Equitable Survival Prediction: A Fairness-Aware Survival Modeling (FASM) Approach

m.liu@u.nus.edu

Mingxuan Liu, MSc[1], Yilin Ning, PhD[1], Haoyuan Wang, MSc[2], Chuan Hong, PhD[2], Mattew Engelhard, PhD[2], Danielle S. Bitterman, MD[3], William G. La Cava, PhD[4], Nan Liu, PhD[1]

[1]Duke-NUS Medical School, National University of Singapore, Singapore; [2]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA; [3]Artificial Intelligence in Medicine Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA; [4]Computational Health Informatics Program, Boston Children's Hospital; Department of Pediatrics, Harvard Medical School, Boston, MA, USA

1. Background & Objective

Background

- As machine learning gains prominence in high-stakes fields like healthcare, concerns about fairness have grown.
- In survival analysis, censoring and time dynamics can further add complexity to fair model development.
- existing fairness approaches often focus on intra-group metrics and overlook disparities in cross-group risk rankings, e.g., high-risk Black patients may be ranked below lower-risk White patients who do not experience the event of mortality.

Objective

- To develop a fairness-aware survival modeling framework, mitigating algorithmic bias regarding both intra—group and cross-group risk rankings over time.

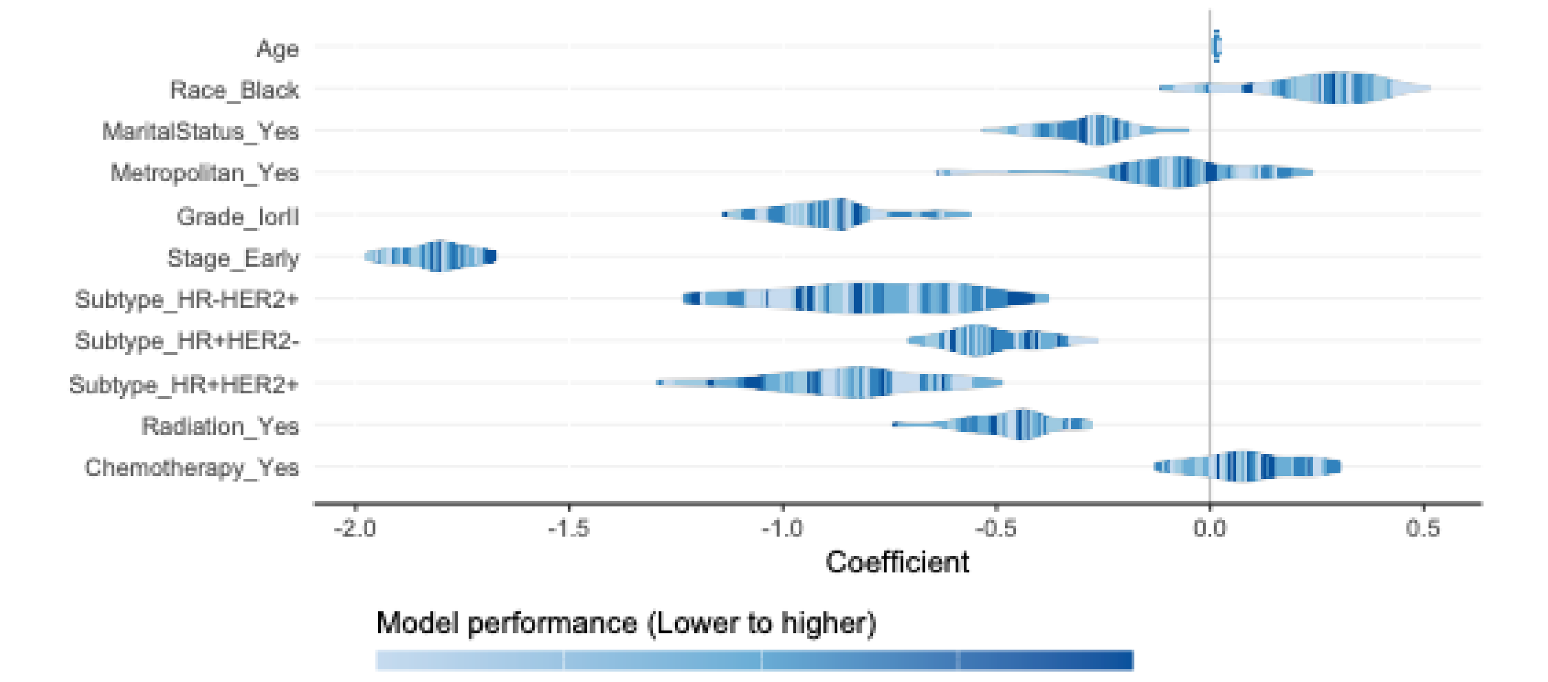
2. Method

❖ **Framework of FASM**

- **Generating nearly-optimal models:**
  - Rashomon Set (RS) of nearly optimal models, extending from "near-optimality"[1], accounting for exclusion cases of sensitive variables[1]

$$R_{(U,S)}\left(\epsilon_0, \beta^*_{(U,S)}, \mathrm{B}_{(U,S)}\right) = \left\{\beta \in \mathrm{B}_{(U,S)} \mid R^2_L\left(f_\beta, Y\right) \geq (1-\epsilon_0) R^2_L\left(f^*_{(U,S)}, Y\right)\right\}$$

Figure 1. Coefficients of nearly-optimal models within the Rashomon set



