AI and Ethics in Healthcare

Shalmali Joshi (Postdoc, Vector Institute)

AI and Ethics: Mathematical Foundations and Algorithms

Fall, 2019 (CSC 2541F)

Overview

- Ethics in healthcare
  - Challenges
  - Bioethics: Foundation of ethics in healthcare
- AI in the mix
  - Limitations of algorithmic fairness
  - Overview of *fairness* in AI and healthcare
  - Where can AI really help?
  - Beyond Classification

Challenges

TECHNOLOGY

Google's Totally Creepy, Totally Legal Health-Data Harvesting

Google is an emerging health-care juggernaut, and privacy laws weren't written to keep up.

SIDNEY FUSSELL NOVEMBER 14, 2019

MORE STORIES

Did Body Cameras Backfire?

SIDNEY FUSSELL

Why Politicians Want Your Smart-TV Data

SIDNEY FUSSELL

The Toxic Bubble of Technical Debt Threatening America

ALEXIS C. MADRIGAL

Privacy

Challenges: Systemic bias, disparity, generalizability

Cause-specific infant mortality rates per 1,000 live births coded according to modified International Collaborative Effort grouping, by Indigenous identity, singleton births, Canada, 2004 through 2006

International Collaborative Effort grouping	Non-Indigenous			Indigenous											
				Total			First Nations			Métis			Inuit		
	Rate	95% confidence interval		Rate	95% confidence interval		Rate	95% confidence interval		Rate	95% confidence interval		Rate	95% confidence interval	
		from	to		from	to		from	to		from	to		from	to
Congenital anomalies	1.3	1.1	1.5	2.2	1.6	3.0	1.9	1.5	2.5	3.1	1.4	6.6	x	x	x
Asphyxia-related conditions	0.6	0.4	0.8	0.6	0.3	1.1	0.4	0.2	0.6	x	x	x	x	x	x
Immaturity-related conditions	1.2	1.0	1.4	1.6	1.0	2.5	2.0	1.2	3.4	x	x	x	x	x	x
Infections	0.3	0.2	0.4	1.4	0.8	2.4	1.0	0.5	1.9	x	x	x	x	x	x
Respiratory conditions	0.8	0.7	0.9	1.2	0.8	1.8	2.2	1.4	3.6	x	x	x	2.5	1.9	3.3
Sudden infant death syndrome	0.4	0.3	0.5	0.4	0.2	0.8	0.7	0.3	1.5	x	x	x	x	x	x
Injuries	0.1	0.0	0.1	0.1	0.0	0.2	0.0	0.0	0.2	0.0	0.0	1.3	0.0	0.0	1.7
Other conditions	0.8	0.7	0.9	1.0	0.7	1.5	x	x	x	x	x	x	x	x	x

< Previous Article | May 2016 Volume 149, Issue 5, Pages 1128–1130 | Next Article >

POINT: Do Randomized Controlled Trials Ignore Needed Patient Populations? Yes

Katherine Courtright, MD*
Pulmonary, Allergy and Critical Care Division, Hospital of the University of Pennsylvania, Philadelphia, PA

PlumX Metrics

DOI: https://doi.org/10.1016/j.chest.2016.01.029 | Check for updates

Harvard Heart Letter

The heart attack gender gap

Heart attacks strike men at younger ages than women. But survival rates are worse in women. Why?

Published: April, 2016

Compared with men, women are less likely to recognize and act upon the symptoms of a heart attack.
Image: zaganDesign/Thinkstock

