} = (T_2, 0, T_2, 1)$
3. $\mathcal{D} = (T_1, 1, C, 0)$
4. $\mathcal{D} = (C, 0, C, 0)$
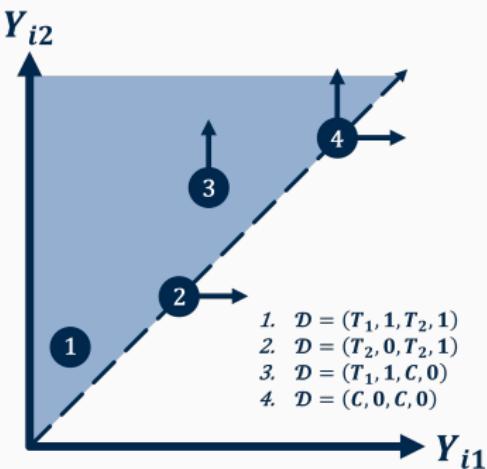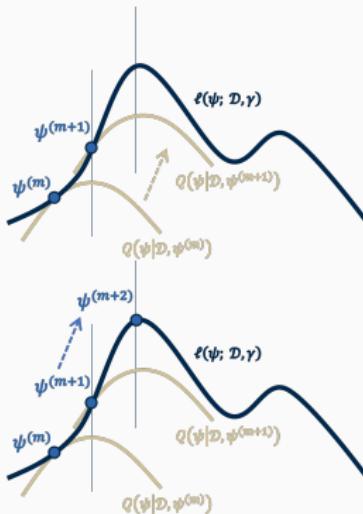
Figure: Observable data

$$\lambda_1\left(t_1\right) = \lim_{\Delta \to 0} \Pr\left[T_1 \in \left[t_1, t_1 + \Delta\right) \mid T_1 \geq t_1, T_2 \geq t_1\right] / \Delta; \quad t_1 > 0$$
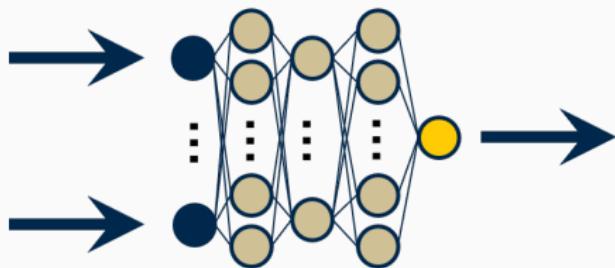
$$\lambda_2\left(t_2\right) = \lim_{\Delta \to 0} \Pr\left[T_2 \in \left[t_2, t_2 + \Delta\right) \mid T_1 \geq t_2, T_2 \geq t_2\right] / \Delta; \quad t_2 > 0$$

$$\lambda_3\left(t_2 \mid t_1\right) = \lim_{\Delta \to 0} \Pr\left[T_2 \in \left[t_2, t_2 + \Delta\right) \mid T_1 = t_1, T_2 \geq t_2\right] / \Delta; \quad t_2 > t_1 > 0$$

- *E Step*: Frailties *estimated* given data and current baseline hazard values

- *M Step*: Baseline hazards and frailty variance *estimated* given current frailty estimates

- But how do we estimate the complex relationships between *risk factors*?

Deep learning has emerged as a powerful tool for *survival prediction*, but limited work has been done on multi-state outcomes, let alone *semi-competing*

In artificial neural networks, *nodes* are connected as a weighted sum of inputs *affine transformations* and *nonlinear activations* [Bauer and Kohler, 2019]

We propose the use of *deep learning* to estimate the *risk functions* for each hazard (i.e, state transition)

E-Step

Calculate

$Q(\psi \mid \mathcal{D}, \psi^{(m)})$

and update

$\gamma_i \mid \mathcal{D}, \psi^{(m)}$

With $\gamma_i \sim \text{Gamma}\left(\frac{1}{\theta}, \frac{1}{\theta}\right)$, calculate expected log-complete data likelihood given the observed data
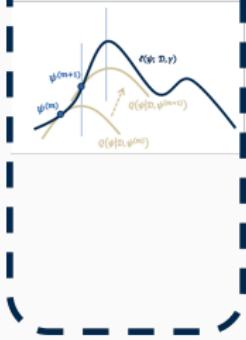
$$Q\left(\psi \mid \mathcal{D}, \psi^{(m)}\right) = \mathbb{E}_{\gamma}\left[\ell\left(\psi; \mathcal{D}, \gamma\right) \mid \mathcal{D}, \psi^{(m)}\right]$$

where $\ell\left(\psi; \mathcal{D}, \gamma\right) = \log L\left(\psi; \mathcal{D}, \gamma\right)$