- **Model selection with fairness**
  - Model Selection Index (MRI), to comprehensively measure model fairness, encompassing both intra- and cross-group fairness metrics and **yielding a fairness-aware model** with the highest ranking.

Performance metric	Description	Fairness metric	Description
C–index	Overall ranking ability: measures whether an individual with an earlier event is assigned a higher risk score than one with a later event or not at all, over the entire follow-up period.	ΔCI	Maximum discrepancy in C–index across subgroups.
$iAUC$	Time-integrated discrimination: averaging time-specific discrimination $AUC(t)$ across the follow-up period.	$\Delta iAUC$	Maximum discrepancy in $iAUC$ across subgroups.
$xCI_{(a,b)}$	Overall cross-group ranking: measures whether an individual i in subgroup a with an earlier event is ranked above individual j in subgroup b with a later event or not at all, over the entire follow-up time period.	ΔxCI	Maximum absolute difference in reciprocal xCI values (e.g., $xCI_{(a,b)}$ and $xCI_{(b,a)}$) across all subgroup pairs.
$xAUC_{(a,b)}(t)$	Time-specific cross-group discrimination: measures whether an individual i in subgroup a with an event before time t is ranked above individual j in subgroup b with an event after time t or not at all.	$\Delta xAUC(t)$	Maximum absolute difference in reciprocal $xAUC$ values at time t across all subgroup pairs.
		$i\Delta xAUC$	Time-integrated disparity, averaging disparities measured by $\Delta xAUC(t)$ across the follow-up period.

❖ **Breast cancer cohort**

- **Historical disparities**
  Despite advances in screening, diagnosis, and treatment for breast cancer, substantial disparities persist in breast cancer outcomes across sociodemographic groups.[2]
- **Dataset**
  invasive breast cancer female patients from Surveillance, Epidemiology, and End Results (SEER) dataset.