:han twice as high
ared with the non-

Other Challenges: Cultural context, informed consent…

Bioethics: Foundation of ethics in healthcare

Nonmaleficence

Beneficence

Justice

Autonomy

AI in the mix: Limitations

- Focus: Systemic bias, disparity, generalizability

Nonmaleficence

Beneficence

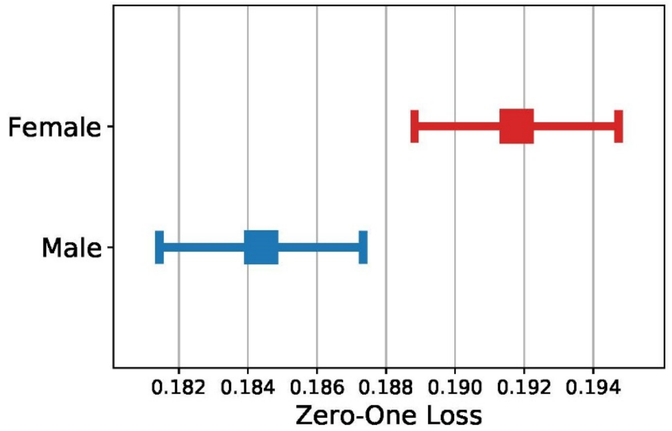
Justice

Autonomy

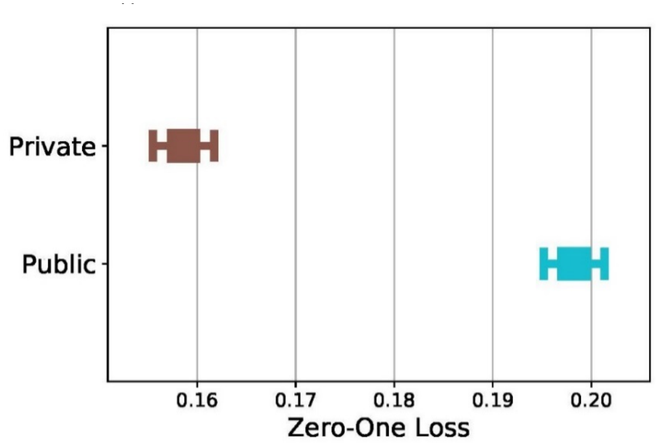
Algorithmic fairness in ML hasn't operationalized these for classification (some exceptions!)

Overview of *fairness* in AI and healthcare

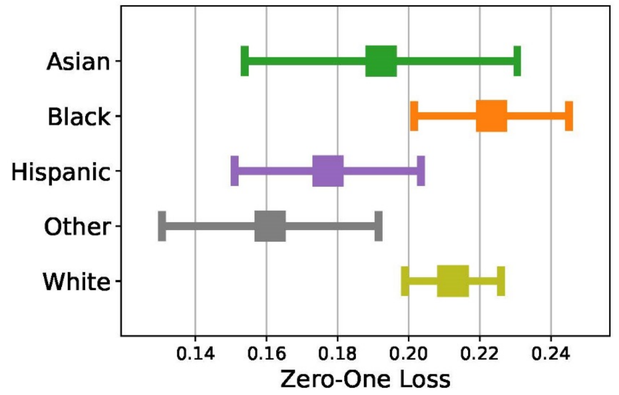
- Identify sources of disparity



95% confidence intervals for error rates in ICU mortality prediction on MIMIC-III clinical notes



95% confidence intervals for error rates in ICU mortality prediction on MIMIC-III clinical notes



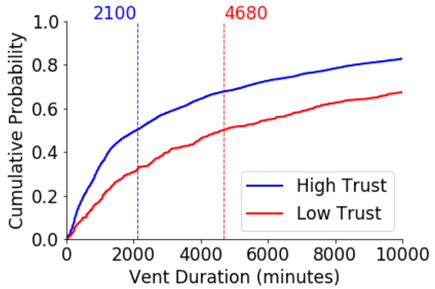
95% confidence intervals for error rates in psychiatric readmission prediction on a New England hospital cohort

Predicting mortality and psychiatric readmission from unstructured clinical notes

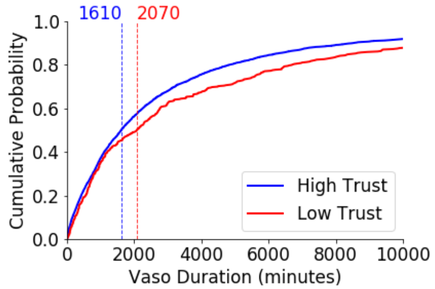
Chen, Irene et al , Can AI Help Reduce Disparities in General Medical and Mental Health Care? AMA Journal of Ethics, 2019