As a useful quantity, we have

$$\mathbb{E}[\gamma_i \mid \mathcal{D}, \psi^{(m)}]$$
$$= \frac{1/\tilde{\theta} + \delta_{i1} + \delta_{i2}}{1/\tilde{\theta} + \tilde{\Lambda}_{01}(Y_{i1})e^{\tilde{h}_1(x_i)} + \tilde{\Lambda}_{02}(y_{i1})e^{\tilde{h}_2(x_i)} + \delta_{i1}\tilde{\Lambda}_{03}(y_{i1}, y_{i2})e^{\tilde{h}_3(x_i)}}$$

Update estimates of the ***baseline hazards***, which resemble Nelson-Aalen type estimators

$$\Delta\Lambda_{01}^{(m+1)}(t) = \frac{\sum_{i=1}^{n} \delta_{i1} I\left[Y_{i1} = t\right]}{\sum_{i=1}^{n} \mathbb{E}[\gamma_i | \mathcal{D}, \psi^{(m)}] I\left[Y_{i1} \geq t\right] \exp\left\{h_1^{(m)}(\boldsymbol{x}_i)\right\}}$$

$$\Delta\Lambda_{02}^{(m+1)}(t) = \frac{\sum_{i=1}^{n} (1 - \delta_{i1}) \delta_{i2} I\left[Y_{i2} = t\right]}{\sum_{i=1}^{n} \mathbb{E}[\gamma_i | \mathcal{D}, \psi^{(m)}] I\left[Y_{i2} \geq t\right] \exp\left\{h_2^{(m)}(\boldsymbol{x}_i)\right\}}$$
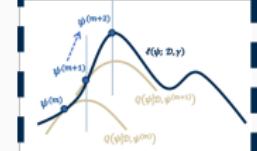
$$\Delta\Lambda_{03}^{(m+1)}(t) = \frac{\sum_{i=1}^{n} \delta_{i1} \delta_{i2} I\left[Y_{i2} - Y_{i1} = t\right]}{\sum_{i=1}^{n} \mathbb{E}[\gamma_i | \mathcal{D}, \psi^{(m)}] \delta_{i1} I\left[Y_{i2} - Y_{i1} \geq t\right] \exp\left\{h_3^{(m)}(\boldsymbol{x}_i)\right\}}$$

M-Step

Maximize

$\psi^{(m+1)}$

with respect to
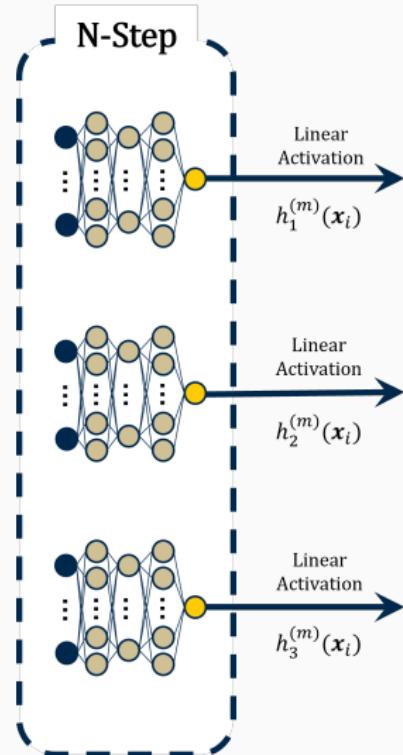
$Q(\psi \mid \mathcal{D}, \psi^{(m)})$

Update the *risk functions* with outputs from three *network sub-architectures*

Sub-networks are *fully-connected feed-forward networks* with linear activations in the final layer

Each sub-network is made up of L layers, with k_l neurons in the lth layer

In *expectation*, we have that the Bivariate Brier Score is equal to

$$\mathbb{E}\left[BBS_c(t)\right] = \text{MSE}(t) + \frac{1}{n}\sum_{i=1}^{n} S_i(t) \cdot [1 - S_i(t)]$$