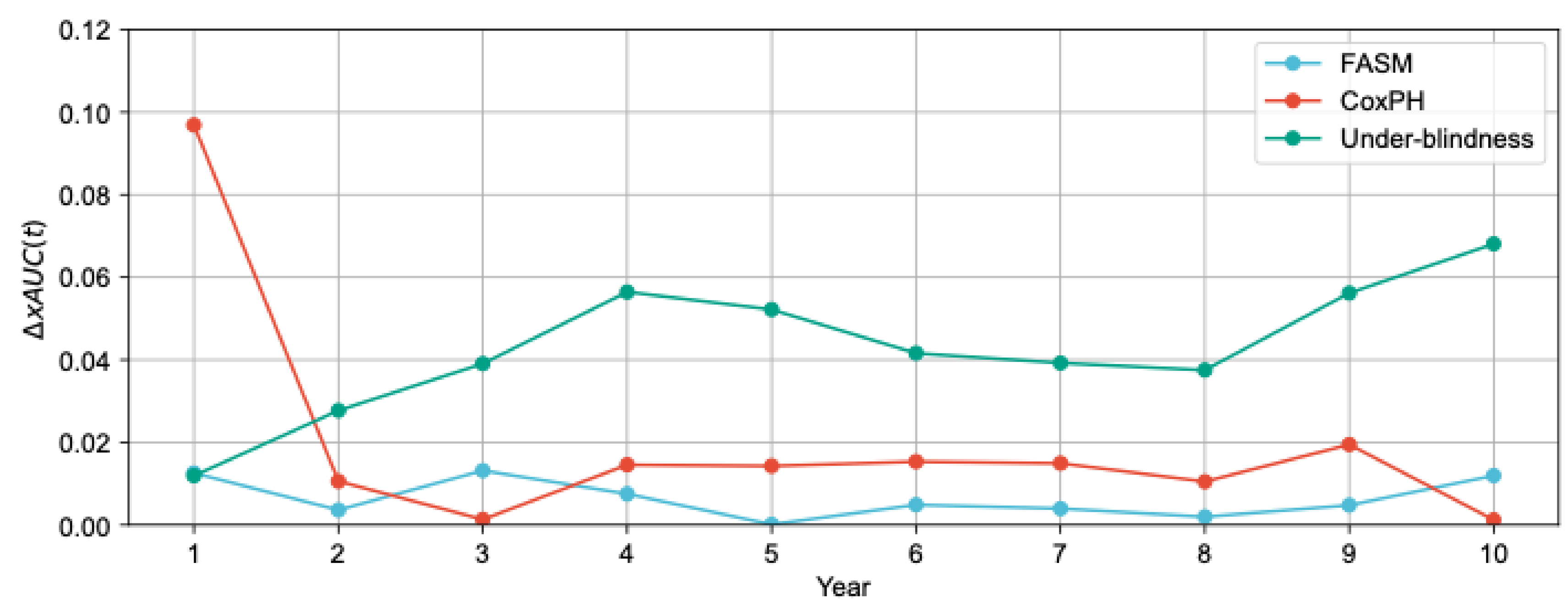
3. Result

Table 1. Two-dimension evaluation (fairness and performance)

	Fairness metrics ↓				Performance metrics ↑	
	$\Delta iAUC$	ΔCI	ΔxCI	$i\Delta xauc$	$iAUC$	CI
CoxPH	0.006	0.016	0.261	0.016	0.833 [0.821, 0.843]	0.766 [0.753, 0.778]
Under-blindness	0.007	0.016	0.163	0.043	0.833 [0.818, 0.845]	0.765 [0.750, 0.779]
FASM	**0.003**	**0.013**	**0.132**	**0.006**	**0.827 [0.815, 0.839]**	**0.758 [0.745, 0.771]**

