A New Deep Learning Approach for Predicting Survival Processes in the Presence of Semi-Competing Risks

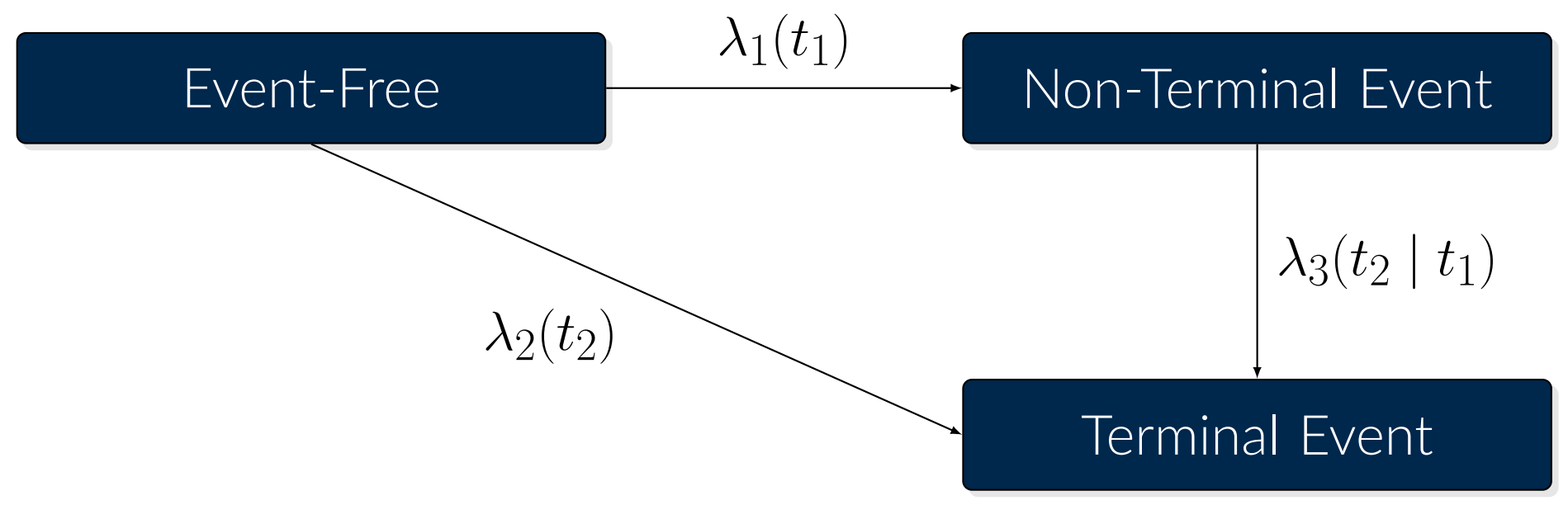
Stephen Salerno, MS [1]   Yi Li, PhD [1]

[1]Department of Biostatistics, University of Michigan

Background

Many survival processes involve a non-terminal (e.g., disease progression) and a terminal (e.g., death) event, which form a semi-competing risk relationship, i.e., the occurrence of the non-terminal event is subject to the terminal event [4]

Figure 1. Illness-Death Model Framework



- Deep learning has emerged as a powerful tool for survival prediction; however, limited work has been done to predict multi-state or competing risk outcomes, let alone semi-competing
- We propose a new deep learning framework for predicting semi-competing risk outcomes based on the illness-death model, a compartment-type model for transitions between states, which allows us to estimate patient-specific transitions and patient frailty
- As deep learning can recover non-linear risk scores, we test our method predicting simulated risk surfaces of varying complexity and exemplify our method with a real data example

Motivation

We apply our method to the Boston Lung Cancer Study Cohort, where we examine time from diagnosis to progression or death among patients with non-small cell lung cancer

- Lung cancer remains one of the leading causes of cancer-related deaths to date, with a 5-year survival rate of approximately 1 in 5 [1]
- Prognosis varies greatly and depends on several individualized risk factors including smoking status, genetic variants, and other comorbid conditions [2, 5]
- Patients diagnosed with lung cancer may experience a disease progression, go into remission, or have a recurrence prior to death

