Assessing Racial Fairness of Dialysis Allocation in End-Stage Renal Disease

Linying Zhang[1], Lauren R. Richter[1], David M. Blei[2], Yixin Wang[3],
Anna Ostropolets[1], Noemie Elhadad[1], George Hripcsak[1]
[1] Department of Biomedical Informatics, Columbia University, New York, NY;
[2] Department of Computer Science and Statistics, Columbia University, New York, NY
[3]Department of Statistics, University of Michigan, Ann Arbor, MI

Background

Assessing fairness in clinical decisions is an important element for improving the quality of care. While the medical ideal is to base decisions on a patient's health condition, this is not the reality. Gender, race, ethnicity, socioeconomic status, and other sensitive attributes can influence clinicians' decision-making process, raising important concerns about inequity in health and health care [1, 2, 3, 4].

There are many perspectives on how to quantify fairness. Some fairness definitions rely on associations in the data (e.g., statistical parity, calibration, and accuracy). Serious concerns have been raised about these fairness definitions because they can't be simultaneously satisfied on a given dataset, except on some edge cases [5, 6, 7]. More researchers are now coming to agree that fairness is fundamentally a causal problem, and causal reasoning and causal inference should be used to quantify fairness from observational data.

In medicine, the idea of fairness is expressed as *health equity*. Health equity can be defined in multiple ways. According to *Communities in Action: Pathways to Health Equity* [8],

> "[h]ealth equity is the state in which everyone has the opportunity to attain full health potential and no one is disadvantaged from achieving this potential because of social position or any other socially defined circumstance".

This work studies the question: how can we use electronic health records (EHRs) to measure health equity? We show how to adapt Imai and Jiang's recently proposed idea of *principal fairness* [9] to the medical domain. As an example, we show how principal fairness can be used to estimate the racial fairness of medical decisions on dialysis initiation for patients with end-stage renal disease.

Methods

Notations and Definitions We introduce principal fairness in the context of clinical decision-making.

Definition 1 *(Principal fairness in clinical decision-making) A decision of medical treatment D_i satisfies principal fairness with respect to the outcome of interest and the sensitive attribute A_i if the decision is independent of the sensitive attribute among patients with the same health potential H_i, i.e., $p(D_i|H_i, A_i) = p(D_i|H_i)$.*

Here, the health potential of a patient i is defined as $H_i = (Y_i(0), Y_i(1))$, where $Y_i(1)$ is the potential outcome that would be observed if a patient is treated, and $Y_i(0)$ is the potential outcome that would be observed if a patient is untreated. For a binary treatment, there are two potential outcomes for each individual, while we always only observe one of the two in an observational dataset. Because potential outcomes are causal quantities, principal fairness is a causal fairness measure.

Principal fairness is essentially saying that if two patients would benefit equally from a treatment, then the decision on whether they receive the treatment should not depend on their sensitive attribute. In this study, the treatment under investigation is dialysis, the outcome of interest is 1-year survival, and the sensitive attribute is race. We can partition the cohort into four groups based on how much a patient benefits from dialysis. Some patients who would always survive regardless of whether they have dialysis or not (i.e., they are in a relatively stable phase of the disease), then these patients are considered to benefit equally from dialysis (called "stable" group). On the contrary, some patients would always die within a year even if they receive dialysis (i.e., they are in the terminal phase of the disease), then patients in this group are considered to benefit equally from dialysis (called "severe" group). There are also patients whose life would be saved by dialysis (survive if receive dialysis, otherwise die, thus called "treatable" group), and patients who do better without dialysis (die if receive dialysis and survive otherwise, thus called "better-without" group). If the probability of receiving dialysis between Black and non-Black patients is the same across all groups, then the decision on dialysis initiation satisfies principal fairness.

Algorithm We developed a Bayesian probabilistic model to estimate the principal fairness measure $\Delta(h)$, defined as

$$\Delta(h) = p\{D_i = 1 | A_i = a, H_i = h\} - p\{D_i = 1 | A_i = a', H_i = h\}, \ \forall h \in \mathcal{H}$$

By estimating $\Delta(h)$ for each group, we can examine whether or not dialysis initiation is made fair across Black and non-Black patients. Because estimating $\Delta(h)$ requires knowing the two potential outcomes, the model first estimates the missing potential outcome (the outcome that would have been observed had the patient been on the treatment that is not his/her observed treatment), for each patient. Because this causal quantity is not always estimable from observational data, we make the following causal assumptions: stable unit treatment value assumption, ignorability, and positivity [10, 9].

