

# Learning Disease Progression Models That Capture Health Disparities

Erica Chiang (eric Chiang@cs.cornell.edu)<sup>1,2</sup>, Divya Shanmugam<sup>1,2</sup>, Ashley Beecy<sup>3,4</sup>, Gabriel Sayer<sup>3,5</sup>, Deborah Estrin<sup>1,2</sup>, Nikhil Garg<sup>1,2</sup>, Emma Pierson<sup>6</sup>

<sup>1</sup> Cornell University   <sup>2</sup> Cornell Tech   <sup>3</sup> NewYork-Presbyterian   <sup>4</sup> Weill Cornell Medical College   <sup>5</sup> Columbia University Irving Medical Center   <sup>6</sup> University of California, Berkeley

## Background

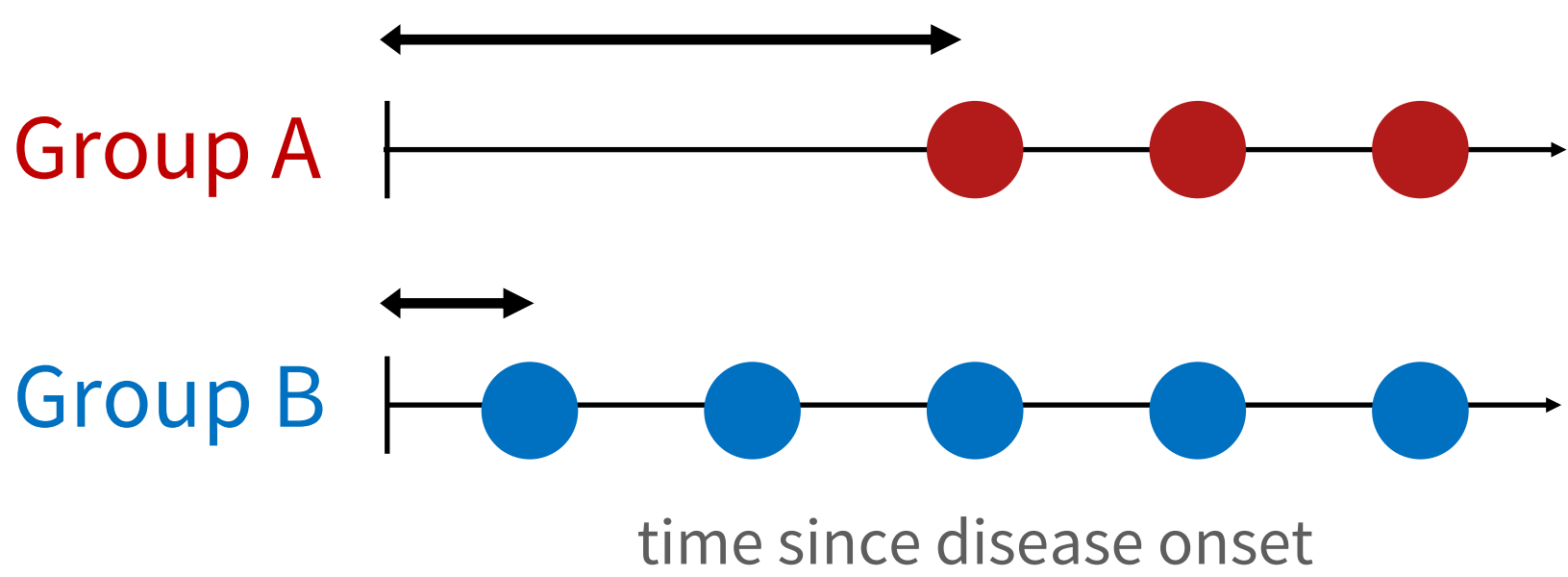
Disease progression models are used to personalize care, detect diseases at earlier stages, and study interventions<sup>[1, 2]</sup>

But these models historically fail to account for disparities that bias the data they are trained on<sup>[3, 4, 5]</sup>