Overview of *fairness* in AI and healthcare

- Identify sources of disparity

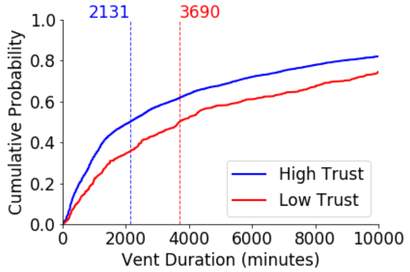


(a) **Mechanical Ventilation**
White: 4810 patients
Black: 510 patients
$p < 0.001$

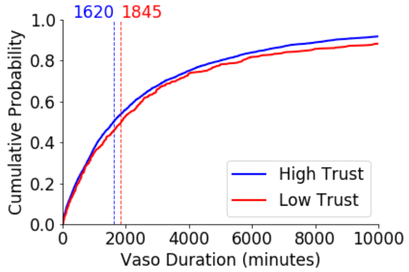


(b) **Vasopressors**
White: 4456 patients
Black: 453 patients
p=0.001

Non-compliance derived cohort and aggressive care reflected in treatment durations



(a) **Mechanical Ventilation**
White: 4810 patients
Black: 510 patients
p<0.001



(b) **Vasopressors**
White: 4456 patients
Black: 453 patients
p=0.059

Autopsy derived cohort and aggressive care reflected in treatment durations

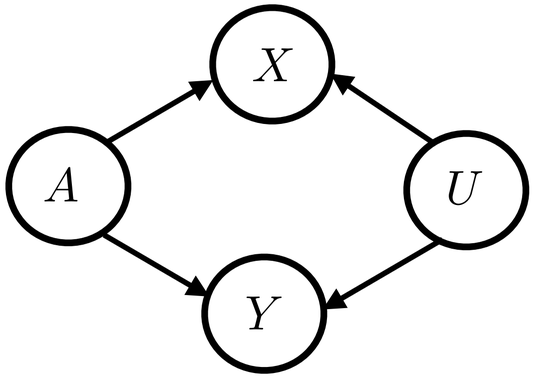
Mistrust between patients and caregivers reflects disparity in End-of-life care

Mistrust reflected in i) non-compliance, and ii) autopsy rates

Boag, Willie et al , Racial Disparities and Mistrust in End-of-Life Care MLHC, 2018

Overview of *fairness* in AI and healthcare

- Algorithmic solutions to fairness in healthcare



$$p(\hat{Y}_{A\leftarrow a}(U)|X=x, Y_{A\leftarrow a}=y, A=a) = p(\hat{Y}_{A\leftarrow a'}(U)|X=x, Y_{A\leftarrow a'}=y, A=a)$$

Individual Equalized odds Counterfactual Fairness (IECF)

$$p(\hat{Y}_{A\leftarrow a}(U)|X=x, A=a) = p(\hat{Y}_{A\leftarrow a'}(U)|X=x, A=a)$$

Counterfactual Fairness (CF)

$$V(h(x,a),y) = 1 - \alpha_0 p(\mathbf{1}[h(x,a)\geq T]=1|Y=0)p(Y=0|X=x,A=a)$$
$$-\alpha_1 p(\mathbf{1}[h(x,a)\geq T]=0|Y=0)p(Y=1|X=x,A=a)$$

Utility of a predictor - reasonable for a clinical policy

Difference: Utility of a predictor under CF is not a function of true outcome and $Y=1$ is preferred

Counterfactual Reasoning for Fair Clinical Risk Prediction (MIMIC-III Mortality Prediction task)

Pfohl, Stephen et al , Counterfactual Reasoning for Fair Clinical Risk Prediction MLHC, 2019

Overview of *fairness* in AI and healthcare

- Algorithmic solutions to fairness in healthcare

$$p(f(X)|A = A_i, Y = Y_k) = p(f(X)|A = A_j, Y = Y_k) \forall A_i, A_j \in \mathcal{A}; Y_k \in \mathcal{Y}$$

Equalized odds for Risk Scoring (Enforce the same ROC Curve for protected groups)

Training procedure:

1. Learn regressor to predict risk
2. Leverage adversarial learning to match group specific distribution of scores

$f : \mathbb{R} \rightarrow [0, 1]$ (parametrized by θ_f)

