# Supplementary Material: Temporal Quilting for Survival Analysis

**Changhee Lee**
UCLA

**William R. Zame**
UCLA

**Ahmed M. Alaa**
UCLA

**Mihaela van der Schaar**
University of Cambridge
UCLA
Alan Turing Institute

## 1 Computational Complexity

By following the greedy sequential approach, we have side-stepped the main challenge of scaling BO to the high dimensionality [1] by reducing the number of computations to maximize the acquisition function from $\mathcal{O}(n^{K\times|\mathcal{M}|})$ to $\mathcal{O}(K \times n^{|\mathcal{M}|})$.

To quantify the computational complexity of training Survival Quilts which can be carried out *off-line*, we first denote the computational complexity of the overall quilting pattern optimization and that of training the $m$-th baseline survival model as $\mathcal{C}_{\text{BO}}$ and $\mathcal{C}_m$ where $m \in \mathcal{M}$, respectively. Then, the computational complexity of training Survival Quilts can be given as $\mathcal{C}_{\text{BO}} + J\sum_{m\in\mathcal{M}}\mathcal{C}_m$. (Recall that $J$ is the number of cross-validations.) Albeit the increased complexity in the training due to the optimization of quilting pattern, the computational complexity of Survival Quilts for prediction – which must be carried out *on-line* – is bounded by the sum of the computational complexity of the baseline survival models in $\mathcal{M}$ for predicting the risk.

## 2 Details on the Datasets

Below we give descriptions of the six datasets we use; statistics on the time-to-event and the number of patients at risk are provided in the following figures.

**MAGGIC:** The Meta Analysis Global Group in Chronic Heart Failure (MAGGIC) performed a literature-based meta-analysis [2], which was used to investigate the clinical characteristics, treatment, and outcomes of younger patients. We extracted a random set of 5,000 patients. Among them, 1,827 (36.5%) were followed until heart failure; the remaining 3137 patients (63.5%) were right-censored. We used a total of 33 features including demographics, medical history, medical treatment, symptom status, clinical variables, and laboratory variables. Figure 1 illustrates statistics on the time-to-event and the number of patients at risk.

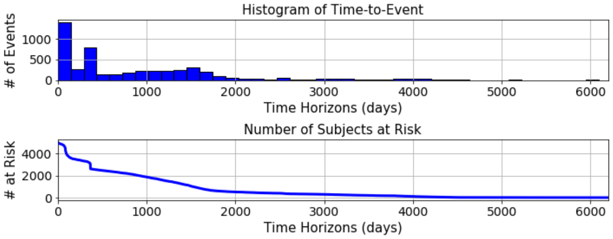


Figure 1: Statistics on the time-to-event and on the number of patients at risk for the MAGGIC dataset.

**SUPPORT:** The purpose of Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) was to improve outcomes for seriously ill hospitalized adults by improving information and decision-making [3]. We consider a total of 9,105 adults hospitalized with one or more of nine life-threatening diagnoses. Among 9,105 patients, 6,201 (68.1%) were followed until death and the remaining 2,904 (31.9%) were right-censored. We used 42 features including demographics, medical history, clinical variables, and laboratory variables. Figure 2 illustrates statistics on the time-to-event and the number of patients at risk.

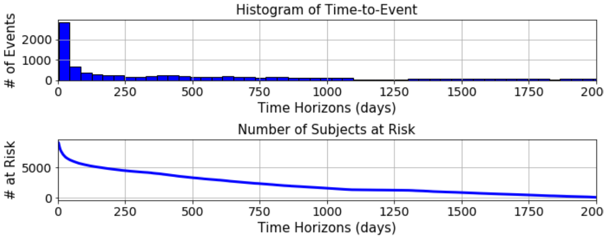


Figure 2: Statistics on the time-to-event and on the number of patients at risk for the SUPPORT dataset.

**METABRIC:** The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data contains gene expression profiles and clinical features used to determine breast cancer subgroups [4]. We consider the total of 1,981 patients in the dataset. Of the total of 1,981 patients, 888 patients (44.8%) were followed until death; the remaining 1,093 patients (55.2%) were right-censored. We used 21 publicly available clinical features including tumor size, number of positive lymph nodes, etc. Figure 3 illustrates statistics on the time-to-event and the number of patients at risk.