- This is the *MSE* of $\pi_i(t)$, plus a *constant piece* with respect to $\pi_i(t)$

- Constant is *irreducible error* incurred by approximating $S_i(t)$

Baseline Hazards
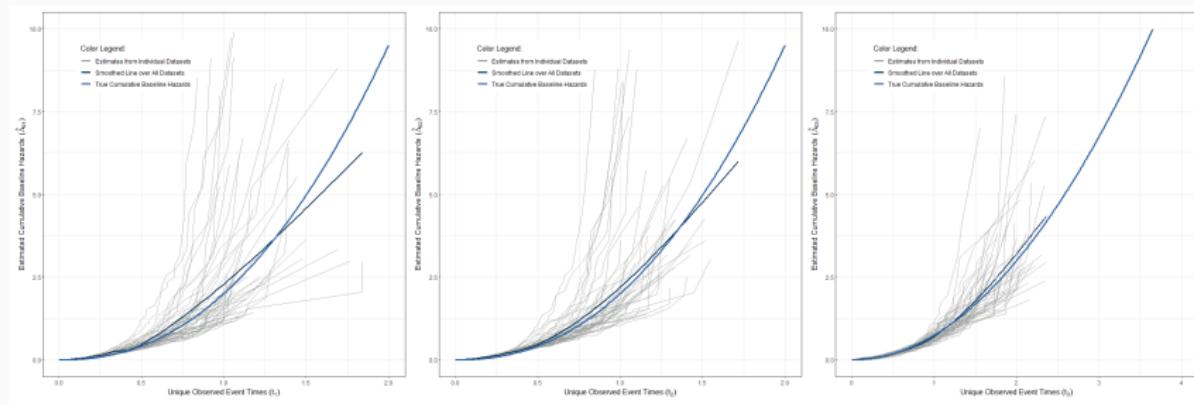
Figure: Estimated cumulative baseline hazard functions based on an example 50 generated datasets with $n = 1,000$, $\theta = 0.5$, log-risk function = non-monotonic, and 50% censoring rates
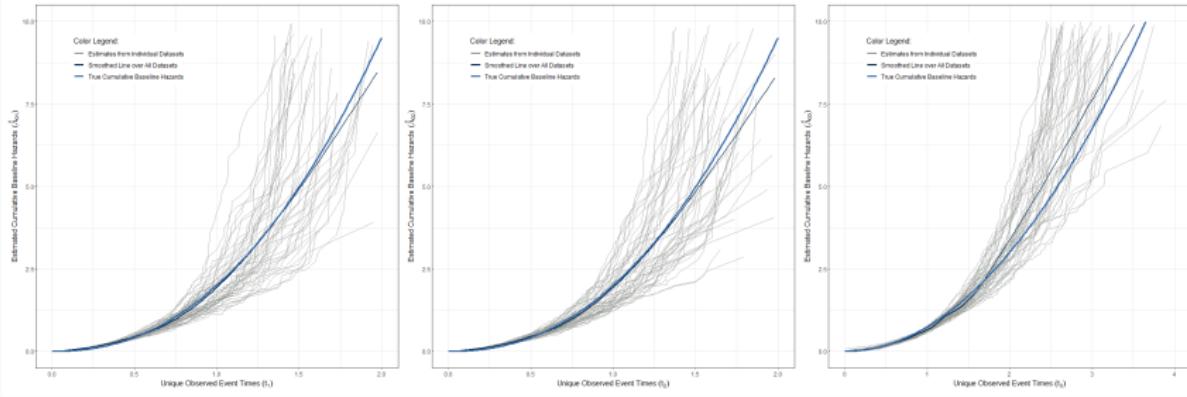
Figure: Estimated cumulative baseline hazard functions based on an example 50 generated datasets with $n = 1,000$, $\theta = 0.5$, log-risk function = non-monotonic, and censoring rate = 0%
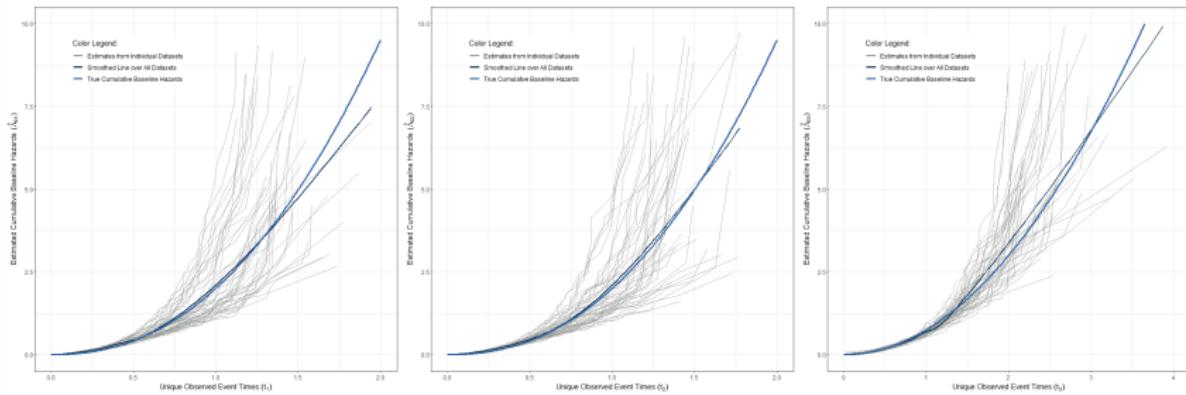
Figure: Estimated cumulative baseline hazard functions based on an example 50 generated datasets with $n = 1,000$, $\theta = 0.5$, log-risk function = non-monotonic, and censoring rate = 25%