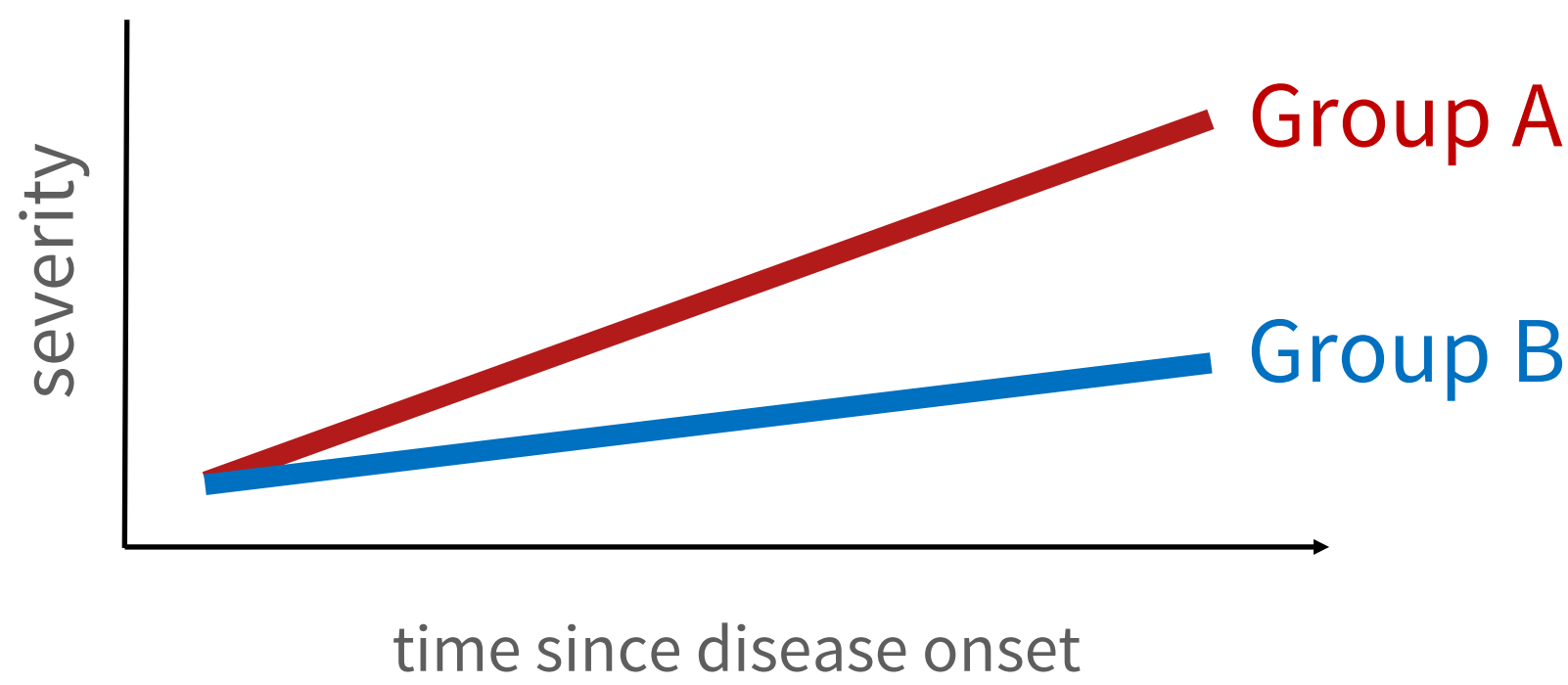
## Axes of Disparities

We define three axes along which we observe and analyze disparities:

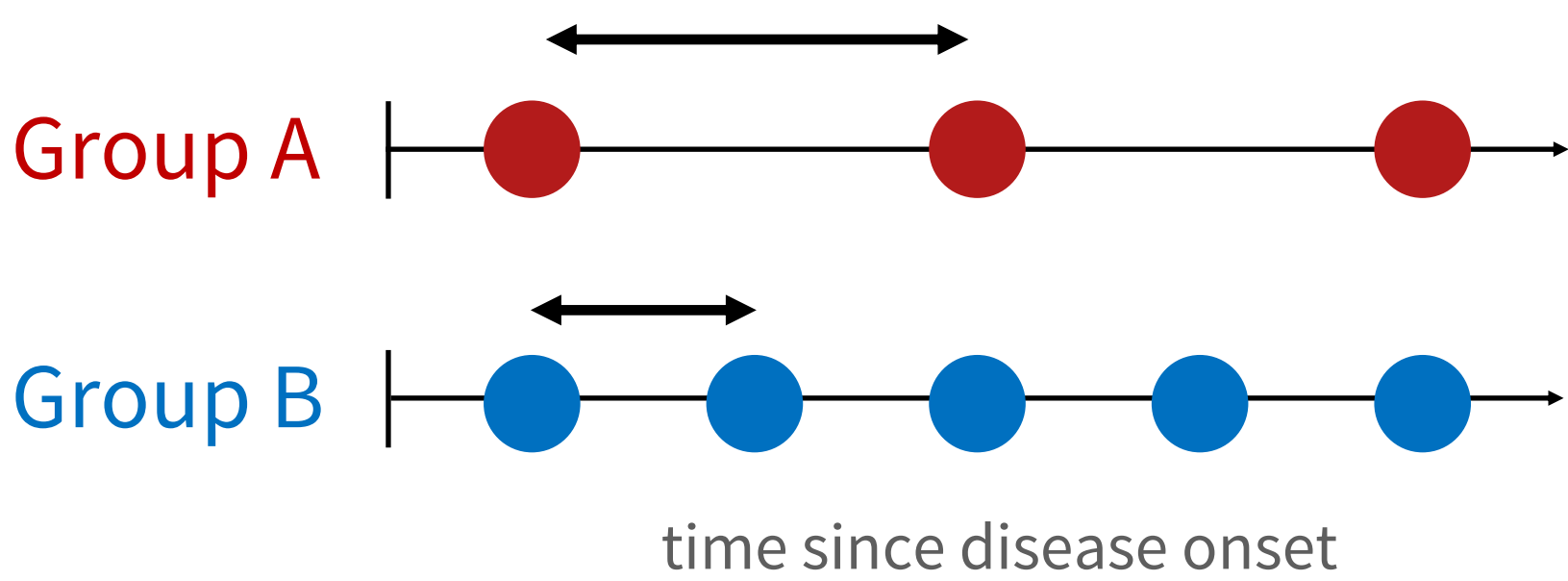
### 1. Disparities in initial severity



### 2. Disparities in progression rate



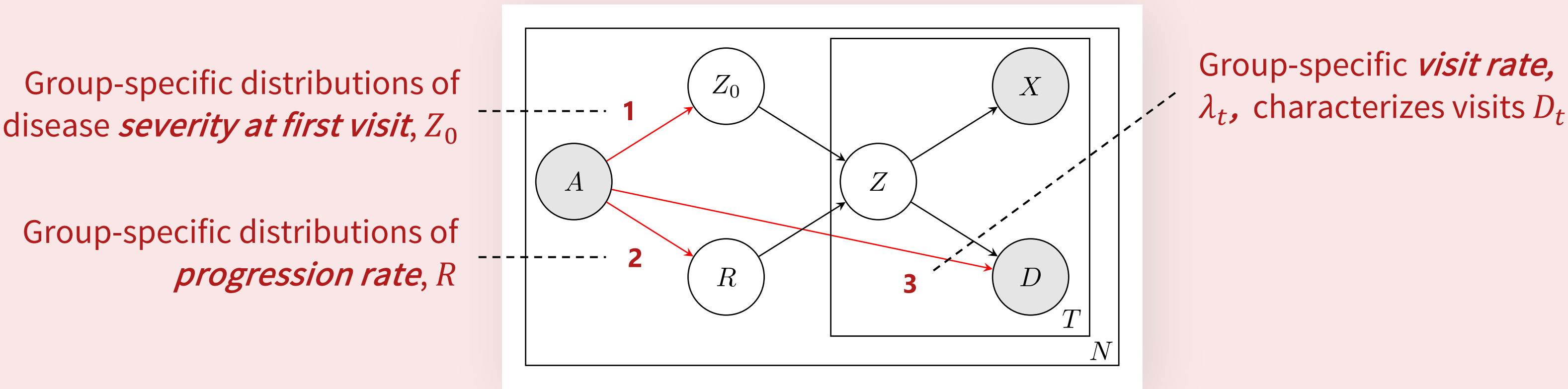
### 3. Disparities in visit frequency



Can we extend a standard disease progression modeling approach to capture health disparities?