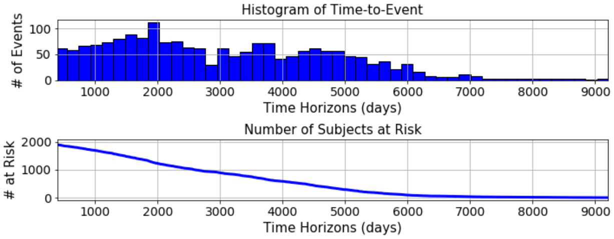


Figure 3: Statistics on the time-to-event and on the number of patients at risk for the METABRIC dataset.

**UNOS:** The United Network for Organ Sharing (UNOS) database[1] holds information on all heart transplants conducted in the US between the years 1985 to 2015. We extracted i) the entire population of 792 patients who were wait-listed to receive a transplant and received a second generation left ventricular assist device (**UNOS-I**), and ii) a randomly selected population of 5,000 patients who underwent a transplant (**UNOS-II**). For UNOS-I, of the total of 792 patients who received heart transplants, 363 (45.8%) were followed until death; the remaining 429 patients (54.2%) were right-censored. We used a total of 16 features including demographic, medical history, and lab variables. Figure 4 illustrates statistics on the time-to-event and the number of patients at risk for UNOS-I.

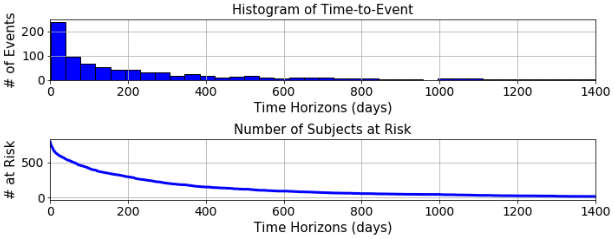


Figure 4: Statistics on the time-to-event and on the number of patients at risk for the UNOS-I dataset.

For UNOS-II, of the total of 5,000 patients who received heart transplants, 2,395 (47.9%) were followed until death; the remaining 2,605 patients (52.1%) were right-censored. We used a total of 50 features (30 recipient-relevant, 9 donor-relevant and 11 donor-recipient compatibility). Figure 5 illustrates statistics on the time-to-event and the number of patients at risk for UNOS-II.

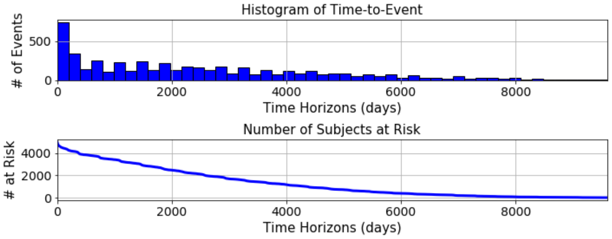


Figure 5: Statistics on the time-to-event and on the number of patients at risk for the UNOS-II dataset.

**BPD:** This was a prospective study on bipolar disorder (BPD) using primary care data collected from United Kingdom electronic health records [5]. We extracted a cohort of 2,510 patients who were diagnosed with BPD and were prescribed lithium or olanzapine as maintenance mood stabilizer treatment in the period 2000-2013. Th event is defined as assigning additional anti-psychotic or mood stabilizer or switching the mood stabilizer treatment (from lithium to olanzapine or vice versa). Among 2,510 patients, 1,999 (79.6%) were followed up until the event, and 511 (20.4%) were censored. We used 48 features including demographics, medical history, medical treatment, symptom status, clinical variables, and laboratory variables. Figure 6 illustrates statistics on the time-to-event and the number of patients at risk.

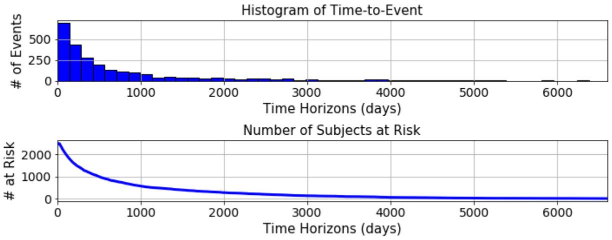


Figure 6: Statistics on the time-to-event and on the number of patients at risk for the BPD dataset.