Table: Estimated frailty variance and one year integrated bivariate Brier score under various simulation settings (50% censoring rate)

	Simulation Settings		Integrated Bivariate Brier Score	
n	Risk	θ	Truth	Neural EM
1,000	Linear	0.5	0.1853 (0.0060)	0.1866 (0.0075)
10,000	Linear	0.5	0.1912 (0.0042)	0.1934 (0.0044)
1,000	Linear	2.0	0.3081 (0.0323)	0.3129 (0.0355)
10,000	Linear	2.0	0.3050 (0.0133)	0.3090 (0.0142)
1,000	Non-Linear	0.5	0.1794 (0.0037)	0.1830 (0.0056)
10,000	Non-Linear	0.5	0.1819 (0.0022)	0.1853 (0.0072)
1,000	Non-Linear	2.0	0.3074 (0.0211)	0.3127 (0.0264)
10,000	Non-Linear	2.0	0.3076 (0.0065)	0.3179 (0.0146)
1,000	Non-Monotonic	0.5	0.1794 (0.0037)	0.1830 (0.0056)
10,000	Non-Monotonic	0.5	0.1819 (0.0022)	0.1853 (0.0072)
1,000	Non-Monotonic	2.0	0.3074 (0.0211)	0.3127 (0.0264)
10,000	Non-Monotonic	2.0	0.3076 (0.0065)	0.3179 (0.0146)

Table: Estimated frailty variance and one year integrated bivariate Brier score under various simulation settings (50% censoring rate)

Simulation Settings		Frailty Variance Estimation			Integrated Bivariate Brier Score		
n	Risk	Truth	Parametric	Neural EM	Truth	Parametric	Neural EM
1,000	Linear	0.5	0.47 (0.17)	0.47 (0.12)	0.1853 (0.0060)	0.1858 (0.0066)	0.1866 (0.0075)
10,000	Linear	0.5	0.51 (0.05)	0.50 (0.05)	0.1912 (0.0042)	0.1912 (0.0046)	0.1934 (0.0044)
1,000	Linear	2.0	1.96 (0.48)	1.87 (0.31)	0.3081 (0.0323)	0.3092 (0.0387)	0.3129 (0.0355)
10,000	Linear	2.0	2.00 (0.16)	1.95 (0.06)	0.3050 (0.0133)	0.3050 (0.0129)	0.3090 (0.0142)
1,000	Non-Linear	0.5	0.47 (0.20)	0.49 (0.12)	0.1794 (0.0037)	0.1808 (0.0032)	0.1830 (0.0056)
10,000	Non-Linear	0.5	0.51 (0.06)	0.49 (0.03)	0.1819 (0.0022)	0.1833 (0.0026)	0.1853 (0.0072)
1,000	Non-Linear	2.0	1.93 (0.28)	1.87 (0.21)	0.3074 (0.0211)	0.3090 (0.0239)	0.3127 (0.0264)
10,000	Non-Linear	2.0	2.06 (0.11)	1.94 (0.09)	0.3076 (0.0065)	0.3092 (0.0077)	0.3179 (0.0146)
1,000	Non-Monotonic	0.5	0.47 (0.20)	0.49 (0.13)	0.1794 (0.0037)	0.1808 (0.0032)	0.1830 (0.0056)
10,000	Non-Monotonic	0.5	0.51 (0.06)	0.49 (0.04)	0.1819 (0.0022)	0.1833 (0.0026)	0.1853 (0.0072)
1,000	Non-Monotonic	2.0	1.93 (0.28)	1.96 (0.21)	0.3074 (0.0211)	0.3090 (0.0239)	0.3127 (0.0264)
10,000	Non-Monotonic	2.0	2.06 (0.11)	1.94 (0.09)	0.3076 (0.0065)	0.3092 (0.0077)	0.3179 (0.0146)

