Understanding Algorithmic Fairness in Health Care: A Proposed Case Study with Three Datasets

Bruna Seewald   Flavio Figueiredo

UFMG

Abstract

Research of algorithmic fairness in machine learning (ML) has widely focused on datasets that explore criminal cases or credit loans. In this work, we present a work in progress that aims to take advantage of advances in health care and characterize ML fairness in-depth for the health domain. Our case study will focus on three datasets, including one large dataset from Brazil, and we intend to raise fairness concerns in these databases as well as identify misdiagnosis in patients with protected attributes. Finally, we aim at proposing solutions (in partnership with health care professionals) to mitigate problems related to ML unfairness.

What is Fairness?

Notation

- Ground truth: $\mathcal{D} = \{(y_i, x_i, \delta_i)\}$
- Hypothesis: $\hat{t}_i = h_\Theta(x_i)$, where $\hat{t}_i \in [0, \infty)$
- Prediction: $\hat{y}_i = \begin{cases} \hat{t}_i & \text{if } \delta_i = 1 \\ c_i & \text{if } \delta_i = 0 \end{cases}$

Random Variables

- $X \rightarrow$ Features from individuals: $P(X = x_i)$
- $G \rightarrow$ A binary feature capturing sensitive groups: $P(G = g)$
- $L \rightarrow$ Legitimate factors: $P(L = l)$
- $S \rightarrow$ Time to event $P(T = t)$
- $Y \rightarrow$ Responses: $P(Y = y_i)$

Some definitions

- **Disparate Treatment**: Individuals from a sensitive group must be treated equally.
  - **Group fairness** [3]: The fraction of positive predictions is equal in both groups.
    - $P(T \geq t | G = g_a) = P(T \geq t | G = g_b)$
  - **Conditional Statistical Parity** [2]: The fraction of positive predictions is equal in both groups controlled by legitimated factors.
    - $P(T \geq t | L = l, G = g_a) = P(T \geq t | L = l, G = g_b)$
- **Disparate Impact**: The impact of decisions must affect groups equally
  - **Equal opportunity error**[1]: A case when the estimated error to the time of the event of interest is equal in both groups.
    - $median(\hat{y}_i - y_i \,|\, G = g_a) = median(\hat{y}_i - y_i \,|\, G = g_b)$

Datasets

- MIMIC-III database [4]
- eICU Collaborative Research Database [5] (ongoing)
- Telehealth Network of Minas Gerais Database [1] (ongoing)

Exploratory analysis

In the first part of this **work in progress**, we are exploring the **MIMIC-III database**, which is a intensive care unit database. Our cohort:

- 3 groups of ICD (International Classification of Diseases)
- Among top 10 causes of death in high-income countries

		Grouped by Ethnicity			
		asian	black	hispanic	white
n		504	2367	738	13548
Gender	F	197 (39.1)	1326 (56.0)	302 (40.9)	5693 (42.0)
	M	307 (60.9)	1041 (44.0)	436 (59.1)	7855 (58.0)
Insurance	Private	100 (19.8)	467 (19.7)	129 (17.5)	3960 (29.2)
	Public	403 (80.0)	1887 (79.7)	599 (81.2)	9553 (70.5)
	Self-Pay	1 (0.2)	13 (0.5)	10 (1.4)	35 (0.3)
Length of Stay		7.5 [4.2,13.0]	7.0 [4.0,12.0]	6.5 [4.1,12.0]	7.5 [4.6,12.7]
Age		66.4 [56.5,76.8]	63.9 [53.9,73.1]	60.2 [50.0,70.5]	68.6 [58.8,77.8]

Figure 1: MIMIC-III Cohort characteristics

Survival Analysis

An important subfield of statistics where the goal is to analyze and model the data and the outcome is the time until the occurrence of an event of interest.

Basic Cox Model

The most used survival analysis method when considering a broader range of conditions. In this model the log-hazard of an individual is a linear function of their static covariates and a population-level baseline hazard that changes over time. Proportional hazard assumption:

$$\lambda(t \mid X_i) = \lambda_0 \exp\Big(\sum_i x_i \beta_i\Big)$$

x_i are patient features: gender, race, disease, emergency, insurance, age and oasis score.