## 3 Additional Experimental Results

In this section, we provide Figures 7 – 11 to illustrate the change in discriminative performance and the quilting pattern for the datasets that are not reported in the manuscript. For the MAGGIC dataset, Survival Quilts reduces the weight for CISF at around $t = 5000$, as the performance of CISF significantly deteriorates. For the METABRIC and UNOS-I, Survival Quilts provide the best performance over almost all the time horizons via assigning smaller weights on survival tree based models

[1] Available at `https://www.unos.org/data/`

Table 1: C-index (mean±std) for the UNOS-I dataset at different time horizons. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Best benchmark** | CISF | CISF | RSF |
| Cox | 0.624±0.02$^{\dagger}$ | 0.622±0.01$^{\dagger}$ | 0.619±0.02 |
| CoxRidge | 0.623±0.02$^{\dagger}$ | 0.619±0.01$^{\dagger}$ | 0.618±0.02 |
| Weibull | 0.624±0.02$^{\dagger}$ | 0.622±0.01$^{\dagger}$ | 0.619±0.02 |
| LogNormal | 0.637±0.02 | 0.631±0.02 | 0.626±0.03 |
| Exponential | 0.626±0.02 | 0.630±0.01 | 0.621±0.02 |
| CoxBoost | 0.620±0.02$^{\dagger}$ | 0.617±0.02$^{\dagger}$ | 0.619±0.03 |
| RSF | 0.635±0.04 | 0.649±0.03 | 0.644±0.03 |
| CISF | 0.661±0.04 | 0.657±0.03 | 0.643±0.03 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.652±0.04 | 0.653±0.03 | 0.645±0.03 |
| exog. $K=2$ | 0.654±0.03 | 0.656±0.03 | 0.645±0.03 |
| exog. $K=3$ | 0.654±0.04 | 0.654±0.04 | 0.642±0.03 |
| **endogenous** | **0.662±0.03** | **0.658±0.02** | **0.651±0.02** |

† indicates p-value < 0.05

in the later time horizons. For the UNOS-II and BPD datasets, Survival Quilts puts all the weights to the best performing benchmark, which is RSF and CoxBoost, respectively, over the time horizons, and, thus, provides the same performance as the best performing benchmark.

In Tables 1 - 3, we report the discriminative performance of the various survival models for the UNOS-I, UNOS-II, and BPD datasets at three different time horizons, representing the 25%, 50%, and 75%-quantiles of time-to-event. For the UNOS-I dataset, the endogenous construction of Survival Quilts provides the best performance because it chooses the time intervals endogenously and allows for different weights in different time intervals. For the UNOS-II and BPD datasets, the performance of Survival Quilts coincides with the best benchmark because it gives full weight to that benchmark. In the tables, we highlight in blue the results for models and time horizons in which the Brier Score constraints are satisfied. Asterisks and daggers indicate that the performance improvements of Survival Quilts are statistically significant at the 0.01 and 0.05 levels, respectively.

In Tables 4 – 9, we report the calibration performance in terms of Brier Score (lower the better) for the six real-world datasets. As seen in the tables, our method achieves the lowest Brier Score for most of the datasets over different time horizons. The values of the baseline survival models are highlighted, which satisfy the calibration constraint which is the median performance of the baseline models.

Table 2: C-index (mean±std) for the UNOS-II dataset at different time horizons. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Best benchmark** | RSF | RSF | RSF |
| Cox | 0.579±0.03$^{*}$ | 0.546±0.02$^{*}$ | 0.535±0.01$^{*}$ |
| CoxRidge | 0.577±0.03$^{*}$ | 0.545±0.02$^{*}$ | 0.534±0.01$^{*}$ |
| Weibull | 0.577±0.03$^{*}$ | 0.545±0.02$^{*}$ | 0.533±0.01$^{*}$ |
| LogNormal | 0.590±0.03$^{*}$ | 0.549±0.02$^{*}$ | 0.537±0.01$^{*}$ |
| Exponential | 0.573±0.03$^{*}$ | 0.545±0.02$^{*}$ | 0.534±0.01$^{*}$ |
| CoxBoost | 0.578±0.02$^{*}$ | 0.544±0.01$^{*}$ | 0.535±0.00$^{*}$ |
| RSF | **0.659±0.03** | **0.627±0.02** | **0.603±0.01** |
| CISF | 0.637±0.02$^{\dagger}$ | 0.589±0.01$^{*}$ | 0.569±0.01$^{*}$ |
| **Survival Quilts** | | | |
| exog. $K=1$ | **0.659±0.03** | **0.627±0.02** | **0.603±0.01** |
| exog. $K=2$ | **0.659±0.03** | **0.627±0.02** | **0.603±0.01** |
| exog. $K=3$ | **0.659±0.03** | **0.627±0.02** | **0.603±0.01** |
| **endogenous** | **0.659±0.03** | **0.627±0.02** | **0.603±0.01** |