DNN-SCR Method

We can model the hazards in Figure 1 as in [6, 7], and others, by extending the Cox model to our semi-competing risks setting. We parameterize each transition in terms of a baseline hazard, a shared frailty term, and the effect of a patient's covariates, x_i:

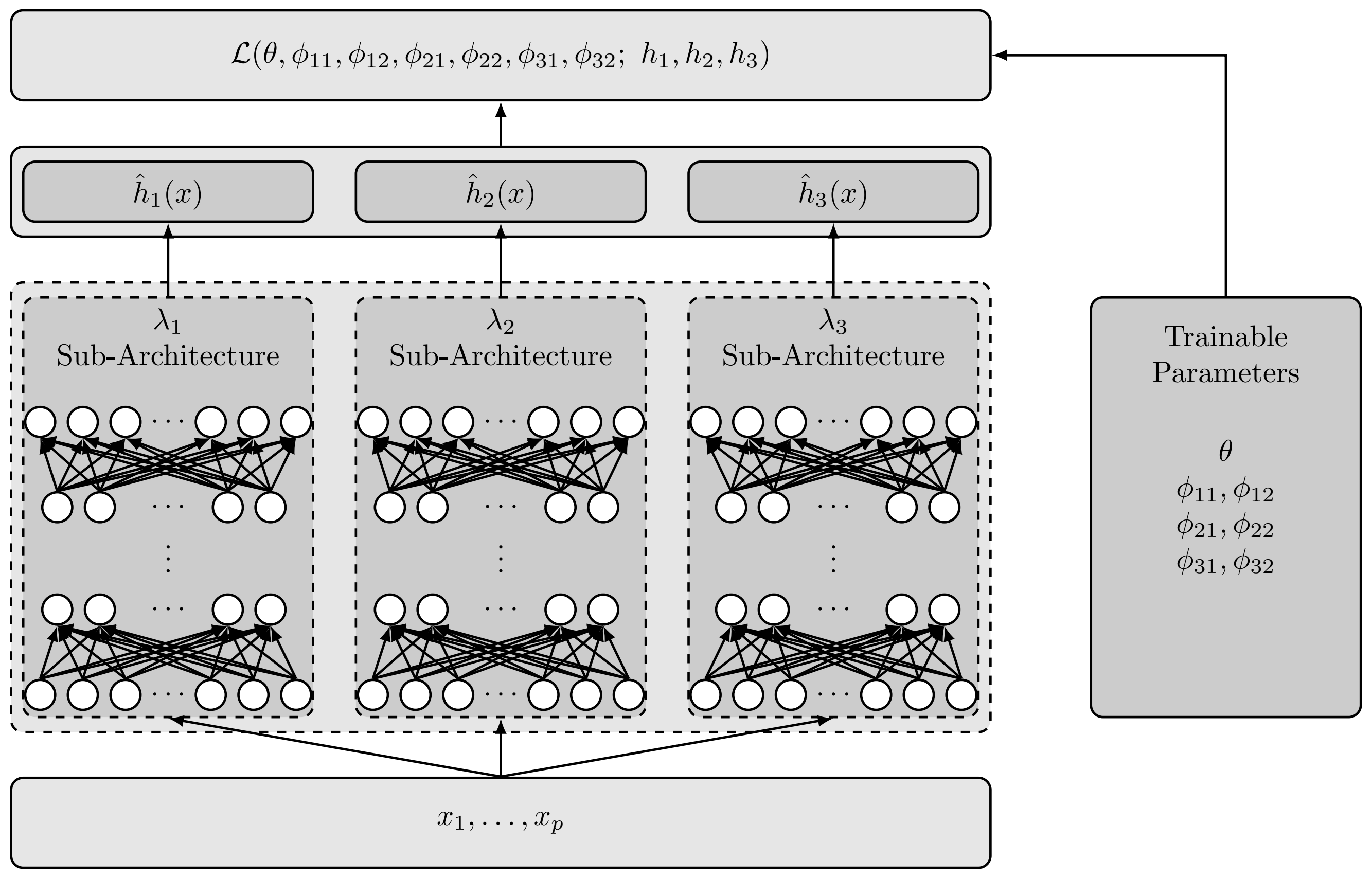
$$\begin{aligned}
\lambda_1(t_1 \mid \gamma_i, x_i) &= \gamma_i \lambda_{01}(t_1) \exp\{h_1(x_i)\}; \quad t_1 > 0 && (1)\\
\lambda_2(t_2 \mid \gamma_i, x_i) &= \gamma_i \lambda_{02}(t_2) \exp\{h_2(x_i)\}; \quad t_2 > 0 && (2)\\
\lambda_3(t_2 \mid t_1, \gamma_i, x_i) &= \gamma_i \lambda_{03}(t_2 \mid t_1) \exp\{h_3(x_i)\}; \quad 0 < t_1 < t_2 && (3)
\end{aligned}$$