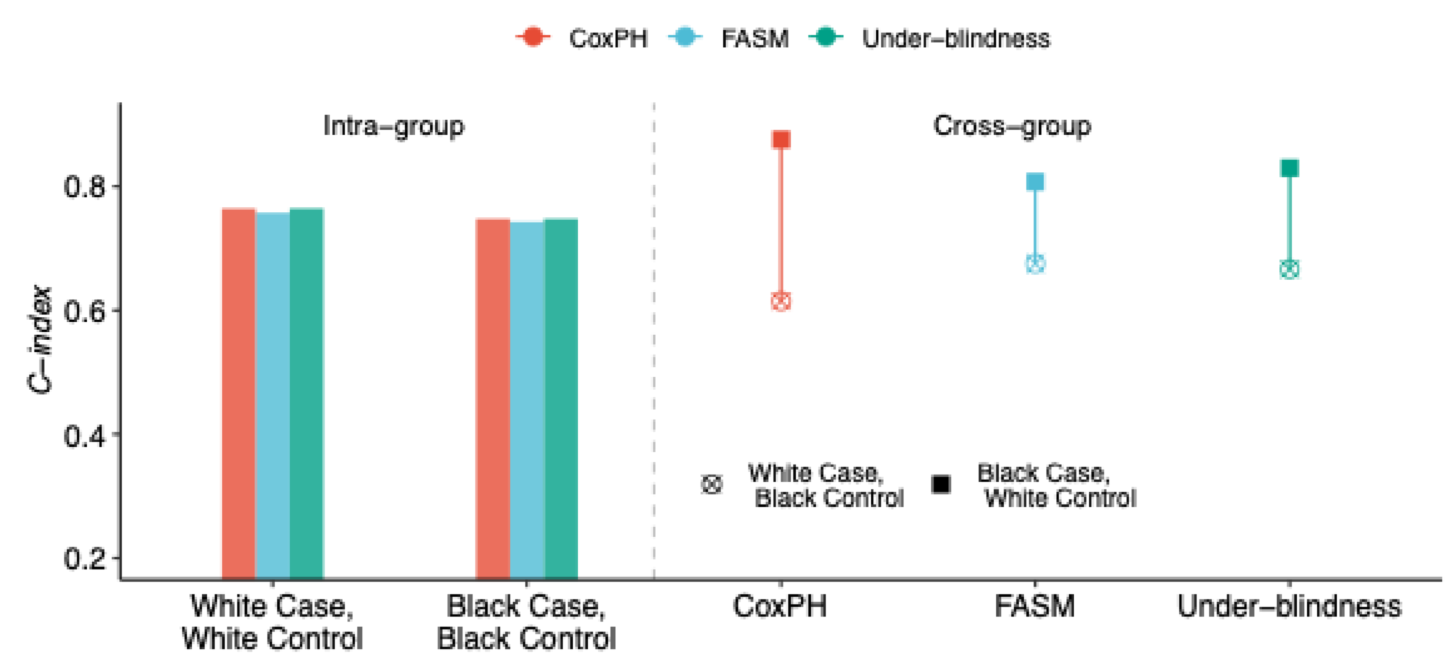
- The FASM model demonstrated the best overall fairness, with the minimal intra-group ranking bias and the smallest cross-group bias.

Figure 2. Comparison of time-dependent cross-group fairness disparities over a 10-year follow-up period across three models.



