The Silent Emergency

Predicting Preterm Birth

Katherine Grillaert, Divya Joshi, Noah Rahman, Alexander Sutherland, and Kristina Zvolanek

GitHub Link

All of Us Link

“The world is facing a silent emergency … of preterm births.” *- UNICEF*

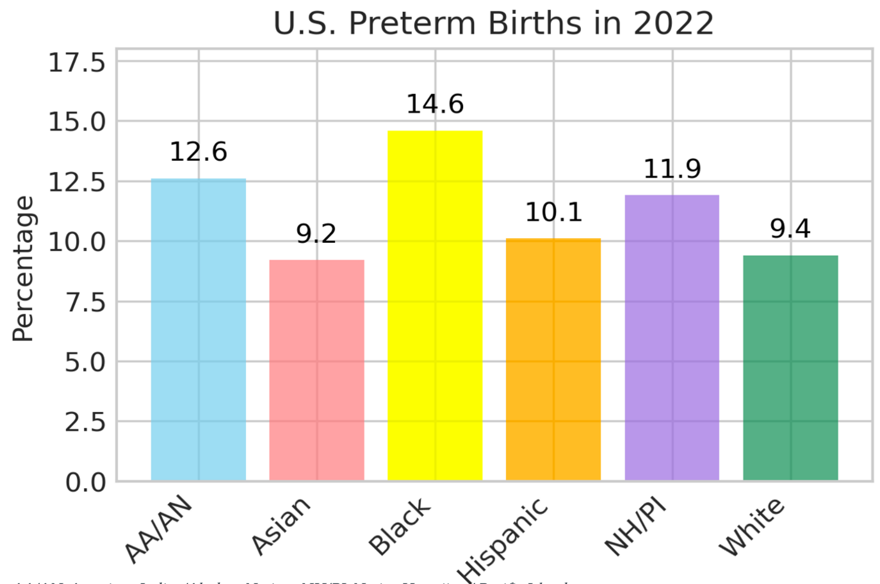
10% of US births are preterm[1]

Risk of death and long-term negative health consequences[2]

Preterm birth results from complex interactions[2]:

- environmental
- biological
- genetic
- behavioral

Current medical efforts focus more on responsive rather than preventative care[3]



AA/AN: American Indian/Alaskan Native; NH/PI: Native Hawaiian / Pacific Islander

Hamilton BE, Martin JA, Osterman MJK. Births: Provisional data for 2022. Vital Statistics Rapid Release; no 28. Hyattsville, MD: National Center for Health Statistics. June 2023. DOI: https://dx.doi.org/10.15620/cdc:127052

Predictive Modeling for Preterm Births

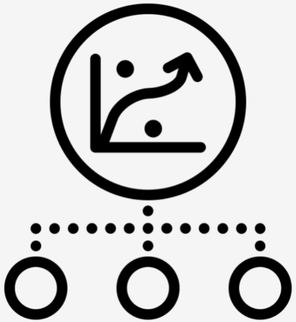
Current gap: Preterm birth is hard to predict!

Predictive models: Could support early and accessible diagnosis[3]

Previous literature: Limited representation in training datasets[4]

Goals of Our Predictive Models:

- Use electronic health records data, specifically social determinants of health associated with preterm birth
- Ensure model's equitable performance across race and ethnicity



Key Stakeholders

Care Receivers
Pregnant individuals
Prospective parents

Care Givers
Medical professionals
Family and friends

Care Facilitators
Hospital systems
Insurance companies

Data Source

NIH All of Us research program:

- Controlled tier of data
- Restricted access
- All data is anonymized

All of Us RESEARCHER WORKBENCH

Trained Models on Two Datasets

Demographics

Health & Lifestyle

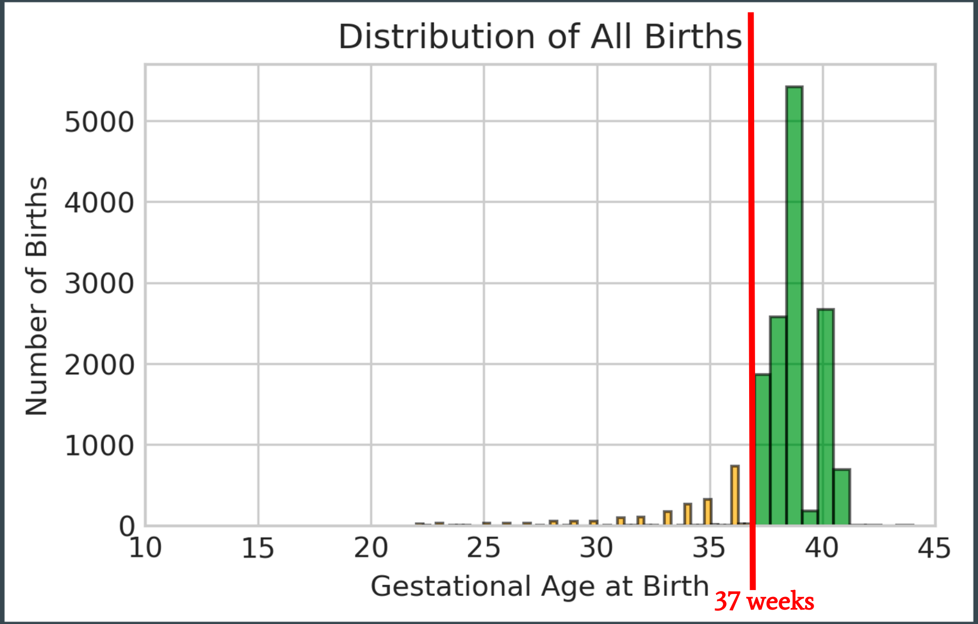
Data origins:

- Electronic health records
- Surveys

Response Variable and Baseline Model

Response variable: binary classes

0: Normal birth
1: Preterm birth



Preterm birth | Normal birth

Baseline model: weighted coin flip

Probability of outcome equal to fraction of preterm birth in dataset

Key Performance Indicators (KPIs)

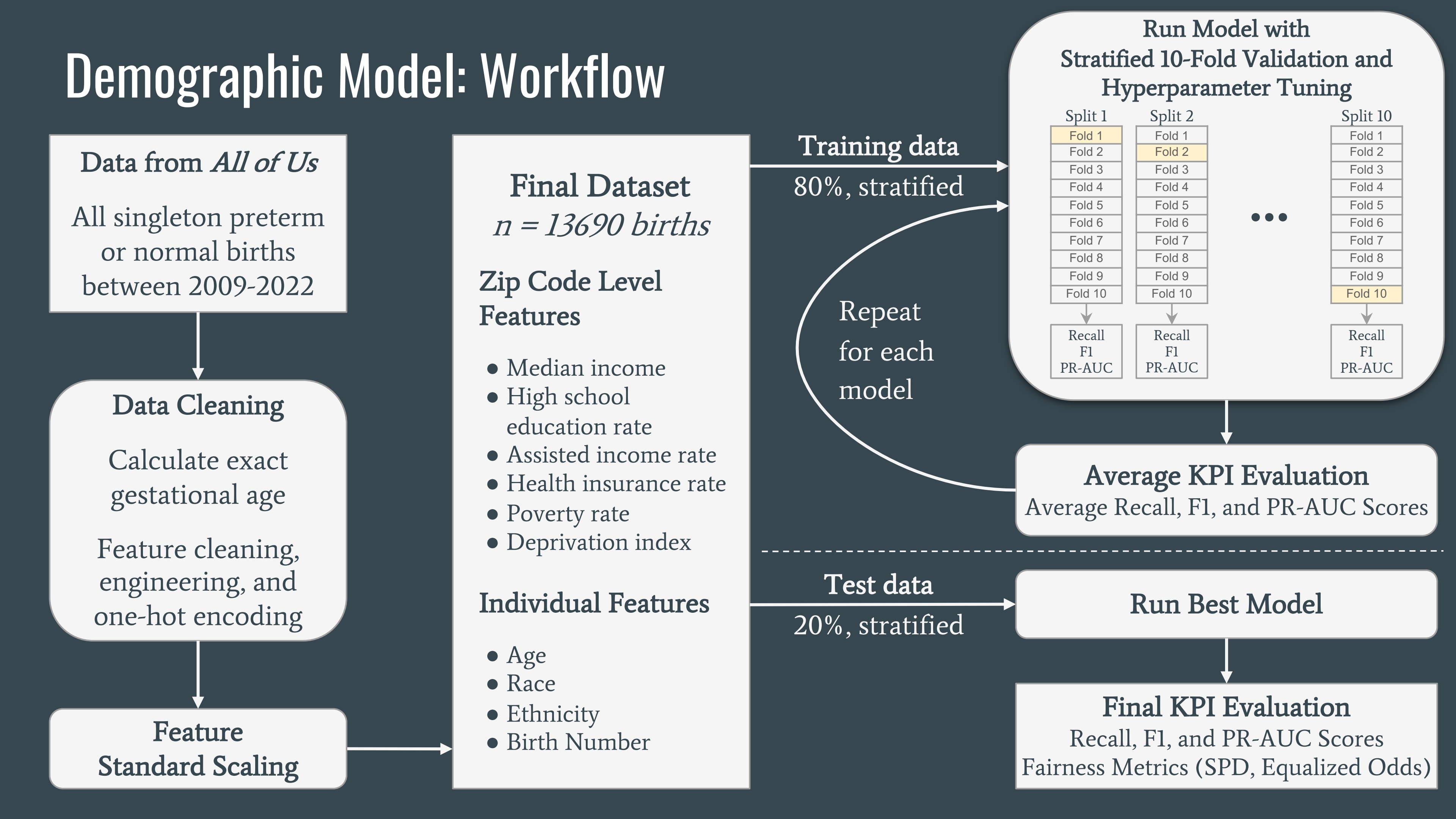
Chose to minimize false negatives and accept more false positives