We assume each γ_i are Gamma distributed with $E[\gamma_i] = 1$ and Var$(\gamma_i) = \theta$. Integrating out γ_i in the conditional likelihood of $\mathcal{D}$, we derive the following log-marginal likelihood function:

$$\begin{aligned}
\mathcal{L}(\theta,\ \phi_{11},\ \ldots,\ \phi_{32},\ h_1(\cdot),\ h_2(\cdot),\ h_3(\cdot) \mid \mathcal{D}) &= \sum_{i=1}^{N} \delta_{i1}\{\log \lambda_{01}(y_{i1}) + h_1(x_i)\}\\
&+ \delta_{i2}(1-\delta_{i1})\{\log \lambda_{02}(y_{i2}) + h_2(x_i)\} + \delta_{i1}\delta_{i2}\{\log \lambda_{03}(y_{i2} - y_{i1}) + h_3(x_i) + \log(1+\theta)\}\\
&- (\theta^{-1} + \delta_{i1} + \delta_{i2}) \log\left[1 + \theta\left\{\Lambda_{01}(y_{i1})e^{h_1(x_i)} + \Lambda_{02}(y_{i1})e^{h_2(x_i)} + \Lambda_{03}(y_{i2} - y_{i1})e^{h_3(x_i)}\right\}\right] \qquad (4)
\end{aligned}$$

We opt for a flexible, non-parametric definition of $h_1(\cdot)$, $h_2(\cdot)$, and $h_3(\cdot)$ to capture potential non-linear dependencies between the covariates by estimating $\hat{h}_g(x_i);\ g = 1,2,3$ non-parametrically as outputs from three neural network sub-architectures:

Figure 2. Schematic of DNN-SCR Architecture


$\mathcal{L}(\theta, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \phi_{31}, \phi_{32};\ h_1, h_2, h_3)$
$\hat{h}_1(x)$
$\hat{h}_2(x)$
$\hat{h}_3(x)$
λ_1 Sub-Architecture
λ_2 Sub-Architecture
λ_3 Sub-Architecture
Trainable Parameters
θ
ϕ_{11}, ϕ_{12}
ϕ_{21}, ϕ_{22}
ϕ_{31}, ϕ_{32}
$x_1, \ldots, x_p$

Implementation

We implement our approach using the `R` interface for the deep learning library **TensorFlow**, with model building and fitting done using **Keras API**:

- Taking advantage of the Keras paradigm for **progressive disclosures of complexity**, we implement our method as a custom Keras model, which has support for built-in training, evaluation, and prediction methods in a **standard, user-friendly workflow**
- Finite parameter training is done via the **GradientTape API** for automatic differentiation in a custom forward pass operation
- Thus, the user need simply **instantiate** the DNN-SCR model with the custom model wrapper function, then proceed with the **typical workflow**