Data, Cohort and Features Data for this study come from Columbia University Irving Medical Center de-identified EHR database in the OHDSI OMOP CDM v5 format. The end-stage renal disease (ESRD) cohort (identified by diagnosis codes) consists of 6,422 patients, among which 3,223 were treated with dialysis and 3,199 were untreated. The index date for the treatment group is the first dialysis visit, and a random clinical visit for the control group.

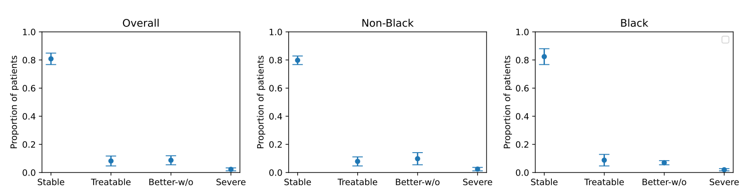
We extracted 728 pre-treatment covariates, including demographics, diagnoses and medications 365 days prior to the index date, and 7 laboratory measurements related to renal function within 60 days prior to the index date. Approximately 1/3 of the patients were Black or African American.

Results and Conclusions

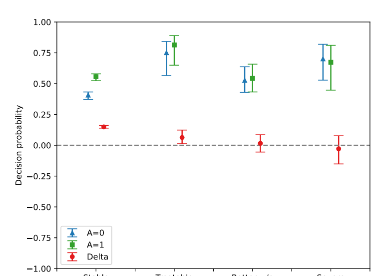
Fig.1(a) shows that 1) about 80% of the ESRD patients would survive for at least 1 year even without dialysis ("stable" group). Although 80% survival rate at first year is expected for patients on dialysis, it is believed that patients would have a shorter life expectancy if not undergo dialysis. This discrepancy could be due to the control group accidentally including patients who received dialysis in reality but didn't have the procedure code in their record.

Fig.1(b) shows that there is no significant racial difference in dialysis allocation in three out of the four groups, except "stable" group where Black patients have a 13% higher probability of receiving dialysis compared to non-Black patients within the same group. Although dialysis is generally not considered for patients in earlier stages according to the clinical guidelines, we did find that more non-Black patients initiated dialysis prior to diagnosis of ESRD, and these patients were excluded from the study. We are expanding the study cohort to include earlier disease stages.

In conclusion, principal fairness is a potential fairness criterion for assessing fairness of clinical decisions.



(a) Health potential group distribution for the entire cohort(left), non-Black patients(middle) and Black patients(right)



(b) Treatment probability ($A = 1$: Black race) and principal fairness delta ($A = 1$ - $A = 0$)

Figure 1: Racial fairness of dialysis allocation in ESRD.

References

1. Dehon E, Weiss N, Jones J, Faulconer W, Hinton E, Sterling S. A Systematic Review of the Impact of Physician Implicit Racial Bias on Clinical Decision Making. Academic Emergency Medicine. 2017;24(8):895-904.
2. FitzGerald C, Hurst S. Implicit Bias in Healthcare Professionals: A Systematic Review. BMC Medical Ethics. 2017 Mar;18(1):19.
3. Aberegg SK, Terry PB. Medical Decision-Making and Healthcare Disparities: The Physician's Role. The Journal of Laboratory and Clinical Medicine. 2004 Jul;144(1):11-7.
4. van Ryn M, Burke J. The Effect of Patient Race and Socio-Economic Status on Physicians' Perceptions of Patients. Social Science & Medicine. 2000 Mar;50(6):813-28.
5. Rahmattalabi A, Xiang A. Promises and Challenges of Causality for Ethical Machine Learning. arXiv:220110683 [cs]. 2022 Jan.
6. Makhlouf K, Zhioua S, Palamidessi C. Survey on Causal-based Machine Learning Fairness Notions. 2020 Oct.
7. Loftus JR, Russell C, Kusner MJ, Silva R. Causal Reasoning for Algorithmic Fairness. arXiv:180505859 [cs]. 2018 May.
8. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States. Communities in Action: Pathways to Health Equity. National Academies Press (US); 2017.
9. Imai K, Jiang Z. Principal Fairness for Human and Algorithmic Decision-Making. arXiv:200510400 [cs, stat]. 2021 Jan.
10. Rubin DB. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology. 1974;66(5):688.