$g : \mathbb{R} \times \mathcal{Y} \rightarrow [0, 1]^k$ (parametrized by θ_g)

$min_{\theta_f} L_{cls} - \lambda L_{adv}$

$min_{\theta_g} L_{adv}$

Atherosclerotic cardiovascular disease risk stratification model

Pfohl, Stephen et al , Creating Fair Models Of Atherosclerotic Cardiovascular Disease AAAI/AIES, 2019

Overview of *fairness* in AI and healthcare

- Focus: Systemic bias, disparity, generalizability

Nonmaleficence

Beneficence

Justice

Autonomy

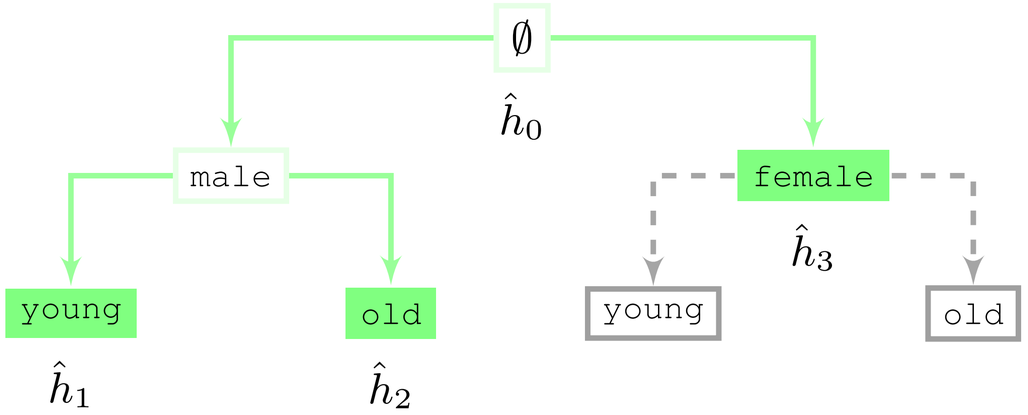
Algorithmic fairness in ML hasn't operationalized these for classification (some exceptions!)

Overview of *fairness* in AI and healthcare

- Operationalizing bioethical principles

Nonmaleficence

Beneficence



Beneficence: Decoupled classifiers — i.e., train a classifier for each group using data from that group

Nonmaleficence: Loosely similar to ***preference guarantees*** — i.e. each group should prefer their assigned model to (i) a pooled model that ignores group membership (***rationality***) and (ii) the model assigned to any other group (***envy-freeness***)