KPIs for evaluating the model:

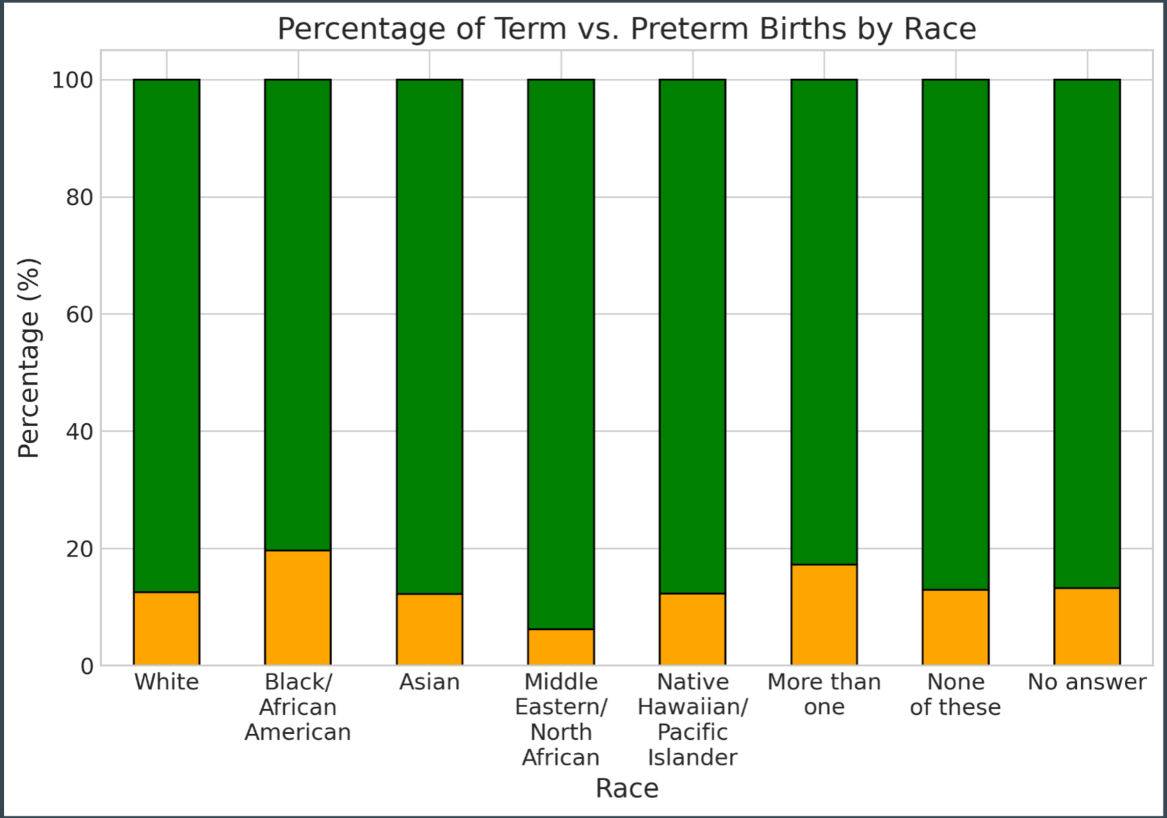
- Recall
- F1
- Precision Recall - Area Under Curve (PR-AUC)