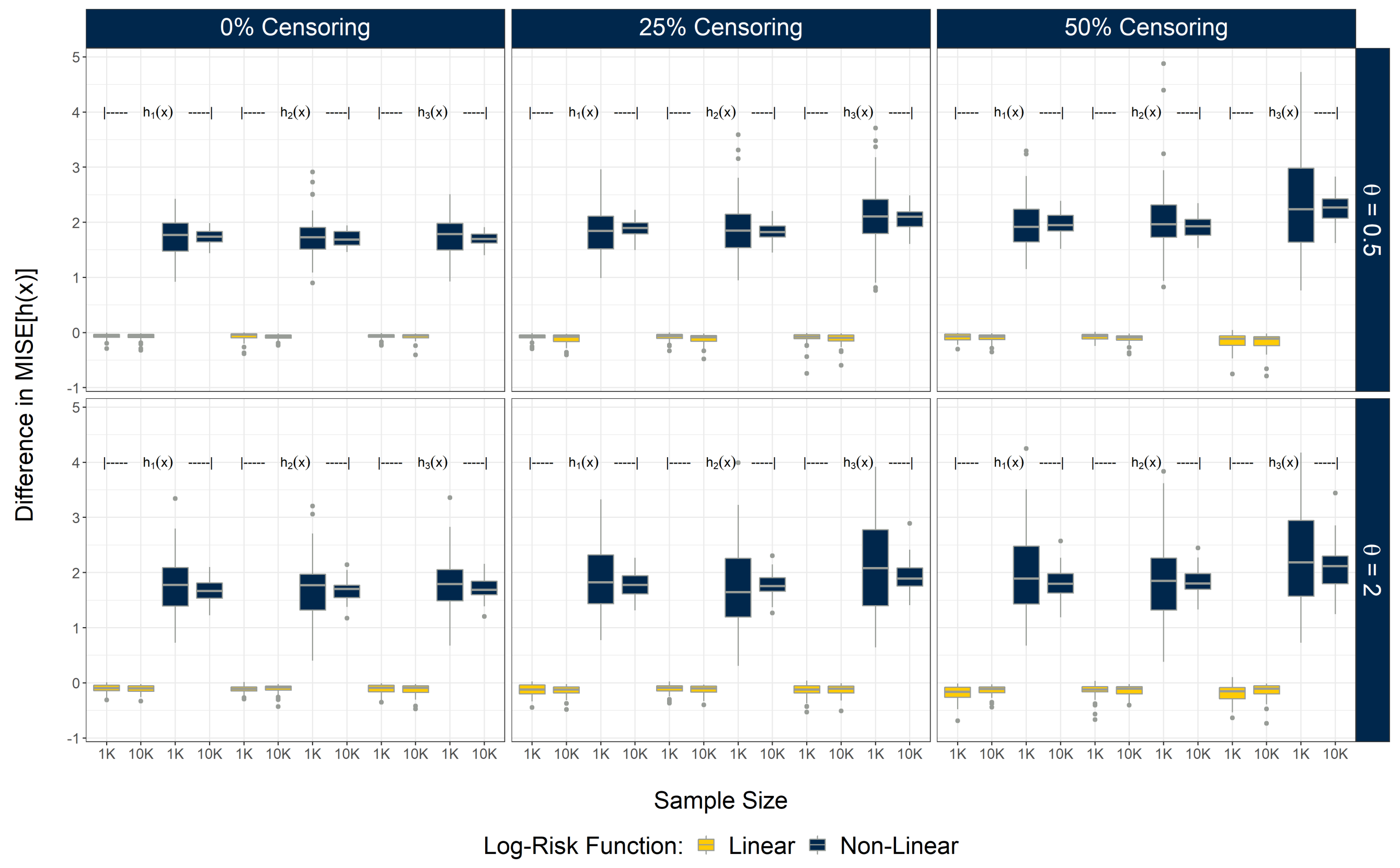
Simulation Study

We generated 50 independent datasets from (4) for each setting, with $x_i \sim N_p(0, I_p)$, varying:

- Sample sizes ($n =$ 1k, 10k)
- Frailty variances ($\theta =$ 0.5, 2)
- Censoring rates (0%, 25%, and 50%)
- Linear ($h_g(x_i) = x_i^{\mathrm{T}}\beta_g$) vs. non-linear ($h_g(x_i) = \log(|x_i|^{\mathrm{T}}\beta_g); \beta_g = 1; g = 1,2,3$) risks

Comparing our method to directly maximizing the log-likelihood function under a semi-Markov assumption with Weibull baseline hazards, Figure 3 shows differences in the mean integrated squared errors for estimating the log-risk surfaces, respectively, under true non-linear risks