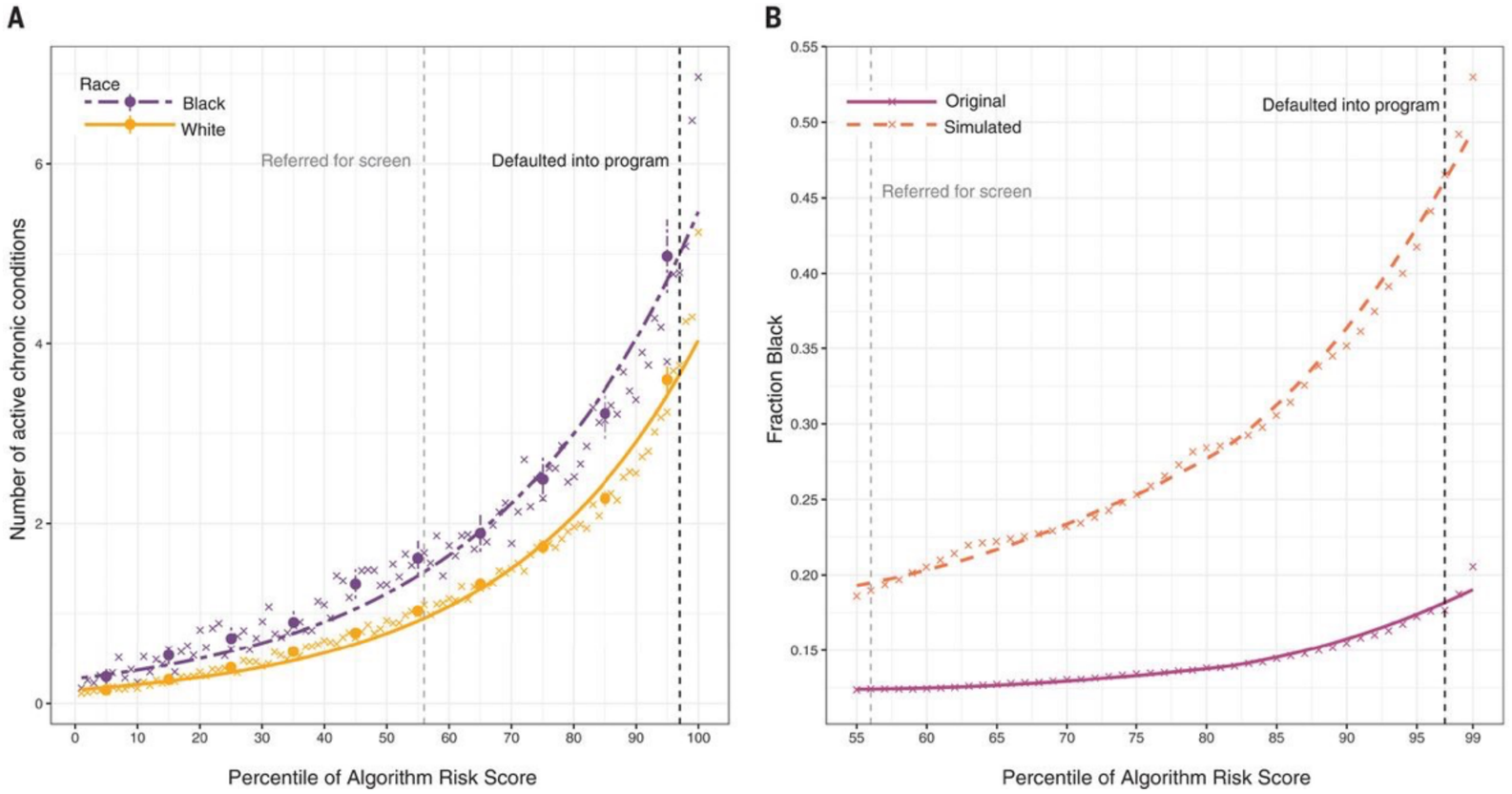
- Reliable risk estimation for **identifiable and intersectional subgroups** in healthcare is critical

Ustun, Berk et al, Fairness without Harm: Decoupled Classifiers with Preference Guarantees, ICML 2018
Hebert-Johnson, Ursula et al, Multicalibration: Calibration for the (Computationally-Identifiable) Masses, ICML 2018

Where can AI/ML really help?

- Audit existing algorithms - Bias in referrals to costly care management programs

Commercial algorithm for targeting patients for "high risk care management" underestimated the needs of black patients

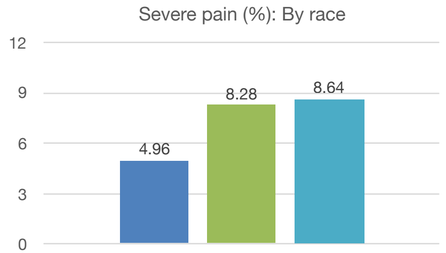


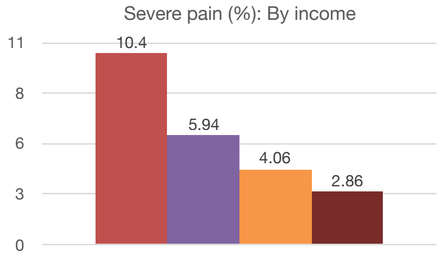
Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 2019.

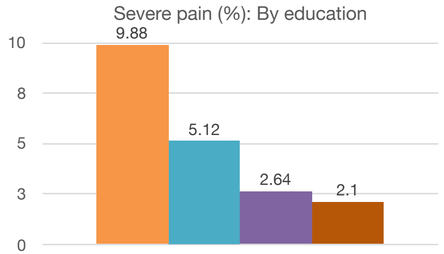
Where can AI/ML really help?