Table: Estimated frailty variance and one year integrated bivariate Brier score under various simulation settings

Simulation Settings			Frailty Variance Estimation			Integrated Bivariate Brier Score		
n	Risk	Cens.	Truth	Parametric	Neural EM	Truth	Parametric	Neural EM
1,000	Linear	0%	0.5	0.49 (0.04)	0.49 (0.07)	0.1600 (0.0035)	0.1605 (0.0035)	0.1618 (0.0035)
10,000	Linear	0%	0.5	0.50 (0.01)	0.49 (0.03)	0.1584 (0.0013)	0.1585 (0.0013)	0.1594 (0.0014)
1,000	Linear	0%	2.0	1.96 (0.11)	1.92 (0.08)	0.1915 (0.0047)	0.1917 (0.0047)	0.1925 (0.0049)
10,000	Linear	0%	2.0	2.01 (0.03)	1.98 (0.04)	0.1911 (0.0016)	0.1911 (0.0016)	0.1921 (0.0015)
1,000	Non-Linear	0%	0.5	0.49 (0.05)	0.50 (0.08)	0.1822 (0.0038)	0.1841 (0.0036)	0.1855 (0.0036)
10,000	Non-Linear	0%	0.5	0.51 (0.01)	0.51 (0.02)	0.1821 (0.0011)	0.1836 (0.0011)	0.1858 (0.0014)
1,000	Non-Linear	0%	2.0	1.97 (0.11)	1.92 (0.07)	0.2244 (0.0022)	0.2258 (0.0022)	0.2271 (0.0024)
10,000	Non-Linear	0%	2.0	2.01 (0.03)	1.95 (0.03)	0.2245 (0.0009)	0.2251 (0.0009)	0.2276 (0.0015)
1,000	Non-Monotonic	0%	0.5	0.49 (0.05)	0.50 (0.08)	0.1822 (0.0038)	0.1841 (0.0036)	0.1855 (0.0036)
10,000	Non-Monotonic	0%	0.5	0.51 (0.01)	0.51 (0.02)	0.1821 (0.0011)	0.1836 (0.0011)	0.1858 (0.0014)
1,000	Non-Monotonic	0%	2.0	1.97 (0.11)	1.95 (0.07)	0.2244 (0.0022)	0.2258 (0.0022)	0.2271 (0.0024)
10,000	Non-Monotonic	0%	2.0	2.01 (0.03)	1.95 (0.03)	0.2245 (0.0009)	0.2251 (0.0009)	0.2276 (0.0015)
1,000	Linear	25%	0.5	0.47 (0.10)	0.47 (0.11)	0.1880 (0.0046)	0.1886 (0.0050)	0.1892 (0.0053)
10,000	Linear	25%	0.5	0.50 (0.05)	0.48 (0.02)	0.1899 (0.0030)	0.1900 (0.0029)	0.1914 (0.0029)
1,000	Linear	25%	2.0	2.00 (0.35)	1.95 (0.20)	0.2967 (0.0200)	0.2970 (0.0228)	0.2999 (0.0224)
10,000	Linear	25%	2.0	2.02 (0.12)	1.96 (0.08)	0.2979 (0.0074)	0.2979 (0.0069)	0.3030 (0.0079)
1,000	Non-Linear	25%	0.5	0.47 (0.14)	0.48 (0.12)	0.1851 (0.0045)	0.1866 (0.0040)	0.1879 (0.0048)
10,000	Non-Linear	25%	0.5	0.52 (0.05)	0.50 (0.04)	0.1858 (0.0010)	0.1874 (0.0012)	0.1893 (0.0025)
1,000	Non-Linear	25%	2.0	2.01 (0.23)	1.89 (0.21)	0.3042 (0.0176)	0.3065 (0.0220)	0.3088 (0.0201)
10,000	Non-Linear	25%	2.0	2.04 (0.10)	1.96 (0.10)	0.3032 (0.0034)	0.3044 (0.0048)	0.3113 (0.0093)
1,000	Non-Monotonic	25%	0.5	0.47 (0.14)	0.47 (0.12)	0.1851 (0.0045)	0.1866 (0.0040)	0.1879 (0.0048)
10,000	Non-Monotonic	25%	0.5	0.52 (0.05)	0.50 (0.04)	0.1858 (0.0010)	0.1874 (0.0012)	0.1893 (0.0025)
1,000	Non-Monotonic	25%	2.0	2.01 (0.23)	1.90 (0.21)	0.3042 (0.0176)	0.3065 (0.0220)	0.3088 (0.0201)
10,000	Non-Monotonic	25%	2.0	2.04 (0.10)	1.91 (0.10)	0.3032 (0.0034)	0.3044 (0.0048)	0.3113 (0.0093)