Figure 3. Difference in mean integrated squared errors (MISE) of $E||\hat{h}_g - h_g||_2^2\ g = 1,2,3$ for Classical MLE - DNN-SCR



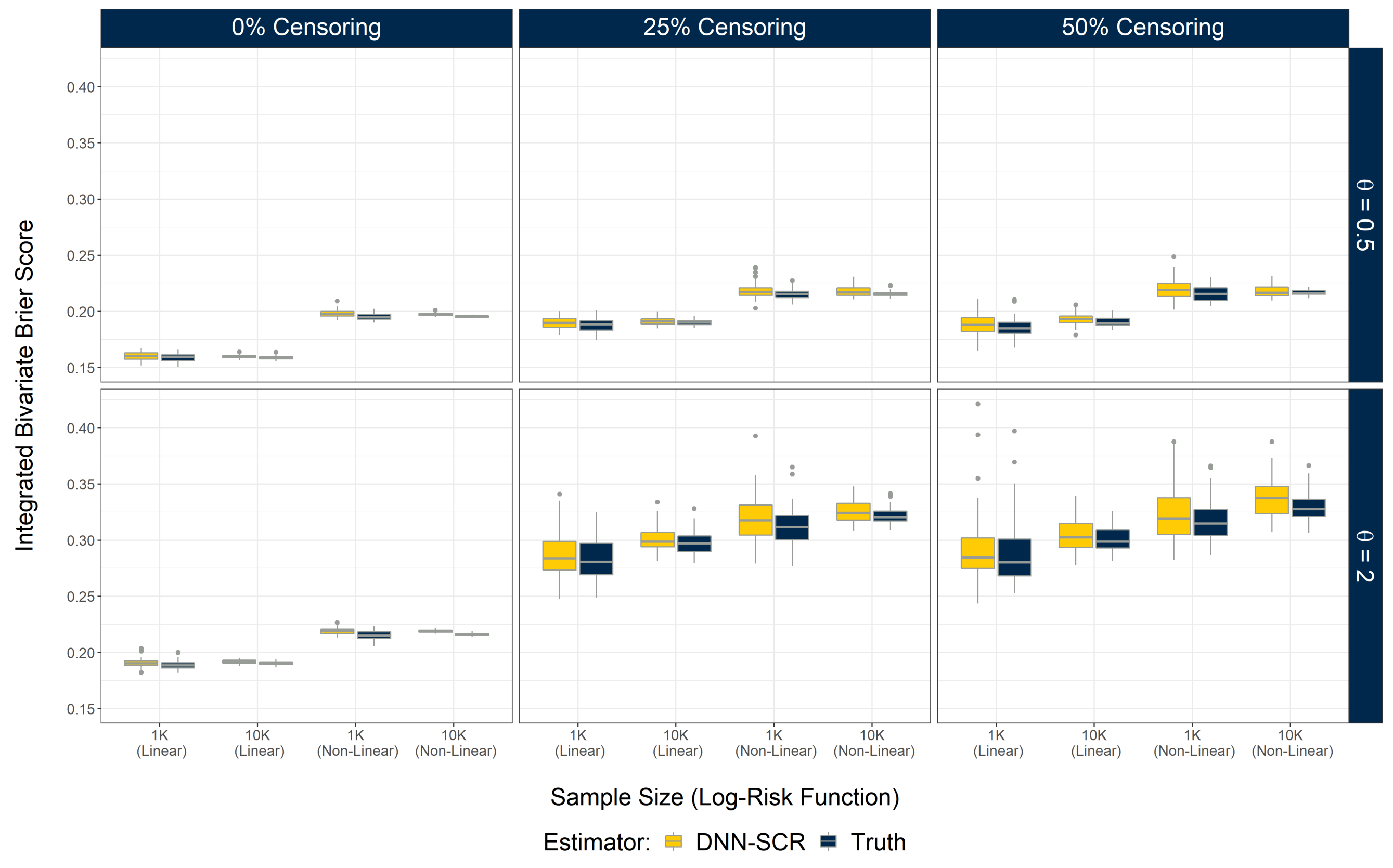
Bivariate Brier Score

As evaluating predictive performance under semi-competing risks has not yet been explored, we extend the Brier Score for right-censored data to the bivariate survival function:

$$\begin{aligned}
BBS_c = &\frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \le t,\ \delta_{i1} = 1,\ Y_{i1} \le Y_{i2}\}}{G_i(Y_{i1})}\\
&+ \frac{\pi_i(t)^2 \cdot \mathbb{I}\{Y_{i1} \le t,\ Y_{i2} \le t,\ \delta_{i1} = 0,\ \delta_{i2} = 1,\ Y_{i1} \le Y_{i2}\}}{G_i(Y_{i2})}\\
&+ \frac{[1 - \pi_i(t)]^2 \cdot \mathbb{I}\{Y_{i1} > t,\ Y_{i2} > t\}}{G_i(t)} \qquad (5)
\end{aligned}$$

We calculate the integrated Bivariate Brier Score for 1-year survival over a sequence of 100 evenly spaced time points in simulation

Figure 4. Integrated Bivariate Brier Score for DNN-SCR versus the true bivariate survival function



Boston Lung Cancer Study Cohort

Our study includes 5,296 patients with non-small cell lung cancer, diagnosed between June 1983 and October 2021 [3]. We investigate time to disease progression and death, where progression might be censored by death or the study endpoint. Included in the model are patient age at diagnosis (years), sex (0: male; 1: female), race (0: other; 1: white), ethnicity (0: non-Hispanic; 1: Hispanic), height (meters), weight (kilograms), smoking status (0: never; 1: former; 2: current), pack-years, cancer stage (1-4), and two indicators of genetic mutations (EGFR and KRAS)

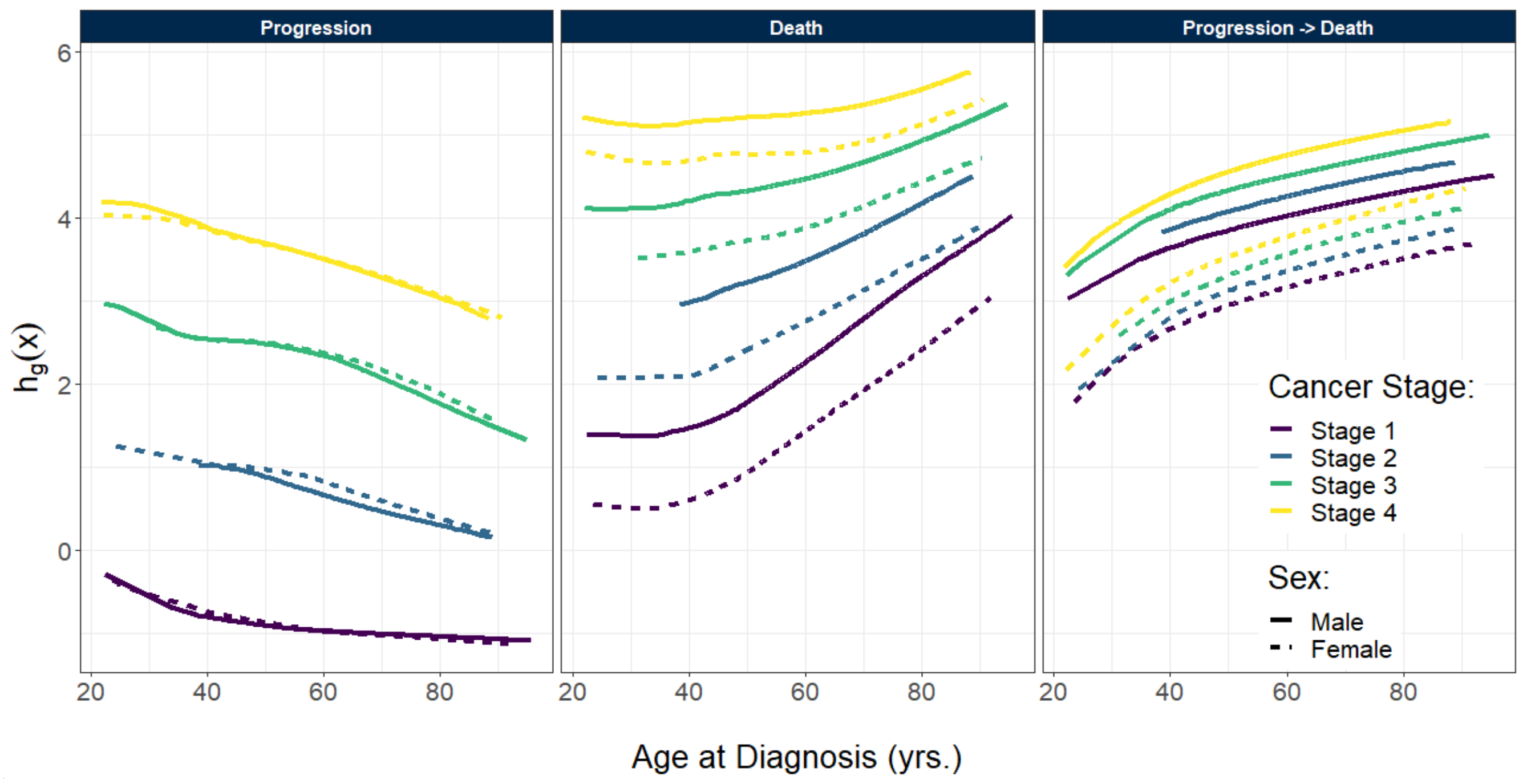
Table 1. Observed Outcomes in the BLCSC Study