∗ indicates p-value < 0.01
† indicates p-value < 0.05

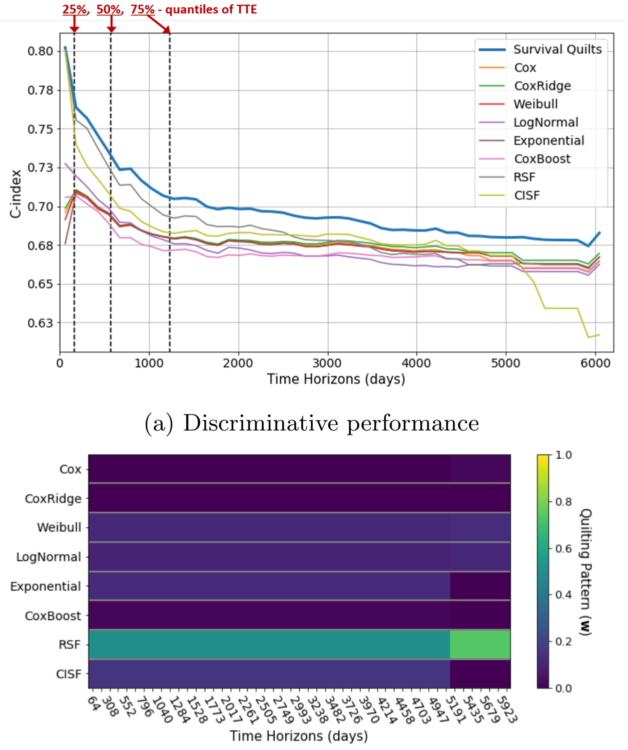
Table 3: C-index (mean±std) for the BPD dataset at different time horizons. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Best benchmark** | CoxBoost | CoxBoost | CoxBoost |
| Cox | 0.631±0.02 | 0.628±0.01 | 0.618±0.01 |
| CoxRidge | 0.632±0.02 | 0.630±0.01 | 0.620±0.01 |
| Weibull | 0.631±0.02 | 0.627±0.01 | 0.618±0.01 |
| LogNormal | 0.633±0.02 | 0.630±0.01 | 0.622±0.01 |
| Exponential | 0.631±0.02 | 0.627±0.01 | 0.617±0.01 |
| CoxBoost | **0.638±0.02** | **0.637±0.02** | **0.626±0.02** |
| RSF | 0.612±0.02 | 0.620±0.01 | 0.614±0.01 |
| CISF | 0.634±0.02 | 0.633±0.01 | 0.623±0.01 |
| **Survival Quilts** | | | |
| exog. $K=1$ | **0.638±0.02** | **0.637±0.02** | **0.626±0.02** |
| exog. $K=2$ | **0.638±0.02** | **0.637±0.02** | **0.626±0.02** |
| exog. $K=3$ | **0.638±0.02** | **0.637±0.02** | **0.626±0.02** |
| **endogenous** | **0.638±0.02** | **0.637±0.02** | **0.626±0.02** |

all p-values > 0.05

Table 4: Brier Score (mean±std) for the MAGGIC dataset at different time horizons. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.092 | 0.159 | 0.210 |
| Cox | 0.092±0.01 | 0.159±0.01 | 0.210±0.01 |
| CoxRidge | 0.092±0.01 | 0.159±0.01 | 0.210±0.01 |
| Weibull | 0.092±0.01 | 0.158±0.01 | 0.210±0.01 |
| LogNormal | 0.092±0.01 | 0.161±0.01 | 0.216±0.01 |
| Exponential | 0.092±0.01 | 0.156±0.01 | 0.208±0.01 |
| CoxBoost | 0.094±0.01 | 0.162±0.01 | 0.213±0.01 |
| RSF | 0.085±0.01 | 0.153±0.01 | 0.207±0.01 |
| CISF | 0.089±0.01 | 0.159±0.01 | 0.213±0.00 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.089±0.01 | 0.155±0.01 | 0.205±0.01 |
| exog. $K=2$ | 0.087±0.01 | 0.153±0.01 | 0.205±0.01 |
| exog. $K=3$ | 0.087±0.01 | 0.154±0.01 | 0.206±0.01 |
| **endogenous** | 0.087±0.01 | 0.153±0.01 | 0.205±0.01 |