- Audit existing algorithms - Understanding and fixing bias in knee pain

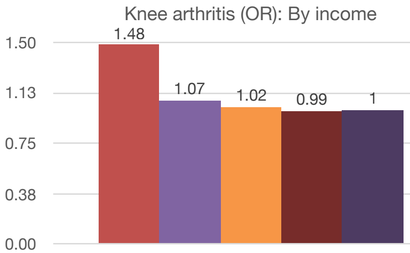
  - Large pain gradients
    - Race
    - Income
    - Education

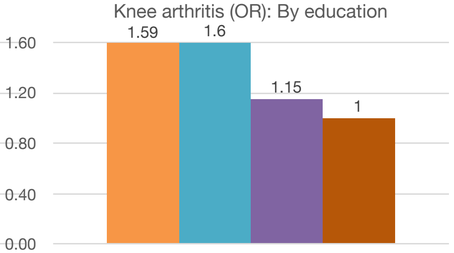




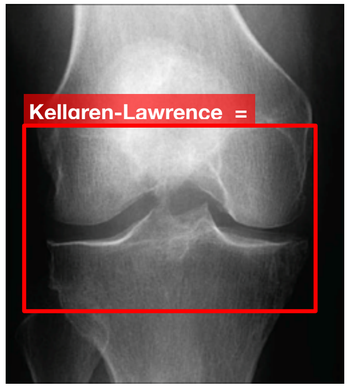


  - Higher prevalence of painful conditions
    - By income
    - By education

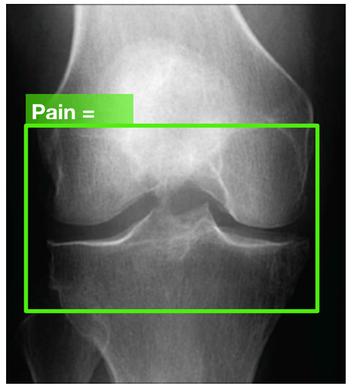




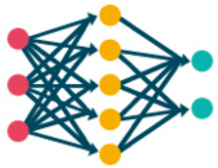
What if instead of learning from the radiologist…



We trained the algorithm to listen to the patient?



Simulation: Who would get surgery…
if the algorithm were in charge, not the doctor?



- Identify patients with severe pain and
  - High disease severity according to human
  - High disease severity according to algorithm

More - Black knees eligible for surgery

Less - Black knees, severe pain but ineligible for surgery

Severe pain + no surgery + high algorithm score = most likely to be on oral pain medicine incl. opiates

Slide courtesy - Ziad Obermeyer from https://blogs.worldbank.org/impactevaluations/machine-learning-pain-relief

Grol-Prokopczyk, *Pain* 2017, Baldassari et al., *Osteoarthritis and Cartilage* 2014

Pierson, Emma, et al. "Using machine learning to understand racial and socioeconomic differences in knee pain" Under Review at JAMA 2019.