	Progression	Censored
Death	111 (2%)	1,916 (36%)
Censored	224 (4%)	3,045 (58%)

We estimate the frailty variance to be 3.55, suggesting that progression is highly correlated with death. iBBS for 5-year survival was 0.178

- Below are the log risk functions for age at diagnosis on each state transition, stratified by sex and initial cancer stage while fixing the other covariates listed previously to be at their sample means or modes
- Younger age and more advanced stage is predictive of higher hazards for progression, while older age is predictive of higher hazards for mortality
- While sex does not seem to play a role in disease progression, male patients are more likely to die than female patients, and advanced stage is associated with higher hazards for all transitions

Figure 5. Hazard functions for the effect of age at diagnosis on each state transition (diagnosis to progression, diagnosis to death, and progression to death), stratified by sex and initial cancer stage



Next Steps

- Our approach fits nicely in a Bayesian paradigm, which would facilitate formulating this as a Bayesian neural network, with individualized risk prediction intervals
- Other specifications of the objective function, particularly a fully non-parametric baseline hazard, may allow for even greater prediction accuracy
- Alternatively, we can consider treating this as a classification problem, predicting survival probabilities directly with a single, sigmoidal output

Conclusions

- We have proposed a novel deep learning approach in the presence of semi-competing risks, a currently unexplored area
- Our method can recover non-linear relationships and potentially higher order interactions between disease progression, survival, and high-dimensional risk factors
- Utilizing existing paradigms for machine learning in `R`, we implement our method in a user-friendly workflow

Acknowledgements

We thank our long-term collaborator, Dr. David C. Christiani, for providing the BLCSC data and Dr. Xinan Wang for helpful discussion of the BLCSC application results. This work is partially supported by the grants from the NIH.

References

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

[2] Michael D Brundage, Diane Davies, and William J Mackillop. Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest*, 122(3):1037–1057, 2002.

[3] David C. Christiani. The Boston lung cancer survival cohort. `http://grantome.com/grant/NIH/U01-CA209414-01A1`, 2017. [Online; accessed November 27, 2018].

[4] Evelyn Fix and Jerzy Neyman. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 23(3):205–241, 1951.

[5] Laurie E Gaspar, Erica J McNamara, E Greer Gay, Joe B Putnam, Jeffrey Crawford, Roy S Herbst, and James A Bonner. Small-cell lung cancer: prognostic factors and changing treatment over 15 years. *Clinical lung cancer*, 13(2):115–122, 2012.

[6] Sebastien Haneuse and Kyu Ha Lee. Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal. *Circulation: Cardiovascular Quality and Outcomes*, 9(3):322–331, 2016.

[7] Jinfeng Xu, John D Kalbfleisch, and Beechoo Tai. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725, 2010.