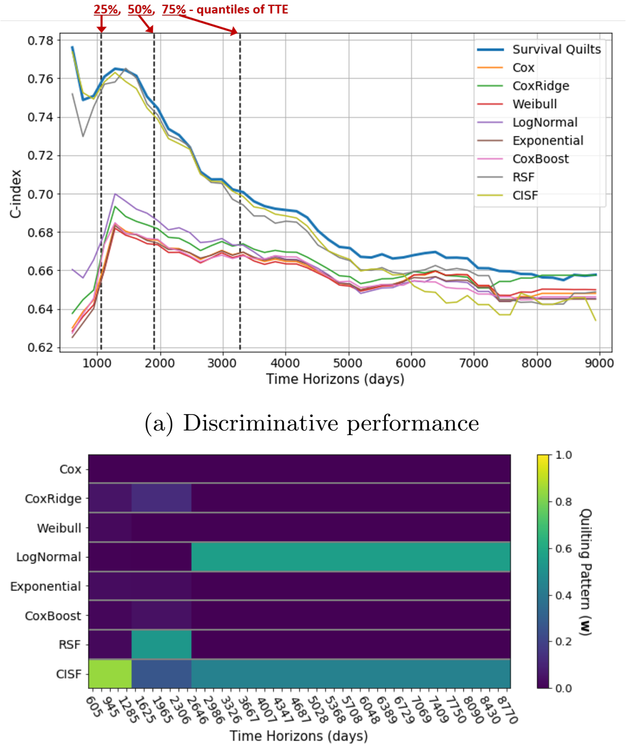
(a) Discriminative performance

(b) Quilting Pattern

Figure 7: Discriminative performance and quilting patterns over time for the MAGGIC dataset. The dotted black lines depict the 25%, 50%, and 75%-quantiles of time-to-event.

Table 5: Brier Score (mean±std) for the SUPPORT dataset. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.065 | 0.167 | 0.193 |
| Cox | 0.065±0.00 | 0.167±0.00 | 0.193±0.00 |
| CoxRidge | 0.065±0.00 | 0.167±0.00 | 0.193±0.00 |
| Weibull | 0.065±0.00 | 0.173±0.00 | 0.194±0.00 |
| LogNormal | 0.064±0.00 | 0.167±0.01 | 0.192±0.00 |
| Exponential | 0.072±0.00 | 0.217±0.01 | 0.210±0.01 |
| CoxBoost | 0.067±0.00 | 0.174±0.00 | 0.202±0.00 |
| RSF | 0.058±0.00 | 0.155±0.00 | 0.188±0.00 |
| CISF | 0.060±0.00 | 0.155±0.00 | 0.188±0.00 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.060±0.00 | 0.155±0.00 | 0.187±0.00 |
| exog. $K=2$ | 0.060±0.02 | 0.157±0.01 | 0.188±0.01 |
| exog. $K=3$ | 0.059±0.00 | 0.156±0.00 | 0.187±0.00 |
| **endogenous** | 0.059±0.00 | 0.154±0.00 | 0.188±0.00 |