Table: Average (SD) mean integrated squared errors (MISE) for the simulated log-risk surfaces, $h_g(\boldsymbol{X}_i)$, for each state transition hazard
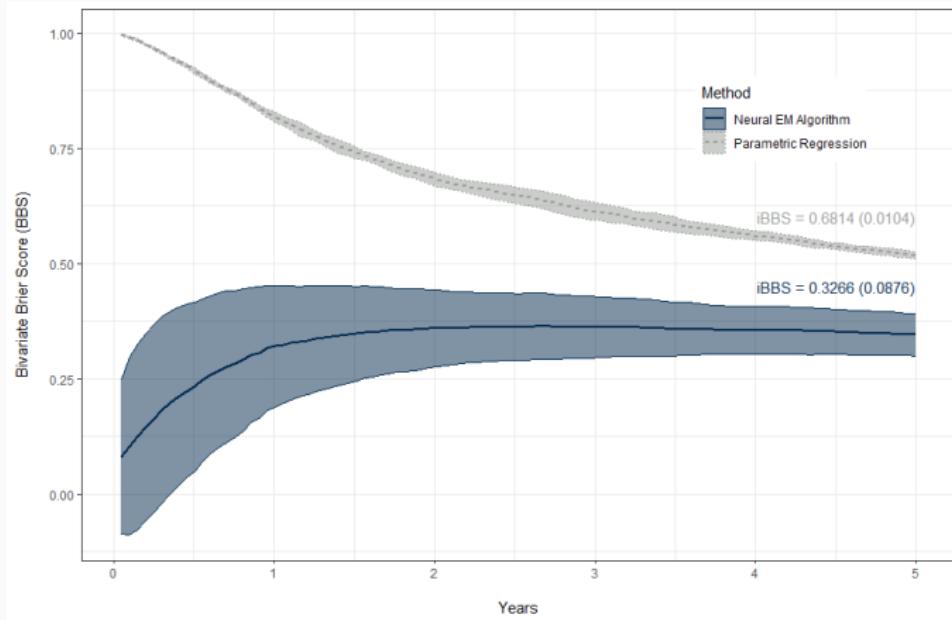
	Simulation Settings			Parametric Approach			Neural EM Algorithm		
n	θ	Risk	Cens.	$h_1(x_i)$	$h_2(x_i)$	$h_3(x_i)$	$h_1(x_i)$	$h_2(x_i)$	$h_3(x_i)$
1,000	0.5	Linear	0%	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.07 (0.07)	0.08 (0.08)	0.07 (0.05)
10,000	0.5	Linear	0%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.08 (0.06)	0.07 (0.05)	0.07 (0.04)
1,000	2.0	Linear	0%	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)	0.12 (0.07)	0.11 (0.08)	0.12 (0.09)
10,000	2.0	Linear	0%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.11 (0.06)	0.11 (0.06)	0.13 (0.09)
1,000	0.5	Non-Linear	0%	0.17 (0.07)	0.15 (0.06)	0.19 (0.08)	0.07 (0.04)	0.09 (0.03)	0.07 (0.01)
10,000	0.5	Non-Linear	0%	0.17 (0.02)	0.19 (0.04)	0.18 (0.02)	0.08 (0.01)	0.08 (0.03)	0.08 (0.03)
1,000	2.0	Non-Linear	0%	0.22 (0.15)	0.27 (0.15)	0.22 (0.18)	0.15 (0.01)	0.10 (0.05)	0.11 (0.04)
10,000	2.0	Non-Linear	0%	0.20 (0.04)	0.19 (0.03)	0.20 (0.06)	0.14 (0.07)	0.14 (0.08)	0.12 (0.05)
1,000	0.5	Non-Monotonic	0%	1.79 (0.32)	1.79 (0.38)	1.83 (0.35)	0.09 (0.05)	0.09 (0.04)	0.09 (0.06)
10,000	0.5	Non-Monotonic	0%	1.82 (0.11)	1.81 (0.14)	1.76 (0.12)	0.07 (0.03)	0.08 (0.03)	0.08 (0.05)