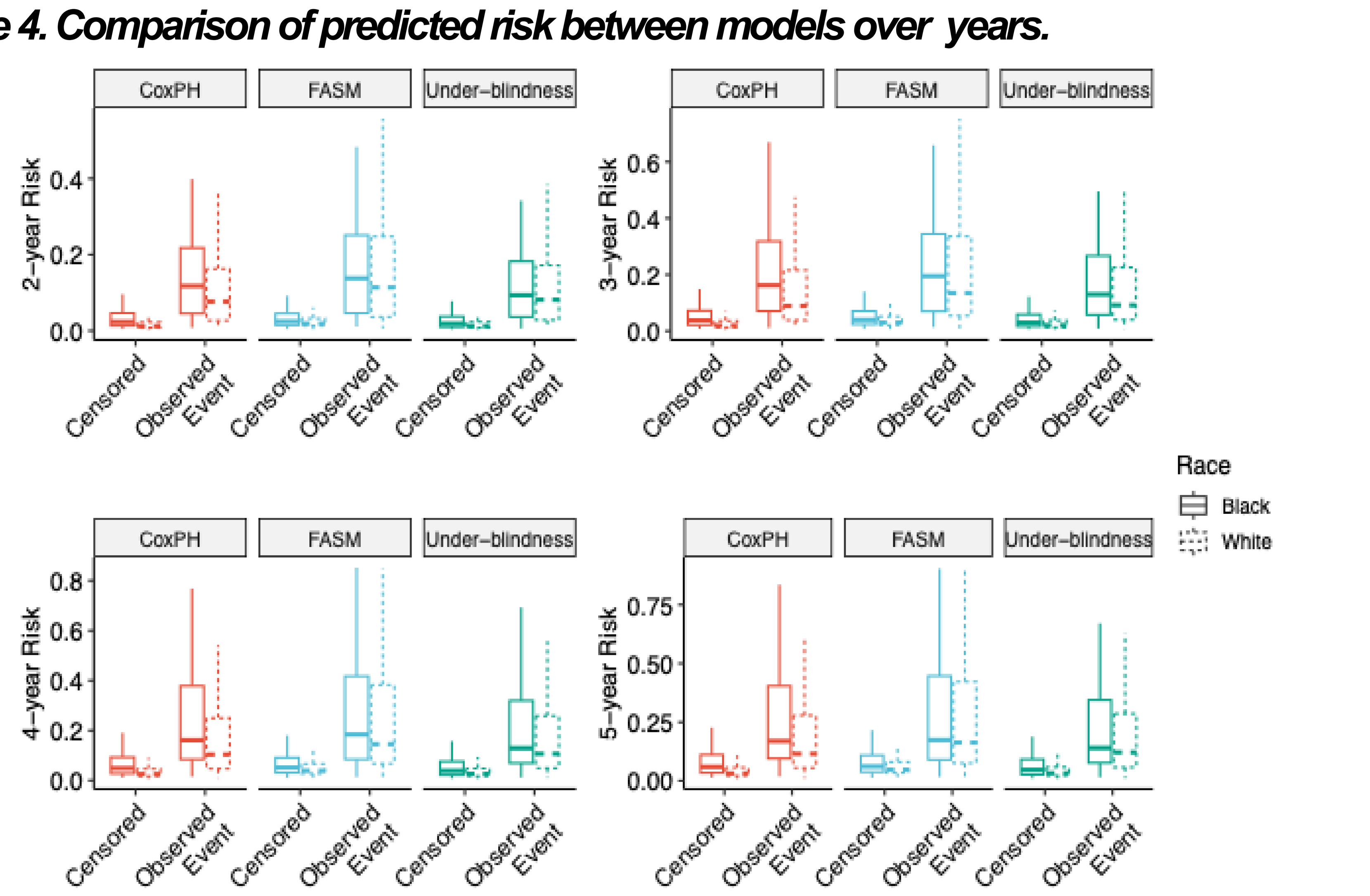
- The FASM model consistently maintains low and stable cross-group ranking disparities over time, with $\Delta xAUC(t)$ remaining below 0.02 across all time points.
- In contrast, the CoxPH model exhibits a sharp disparity in the first year. The Under-blindness model exhibits worsening disparities over time.

Figure 3. Comparison of intra-group and cross-group C–index across models.



- All three models showed comparable performance for both White and Black subgroups.
- The CoxPH model demonstrated marked directional bias, with much lower C–index when comparing White cases to Black controls and disproportionately higher C–index for Black cases versus White controls.
- FASM notably reduced this cross-group ranking disparity, improving the fairness of risk rankings across racial subgroups without compromising intra-group predictive performance.

Figure 4. Comparison of predicted risk between models over years.



- FASM successfully reduced this discrepancy, yielding similar risk predictions across racial groups while maintaining the distinction between censored and event cases.

4. Conclusion & Future work

- The interpretable framework of FASM can produce a fairness-aware model, which can
  - maintain the model performance achieved by the fairness-unaware model, i.e., the standardized CoxPH model
  - improve model fairness in terms of both intra- and cross-group fairness over time
- **Future work**
  - To extend FASM to other interpretable machine learning models
  - To validate in various clinical scenarios

Reference

1. Dong J, Rudin C. Exploring the cloud of variable importance for the set of all good models. Nature Machine Intelligence 2020;2(12):810-824.
2. Jatoi I, Sung H, Jemal A. The Emergence of the Racial Disparity in U.S. Breast-Cancer Mortality. New England Journal of Medicine. 2022;386(25):2349-2352.