Beyond Classification

- Ignoring sources of implicit bias in observational healthcare data

Model performance

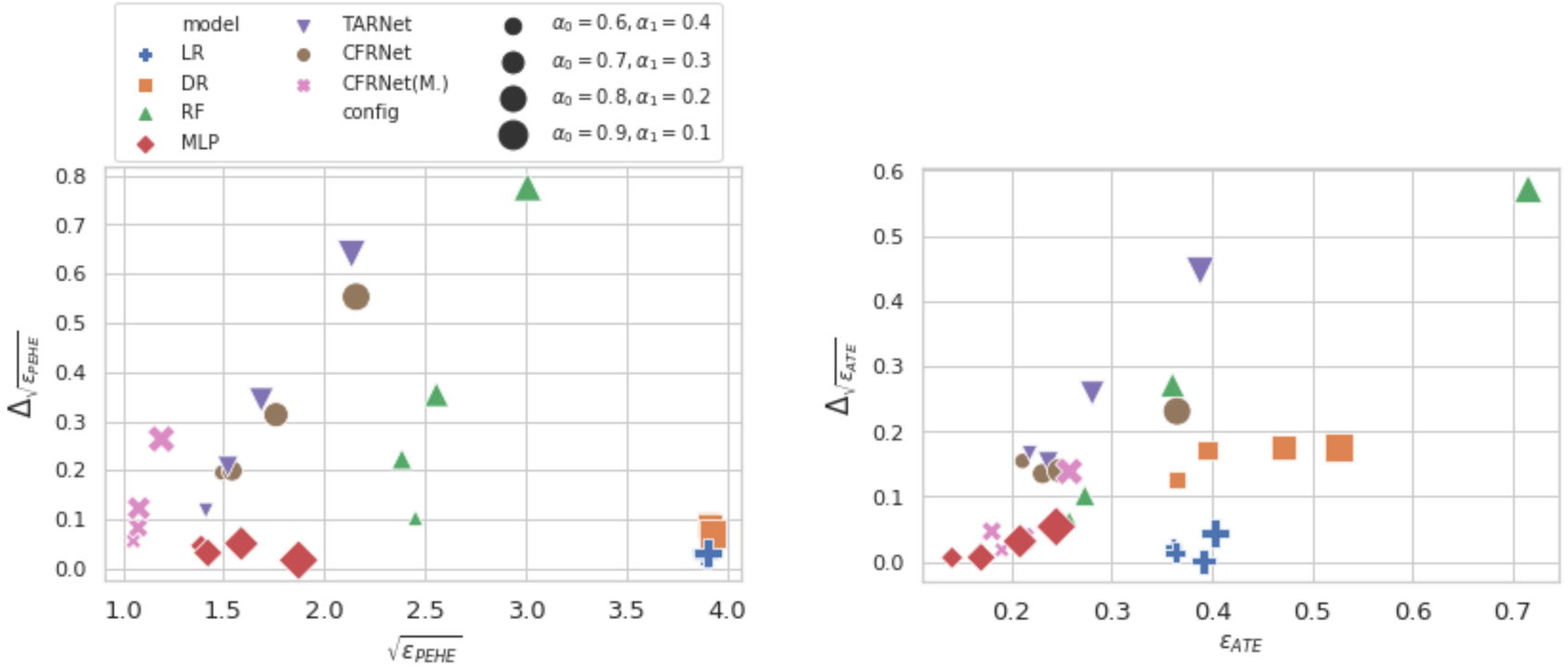
$$\sqrt{\epsilon_{PEHE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_1}(\mathbf{x}_i) - \hat{y_0}(\mathbf{x}_i) - (y_1(\mathbf{x}_i) - y_0(\mathbf{x}_i)))^2}$$

$$\epsilon_{ATE} = |\frac{1}{n}\sum_{i=1}^{n}(\hat{y_1}(\mathbf{x}_i) - \hat{y_0}(\mathbf{x}_i)) - \frac{1}{n}\sum_{i=1}^{n}(y_1(\mathbf{x}_i) - y_0(\mathbf{x}_i))|$$

Disparity in causal effect estimation

$$\Delta_{\sqrt{\epsilon_{PEHE}}} = |\sqrt{\epsilon_{PEHE}}_{A=0} - \sqrt{\epsilon_{PEHE}}_{A=1}|$$

$$\Delta_{\epsilon_{ATE}} = |\epsilon_{ATE\,A=0} - \epsilon_{ATE\,A=1}|$$



- Always consider implicit bias when doing covariate selection for causal effect estimation
- Conventional propensity scoring models - Protected attribute inclusion improves effect estimations unless there is exclusive treatment disparity
- Flexible models like deep neural networks are more amenable to misspecification of generative assumptions - but use all covariates!

Yi et al, Fair and Robust Treatment Effect Estimates:Estimation Under Treatment and Outcome Disparitywith Deep Neural Models, Fair ML for Health, NeurIPS 2018

Conclusion

- To tackle disparities:
  - Leverage mathematical foundations of AI for better science in healthcare
  - Formulate the right problem/task (think beyond models)
    - Heart attack gender gap
    - Endometriosis diagnosis delays
    - Black and indigenous infant and maternal mortality
  - Operationalize bioethical principles as fairness metrics for evaluation

Chen, Irene et al Treating Health Disparities with AI Nature Medicine, 2019 (to appear)