1,000	2.0	Non-Monotonic	0%	1.89 (0.49)	1.86 (0.53)	1.97 (0.52)	0.15 (0.05)	0.13 (0.06)	0.14 (0.07)
10,000	2.0	Non-Monotonic	0%	1.82 (0.18)	1.80 (0.18)	1.85 (0.17)	0.14 (0.04)	0.12 (0.03)	0.14 (0.06)
1,000	0.5	Linear	25%	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.11 (0.10)	0.10 (0.10)	0.13 (0.12)
10,000	0.5	Linear	25%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.12 (0.08)	0.12 (0.05)	0.12 (0.10)
1,000	2.0	Linear	25%	0.03 (0.02)	0.02 (0.02)	0.03 (0.03)	0.15 (0.10)	0.13 (0.08)	0.16 (0.11)
10,000	2.0	Linear	25%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.14 (0.09)	0.12 (0.06)	0.14 (0.10)
1,000	0.5	Non-Linear	25%	0.18 (0.12)	0.15 (0.06)	0.32 (0.30)	0.10 (0.04)	0.08 (0.02)	0.07 (0.02)
10,000	0.5	Non-Linear	25%	0.19 (0.03)	0.21 (0.05)	0.21 (0.06)	0.08 (0.03)	0.08 (0.04)	0.09 (0.04)
1,000	2.0	Non-Linear	25%	0.29 (0.24)	0.31 (0.20)	0.38 (0.29)	0.14 (0.05)	0.11 (0.05)	0.12 (0.04)
10,000	2.0	Non-Linear	25%	0.21 (0.04)	0.20 (0.03)	0.19 (0.06)	0.12 (0.04)	0.14 (0.06)	0.19 (0.19)
1,000	0.5	Non-Monotonic	25%	1.97 (0.47)	2.02 (0.51)	2.18 (0.60)	0.10 (0.08)	0.10 (0.08)	0.12 (0.08)
10,000	0.5	Non-Monotonic	25%	1.92 (0.16)	1.91 (0.16)	2.16 (0.17)	0.09 (0.04)	0.09 (0.05)	0.11 (0.06)
1,000	2.0	Non-Monotonic	25%	2.00 (0.62)	1.97 (0.69)	2.25 (0.75)	0.13 (0.07)	0.15 (0.08)	0.13 (0.06)
10,000	2.0	Non-Monotonic	25%	1.85 (0.20)	1.85 (0.21)	2.12 (0.27)	0.10 (0.05)	0.11 (0.06)	0.11 (0.05)