## A model that captures three key disparities

We model dependence on a demographic vector  $A$  to capture disparities along our axes:



## 1 More accurate severity estimates

*Theorems 3-5 (informal). In the presence of health disparities, a model that fails to account for any of the three disparities will produce biased estimates of severity.*

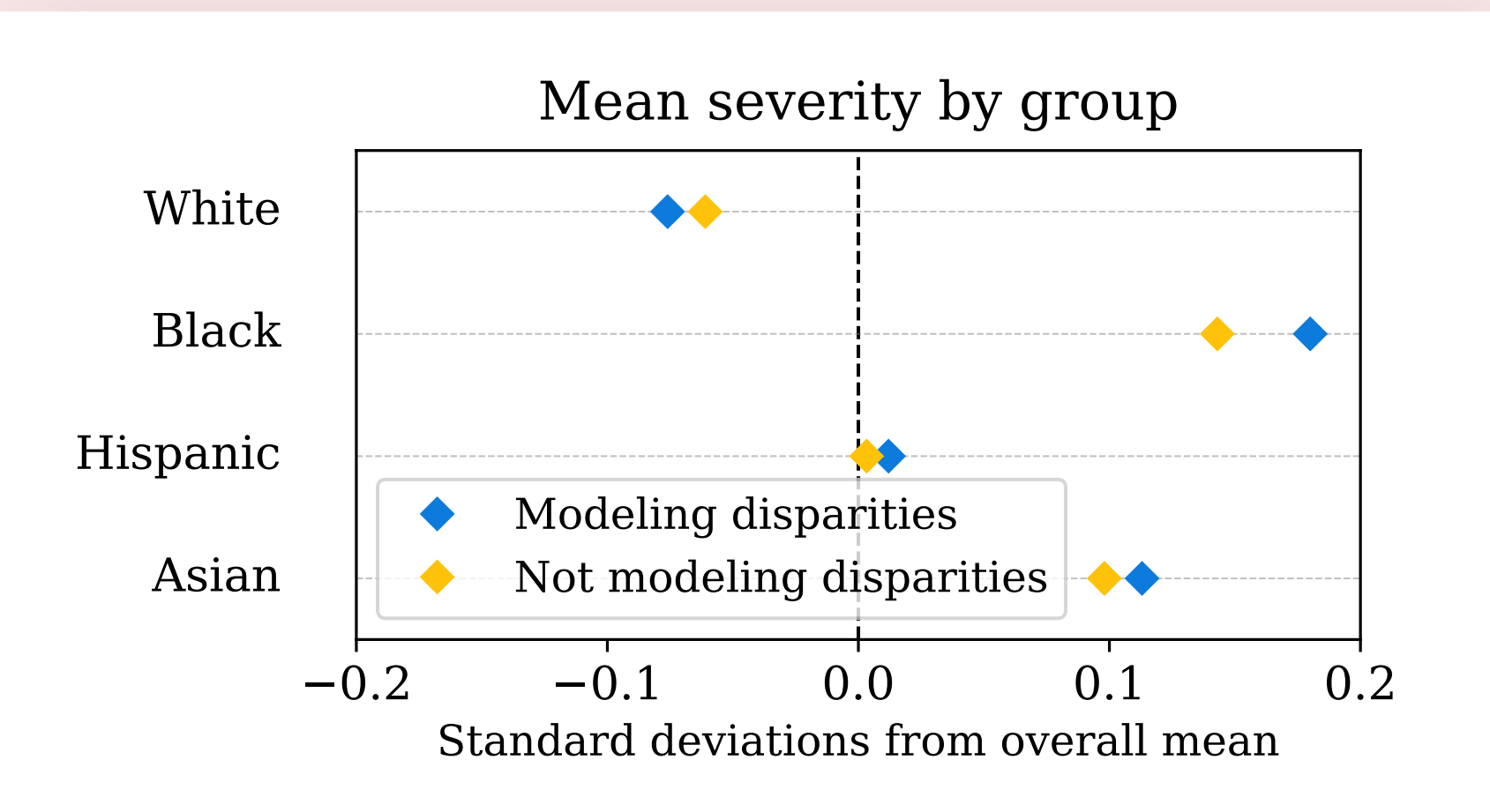
Our model accounts for all three disparities, leading to less biased severity estimates:

### Synthetic data

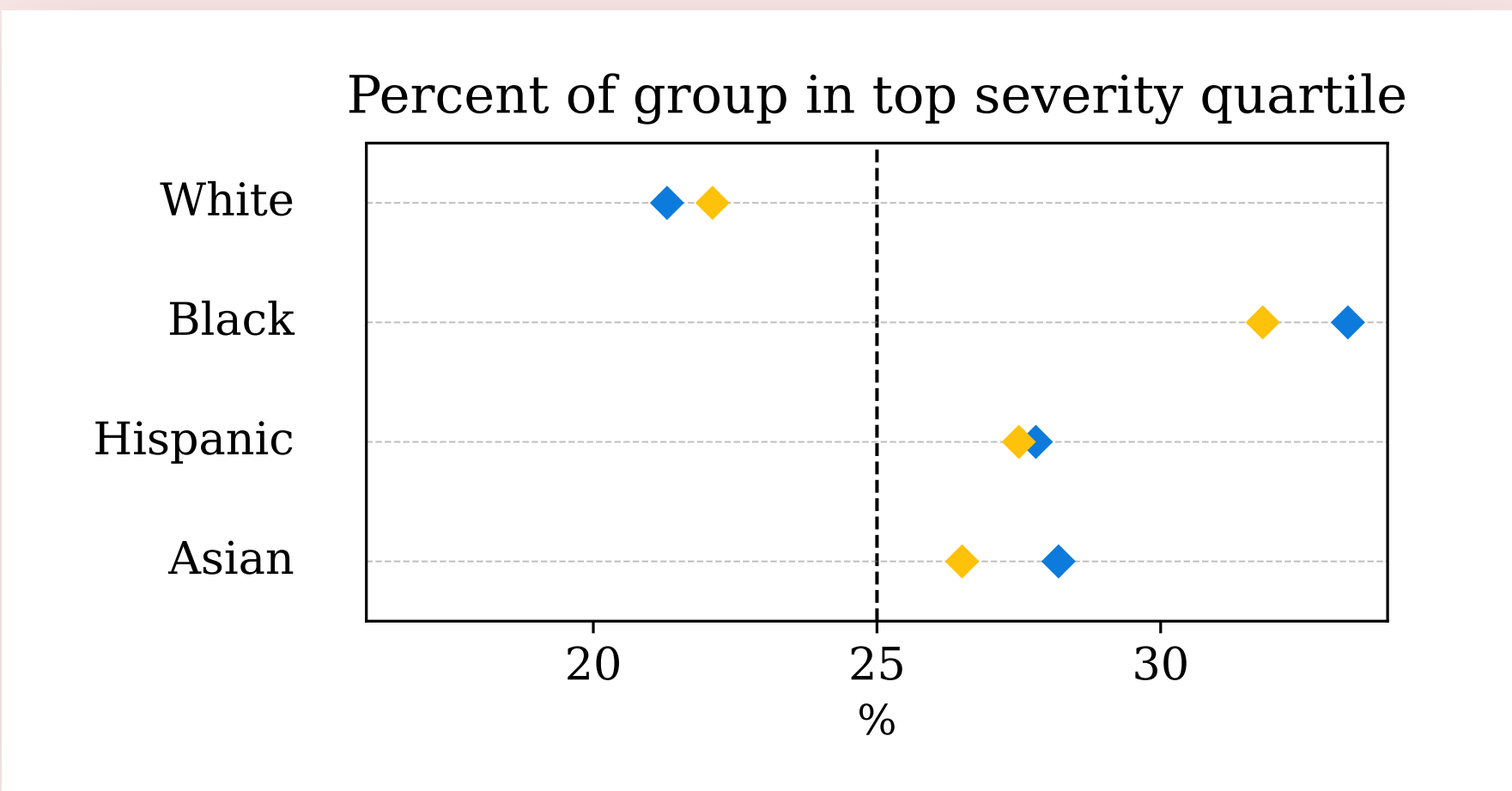
	Model that fails to account for disparities in...			
	Full model	Initial severity	Progression rate	Visit frequency
Underserved group bias	-0.02	-0.89	-0.04	-0.37
Non-underserved group bias	0.01	+1.02	+0.20	+0.33