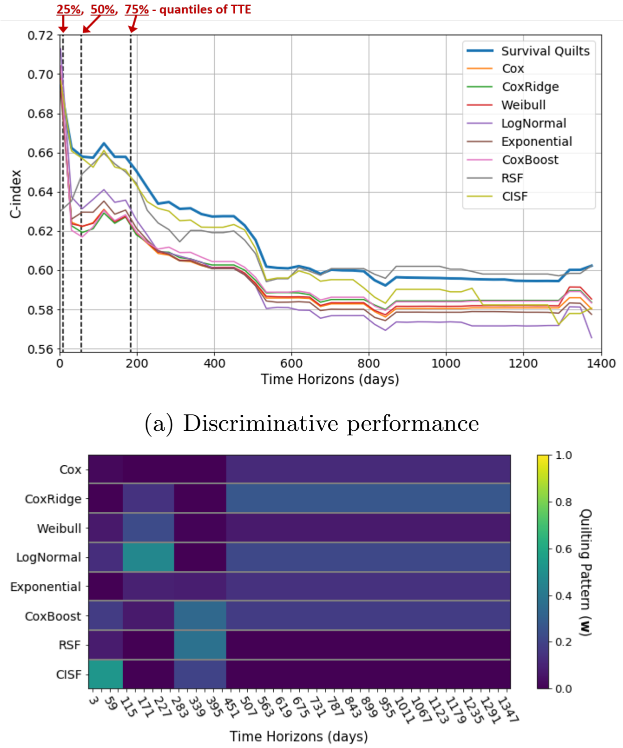
(a) Discriminative performance

(b) Quilting Pattern

Figure 8: Discriminative performance and quilting patterns over time for the METABRIC dataset. The dotted black lines depict the 25%, 50%, and 75%-quantiles of time-to-event.

Table 6: Brier Score (mean±std) for the METABRIC dataset. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.102 | 0.163 | 0.204 |
| Cox | 0.102±0.01 | 0.164±0.01 | 0.202±0.01 |
| CoxRidge | 0.101±0.01 | 0.161±0.00 | 0.198±0.01 |
| Weibull | 0.103±0.01 | 0.164±0.01 | 0.201±0.01 |
| LogNormal | 0.102±0.01 | 0.163±0.01 | 0.204±0.00 |
| Exponential | 0.105±0.01 | 0.165±0.00 | 0.205±0.00 |
| CoxBoost | 0.101±0.01 | 0.164±0.01 | 0.205±0.00 |
| RSF | 0.095±0.01 | 0.152±0.01 | 0.203±0.01 |
| CISF | 0.097±0.01 | 0.158±0.01 | 0.205±0.00 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.096±0.01 | 0.155±0.01 | 0.201±0.00 |
| exog. $K=2$ | 0.096±0.01 | 0.155±0.01 | 0.201±0.00 |
| exog. $K=3$ | 0.097±0.01 | 0.155±0.01 | 0.204±0.01 |
| **endogenous** | 0.096±0.01 | 0.155±0.01 | 0.201±0.01 |



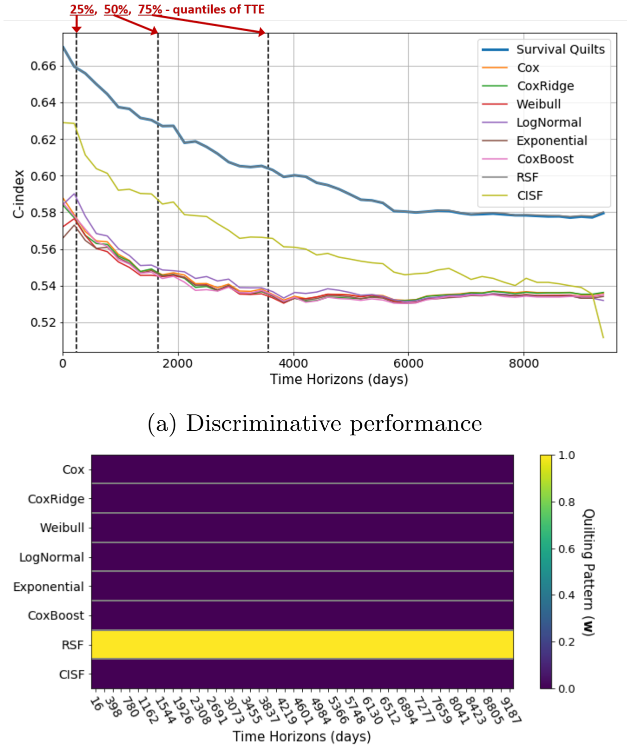
(a) Discriminative performance

(b) Quilting Pattern

Figure 9: Discriminative performance and quilting patterns over time for the UNOS-I dataset. The dotted black lines depict the 25%, 50%, and 75%-quantiles of time-to-event.