[1] In a classifier task, this is also knows as the **false negative error rate balance**

First Results

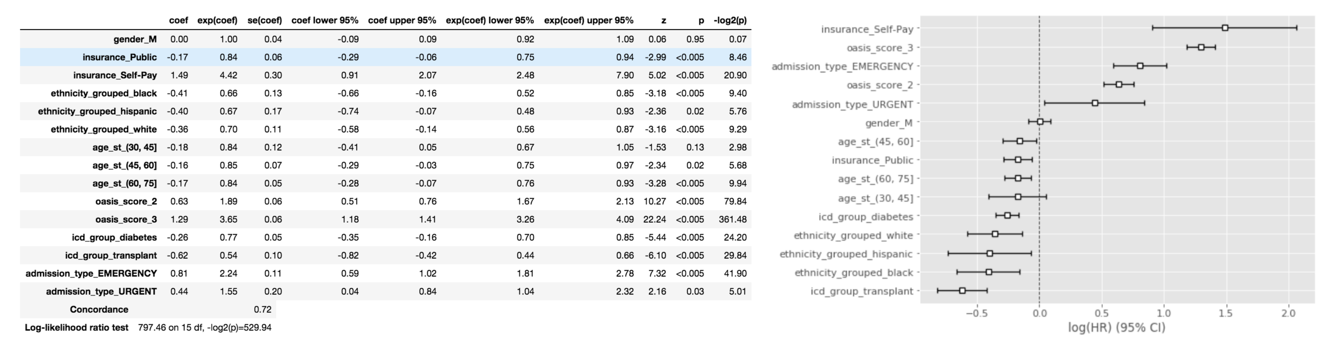

Figure 2: Model and coefficients

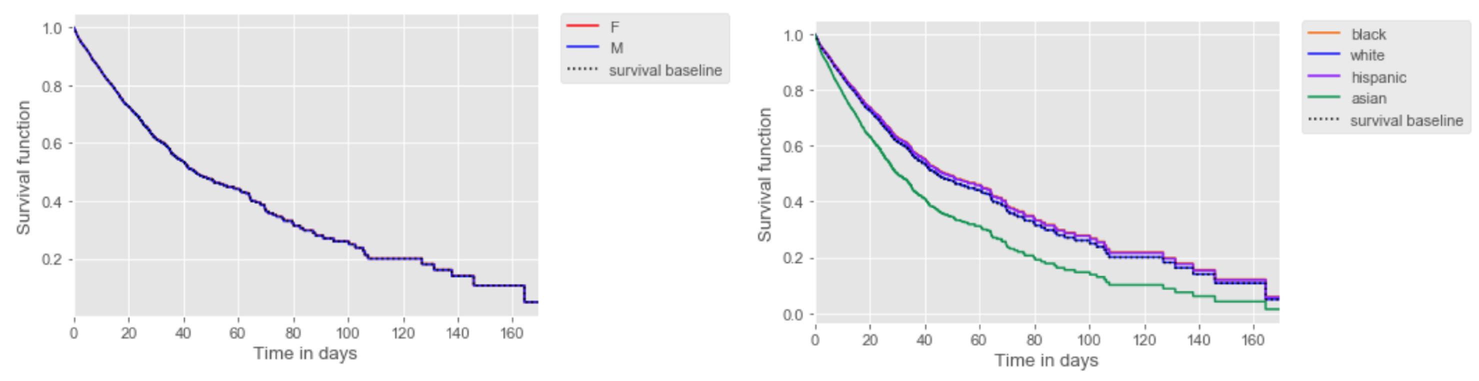

Figure 3: Survival analysis and Group Fairness

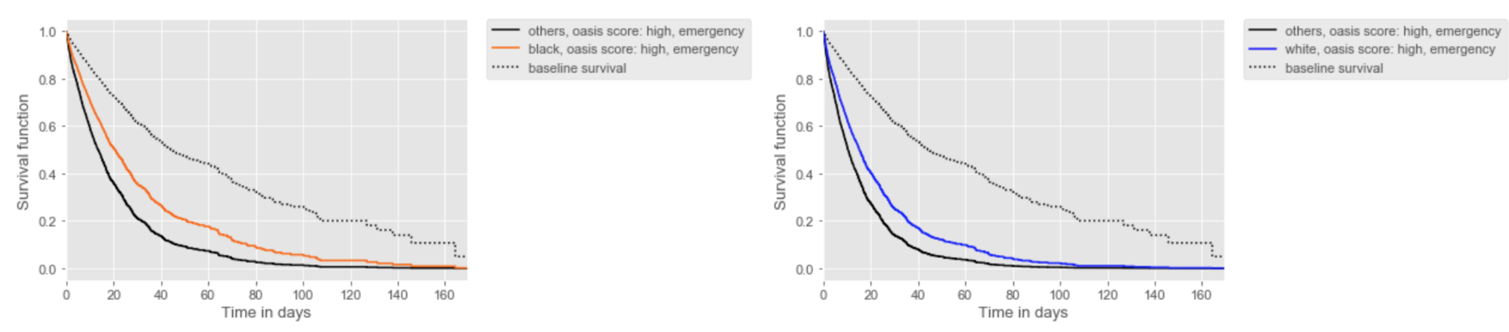

Figure 4: Survival analysis and Conditional Statistical Parity

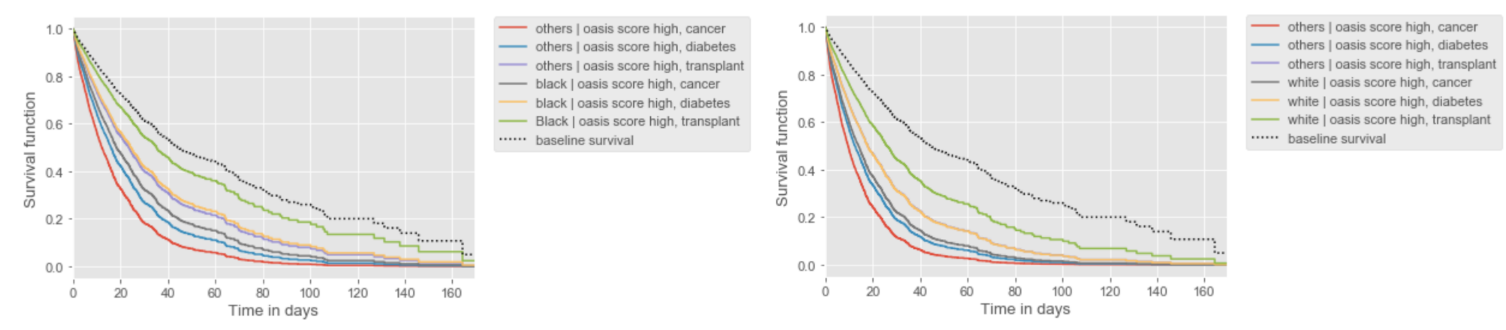

Figure 5: Another case of survival analysis and Conditional Statistical Parity (with more legitimated factors)

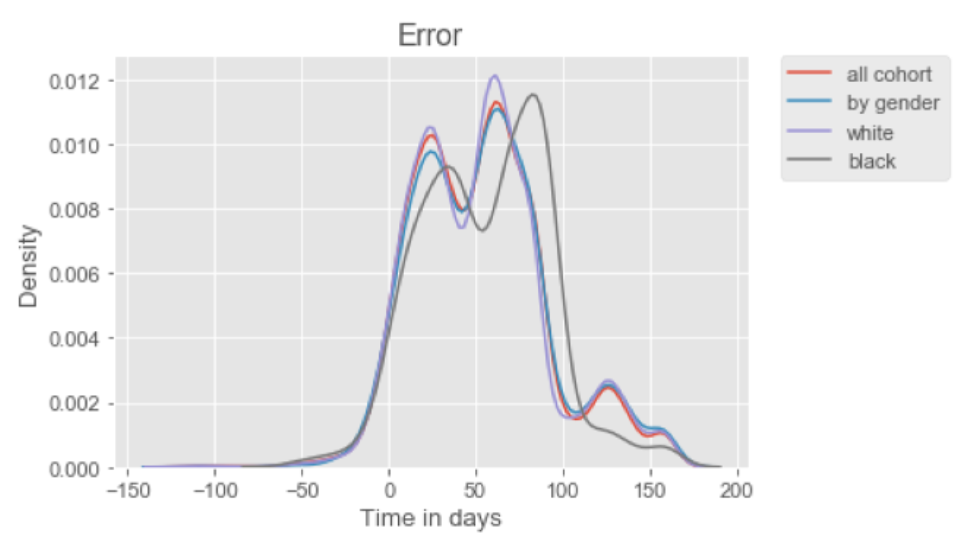

Figure 6: Survival analysis and Equal Opportunity Error

Expected Contributions and Future Work

- Include eletronic health records (EHR) (+features).
- Apply machine learning methods (e.g. K Fold CV, Random Survival Forests).
- Compare our findings with mistrust metrics reported in existing literature.
- Bridging findings from other spheres (e.g criminal justice) to a different application and cultural domains.

References

[1] Maria Beatriz Figueira Alkmim et al. Improving patient access to specialized health care: the telehealth network of minas gerais, brazil. *Bulletin of the World Health Organization*, 90:373–378, 2012.

[2] Sam Corbett-Davies et al. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806. ACM, 2017.

[3] Cynthia Dwork et al. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226. ACM, 2012.

[4] Alistair EW Johnson et al. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

[5] Tom J Pollard et al. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific data*, 5, 2018.

Source Code: `https://github.com/bseewald/fairness-in-health`

Acknowledgements: This work has been partially supported by the project ATMOSPHERE (atmosphere-eubrazil.eu). We also thank LXAI for the travel grants to present this work. Funding was provided by the authors' individual grants from CNPq and CAPES.