- Generated 1,000 **independent datasets** of size $n = 1,000$

- Assumed **Weibull** baseline hazards with $\phi_{g1} = 1.5$, $\phi_{g2} = 0.2$, and $\theta = 0.5$

- Varied whether **covariate** generated and **censoring rate** of 0% vs. 50%

Table: Mean (SD) integrated Bivariate Brier Score under various data generation settings

Simulation Settings		1-Year iBBS	
Covariate	Censoring	True	Estimated
No	No	0.0187 (0.0068)	0.0199 (0.0073)
Yes	No	0.0181 (0.0067)	0.0205 (0.0077)
No	Yes	0.0206 (0.0067)	0.0219 (0.0072)
Yes	Yes	0.0195 (0.0066)	0.0221 (0.0075)

M | BIOSTATISTICS

Figure: Average (SD) Bivariate Brier score (BBS) for our Neural EM Algorithm (blue, solid line) versus a semi-competing regression model (gray, dashed line), with 5-fold cross-validation. Integrated (BBS) was taken over 100 evenly spaced time points from time zero to five years post-diagnosis.

Current and Future Work

- Due to **dependent censoring**, progression or other non-terminal events as outcomes are often forgone, despite being of clinical interest

- We are investigating whether we can achieve high predictive accuracy for dependently-censored events with a simplified **objective function**

- We are currently working on a **jackknife pseudo-value approach**:

  - Reduces the problem to a logistic loss function

  - Consistent and does not imposes common assumptions across time points

  - Targets are more interpretable quantities such as survival probabilities or restricted mean survival times

- We formulate the semi-competing outcomes based on the ***Clayton copula model*** to construct our outcome pseudo-values:

  - Equivalent to the ***gamma frailty conditional Markov model*** used here, under some simplifying assumptions

  - Admits an ***explicit form*** and can model dependence for arbitrarily high dimensions with one parameter

- This approach facilitates the development of ***causal estimators*** for targets such as marginal survival probabilities and restricted mean survival times

- We will further develop of ***individualized prediction intervals***

- We will ***validate*** our approach in simulation and ***apply*** it to the BLCS

[Amir et al., 2012] Amir, E., Seruga, B., Kwong, R., Tannock, I. F., and Ocaña, A. (2012).
Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer?
European Journal of Cancer, 48(3):385–388.

[Ashworth et al., 2014] Ashworth, A. B., Senan, S., Palma, D. A., Riquet, M., Ahn, Y. C., Ricardi, U., Congedo, M. T., Gomez, D. R., Wright, G. M., Melloni, G., et al. (2014).
An individual patient data metaanalysis of outcomes and prognostic factors after treatment of oligometastatic non−small-cell lung cancer.
Clinical lung cancer, 15(5):346–355.

[Bade and Cruz, 2020] Bade, B. C. and Cruz, C. S. D. (2020).
Lung cancer 2020: epidemiology, etiology, and prevention.
Clinics in chest medicine, 41(1):1–24.

[Bauer and Kohler, 2019] Bauer, B. and Kohler, M. (2019).
On deep learning as a remedy for the curse of dimensionality in nonparametric regression.
The Annals of Statistics, 47(4):2261–2285.

[Brier et al., 1950] Brier, G. W. et al. (1950).
Verification of forecasts expressed in terms of probability.
Monthly weather review, 78(1):1–3.

[Chakravarty and Sridhara, 2008] Chakravarty, A. and Sridhara, R. (2008).
Use of progression-free survival as a surrogate marker in oncology trials: some regulatory issues.
Statistical Methods in Medical Research, 17(5):515–518.

[Christiani, 2017] Christiani, D. C. (2017).
The Boston lung cancer survival cohort.
`http://grantome.com/grant/NIH/U01-CA209414-01A1.`
[Online; accessed November 12, 2022].

[Fine et al., 2001] Fine, J. P., Jiang, H., and Chappell, R. (2001).
On semi-competing risks data.
Biometrika, 88(4):907–919.

[Inamura and Ishikawa, 2010] Inamura, K. and Ishikawa, Y. (2010).
Lung cancer progression and metastasis from the prognostic point of view.
Clinical & experimental metastasis, 27(6):389–397.

[Jazić et al., 2016] Jazić, I., Schrag, D., Sargent, D. J., and Haneuse, S. (2016). **Beyond composite endpoints analysis: semicompeting risks as an underutilized framework for cancer research.** *JNCI: Journal of the National Cancer Institute*, 108(12).

[Poggio et al., 2017] Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., and Liao, Q. (2017). **Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review.** *International Journal of Automation and Computing*, 14(5):503–519.