Table 7: Brier Score (mean±std) for the UNOS-I dataset. Blue highlighting indicates that the Brier Score constraints are satisfied.

| Models | Time-Horizons (quantiles) | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.158 | 0.179 | 0.226 |
| Cox | 0.157±0.02 | 0.178±0.02 | 0.227±0.01 |
| CoxRidge | 0.157±0.02 | 0.179±0.02 | 0.226±0.01 |
| Weibull | 0.159±0.02 | 0.179±0.02 | 0.227±0.01 |
| LogNormal | 0.157±0.02 | 0.178±0.02 | 0.225±0.01 |
| Exponential | 0.177±0.02 | 0.195±0.03 | 0.233±0.02 |
| CoxBoost | 0.158±0.02 | 0.181±0.02 | 0.229±0.01 |
| RSF | 0.158±0.02 | 0.175±0.02 | 0.222±0.01 |
| CISF | 0.153±0.02 | 0.173±0.02 | 0.217±0.02 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.154±0.02 | 0.173±0.02 | 0.219±0.01 |
| exog. $K=2$ | 0.155±0.02 | 0.174±0.02 | 0.218±0.01 |
| exog. $K=3$ | 0.154±0.02 | 0.174±0.02 | 0.220±0.01 |
| **endogenous** | 0.153±0.02 | 0.173±0.02 | 0.217±0.01 |



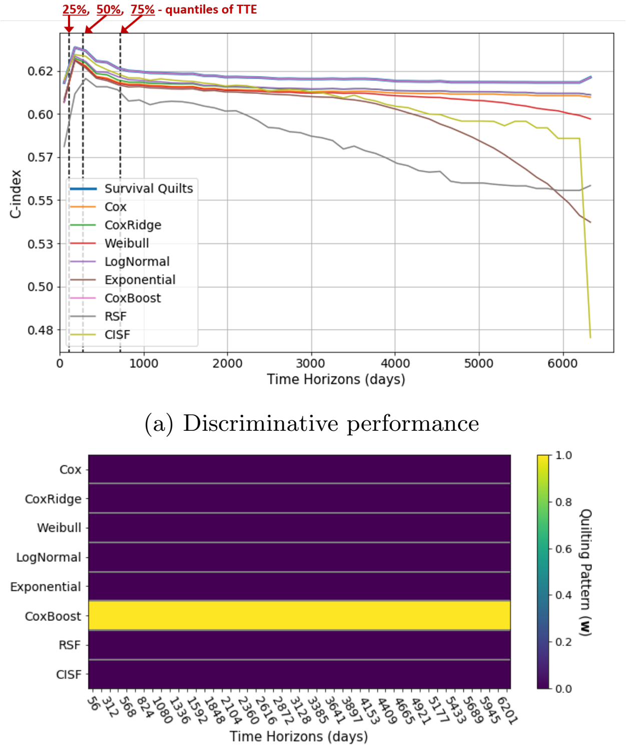
(a) Discriminative performance

(b) Quilting Pattern

Figure 10: Discriminative performance and quilting patterns over time for the UNOS-II dataset. The dotted black lines depict the 25%, 50%, and 75%-quantiles of time-to-event.

Table 8: Brier Score (mean±std) for the UNOS-II dataset. Blue highlighting indicates that the Brier Score constraints are satisfied.

| Models | Time-Horizons (quantiles) | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.103 | 0.196 | 0.247 |
| Cox | 0.102±0.01 | 0.197±0.00 | 0.249±0.01 |
| CoxRidge | 0.102±0.01 | 0.196±0.00 | 0.247±0.01 |
| Weibull | 0.103±0.01 | 0.204±0.00 | 0.253±0.01 |
| LogNormal | 0.103±0.01 | 0.209±0.00 | 0.250±0.01 |
| Exponential | 0.108±0.01 | 0.197±0.00 | 0.250±0.01 |
| CoxBoost | 0.103±0.01 | 0.195±0.00 | 0.244±0.01 |
| RSF | 0.097±0.00 | 0.185±0.00 | 0.235±0.01 |
| CISF | 0.101±0.01 | 0.192±0.00 | 0.241±0.01 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.097±0.00 | 0.185±0.00 | 0.235±0.01 |
| exog. $K=2$ | 0.097±0.00 | 0.185±0.00 | 0.235±0.01 |
| exog. $K=3$ | 0.097±0.00 | 0.185±0.00 | 0.235±0.01 |
| **endogenous** | 0.097±0.00 | 0.185±0.00 | 0.235±0.01 |



(a) Discriminative performance

(b) Quilting Pattern

Figure 11: Discriminative performance and quilting patterns over time for the BPD dataset. The dotted black lines depict the 25%, 50%, and 75%-quantiles of time-to-event.