Failing to account for disparities produces biased severity estimates; our model addresses this

### Real data



Accounting for disparities meaningfully shifts severity estimates for all racial/ethnic groups



Failing to account for disparities biases classification of “high-risk patients” away from groups with higher severity

## 2 Fine-grained descriptions of disparities

Our model remains provably and empirically identifiable while accounting for multiple types of disparities

+

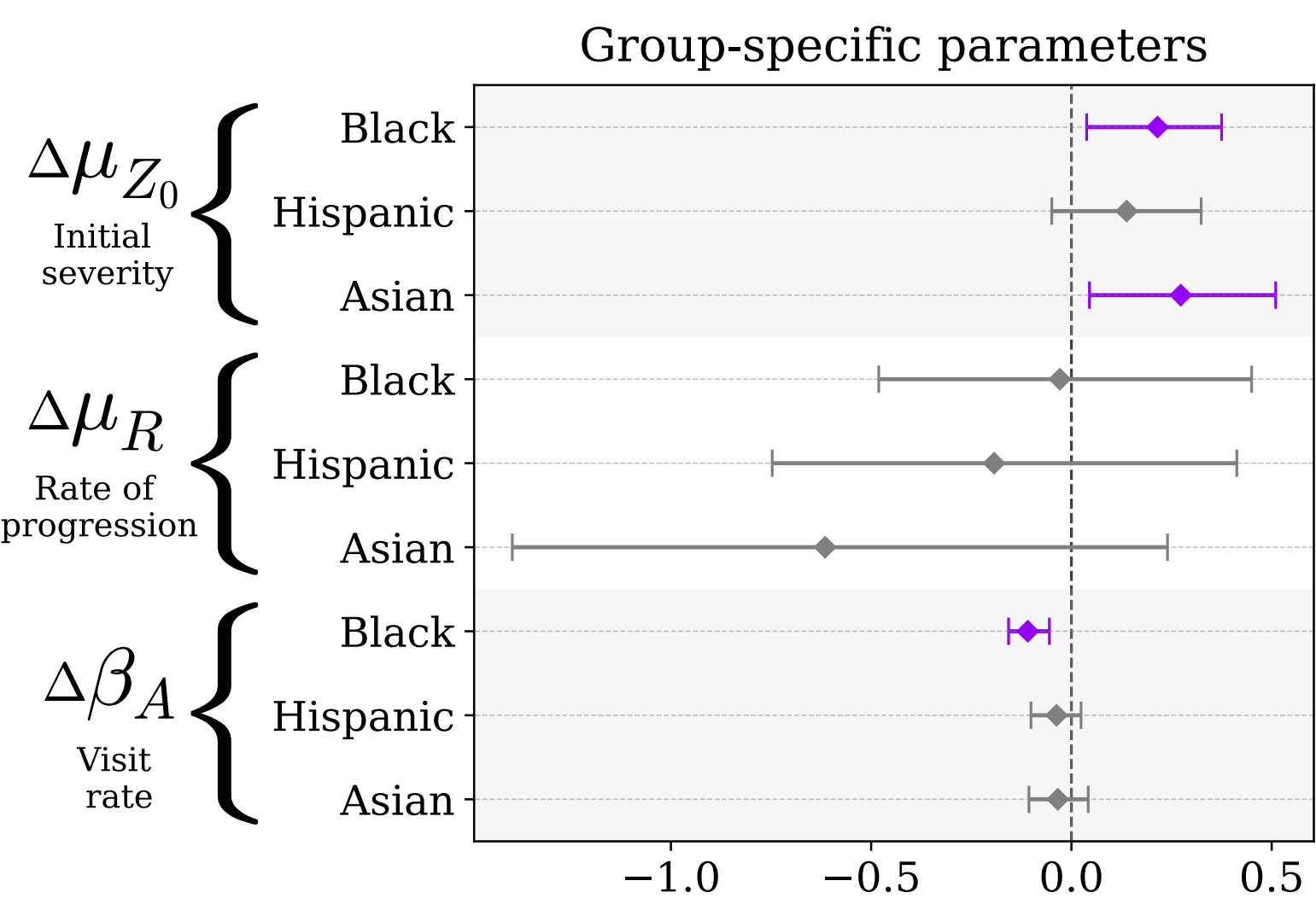
Interpretable parameters

→

Quantify existence of disparities

## Disparities in heart failure progression

We fit our model on NYP heart failure data of 2942 patients and interpret model parameters. We find racial disparities in initial severity and visit frequency:

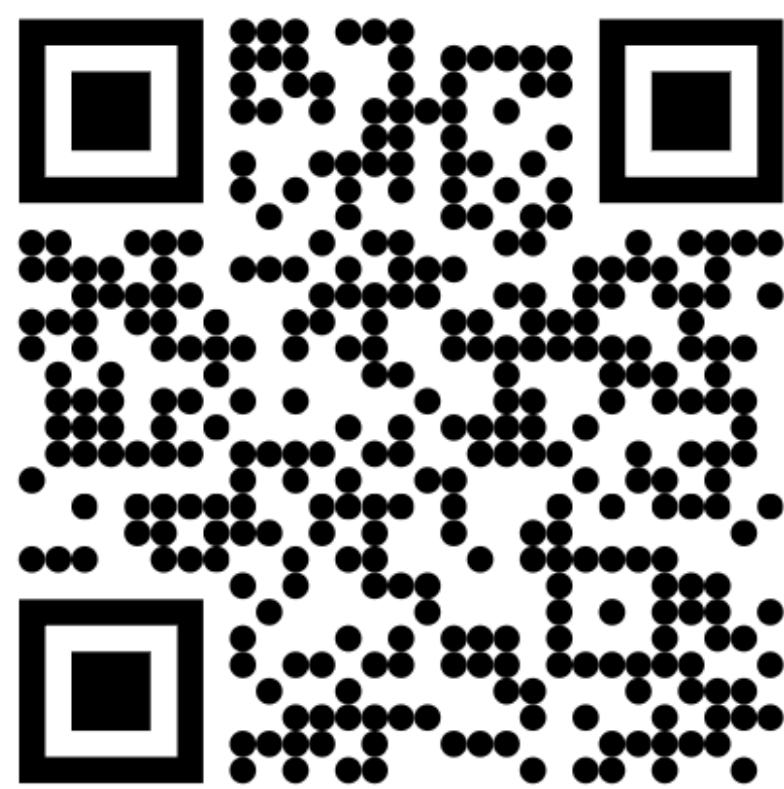


Inferred model parameters with 95% confidence intervals

## Key Contributions

1. Develop an **interpretable, identifiable** disease progression model that accounts for 3 key disparities
2. Prove that failing to account for disparities leads to **biased estimates of severity**
3. Characterize **fine-grained disparities** in a real-world heart failure dataset

Full paper



## References

- [1] Mould et al., 2007. Using disease progression models...
- [2] Romero et al., 2015. The future is now: Model-based clinical trial design...
- [3] Weaver et al., 2010. Forgoing medical care because of cost...
- [4] Reilly, 2021. Health disparities and access to healthcare...
- [5] Yearby, 2018. Racial disparities in health status and access...