Table 9: Brier Score (mean±std) for the BPD dataset. Blue highlighting indicates that the Brier Score constraints are satisfied.

| **Models** | **Time-Horizons (quantiles)** | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Median** | 0.205 | 0.229 | 0.215 |
| Cox | 0.203±0.01 | 0.227±0.01 | 0.216±0.01 |
| CoxRidge | 0.203±0.01 | 0.227±0.01 | 0.215±0.01 |
| Weibull | 0.206±0.01 | 0.231±0.01 | 0.219±0.01 |
| LogNormal | 0.204±0.01 | 0.228±0.01 | 0.214±0.01 |
| Exponential | 0.215±0.02 | 0.241±0.01 | 0.225±0.01 |
| CoxBoost | 0.202±0.01 | 0.226±0.01 | 0.212±0.01 |
| RSF | 0.209±0.01 | 0.230±0.00 | 0.215±0.01 |
| CISF | 0.205±0.01 | 0.229±0.01 | 0.215±0.01 |
| **Survival Quilts** | | | |
| exog. $K=1$ | 0.202±0.01 | 0.226±0.01 | 0.212±0.01 |
| exog. $K=2$ | 0.202±0.01 | 0.226±0.01 | 0.212±0.01 |
| exog. $K=3$ | 0.202±0.01 | 0.226±0.01 | 0.212±0.01 |
| **endogenous** | 0.202±0.01 | 0.226±0.01 | 0.212±0.01 |

## References

[1] Kirthevasan Kandasamy, Jeff Schneider, and Barnabás Póczos. High dimensional bayesian optimisation and bandits via additive models. *In Proceedings of the 32th International Conference on Machine Learning (ICML 2015)*, 2015.

[2] Chih M. Wong, Nathaniel M. Hawkins, Mark C. Petrie, Pardeep S. Jhund, Roy S. Gardner, Cono A. Ariti, Katrina K. Poppe, Nikki Earle, Gillian A. Whalley, Iain B. Squire, Robert N. Doughty, and John J.V. McMurray. Heart failure in younger patients: the meta-analysis global group in chronic heart failure (MAGGIC). *European Heart Journal*, 35(39):2714–2721, June 2014.

[3] William A. Knaus, Frank E. Harrell, Joanne Lynn, Lee Goldman, Russell S. Phillips, Alfred F. Connors, Neal V. Dawson, William J. Fulkerson, Robert M. Califf, Norman Desbiens, Peter Layde, Robert K. Oye, Paul E. Bellamy, Rosemarie B. Hakim, and Douglas P. Wagner. The SUPPORT prognostic model. objective estimates of survival for seriously ill hospitalized adults. study to understand prognoses and preferences for outcomes and risks of treatments. *Annals of Internal Medicine*, 122(3): 191–203, February 1995.

[4] Christina Curtis, Sohrab P. Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M. Rueda, Mark J. Dunning, Doug Speed, Andy G. Lynch, Shamith Samarajiwa, Yinyin Yuan, Stefan Gräf, Gavin Ha, Gholamreza Haffari, Ali Bashashati, Roslin Russell, Steven McKinney, METABRIC Group, Anita Langerød, Andrew Green, Elena Provenzano, Gordon Wishart, Sarah Pinder, Peter Watson, Florian Markowetz, Leigh Murphy, Ian Ellis, Arnie Purushotham, Anne-Lise Børresen-Dale, James D. Brenton, Simon Tavaré, Carlos Caldas, and Samuel Aparicio. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486:346–352, June 2012.

[5] Joseph F. Hayes, Louise Marston, Kate Walters, John R. Geddes, Michael King, and David P. J. Osborn. Adverse renal, endocrine, hepatic, and metabolic events during maintenance mood stabilizer treatment for bipolar disorder: A population-based cohort study. *PLOS Medicine*, 13(8):e1002058, August 2016.