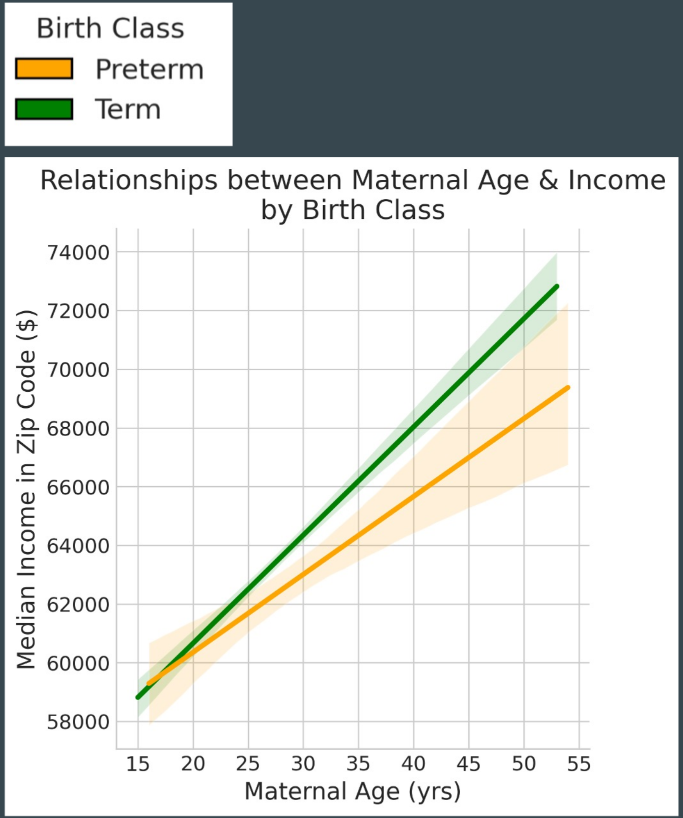
KPIs for evaluating fairness across demographic groups:

- Statistical Parity Difference (SPD)
- Equalized Odds


Demographic Model: Workflow
Data from *All of Us*
All singleton preterm or normal births between 2009-2022
Data Cleaning
Calculate exact gestational age
Feature cleaning, engineering, and one-hot encoding
Feature Standard Scaling
Final Dataset
n = 13690 births
Zip Code Level Features
Median income
High school education rate
Assisted income rate
Health insurance rate
Poverty rate
Deprivation index
Individual Features
Age
Race
Ethnicity
Birth Number
Training data
80%, stratified
Repeat for each model
Run Model with Stratified 10-Fold Validation and Hyperparameter Tuning
Split 1
Split 2
Split 10
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Fold 6
Fold 7
Fold 8
Fold 9
Fold 10
Recall
F1
PR-AUC
Average KPI Evaluation
Average Recall, F1, and PR-AUC Scores
Test data
20%, stratified
Run Best Model
Final KPI Evaluation
Recall, F1, and PR-AUC Scores
Fairness Metrics (SPD, Equalized Odds)

Demographic Models: Exploratory Data Analysis





Demographic Models: Results

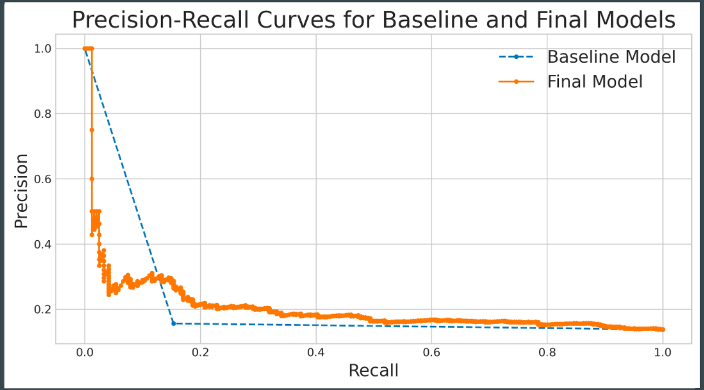
Models explored:

- AdaBoost Decision Tree (n=50)
- SVC with class weights (**0:** 1, **1:** 5.75)

Best results: SVC with class weights

Challenges:

- Primarily zip code level data
- Over 6x more term births than preterm births

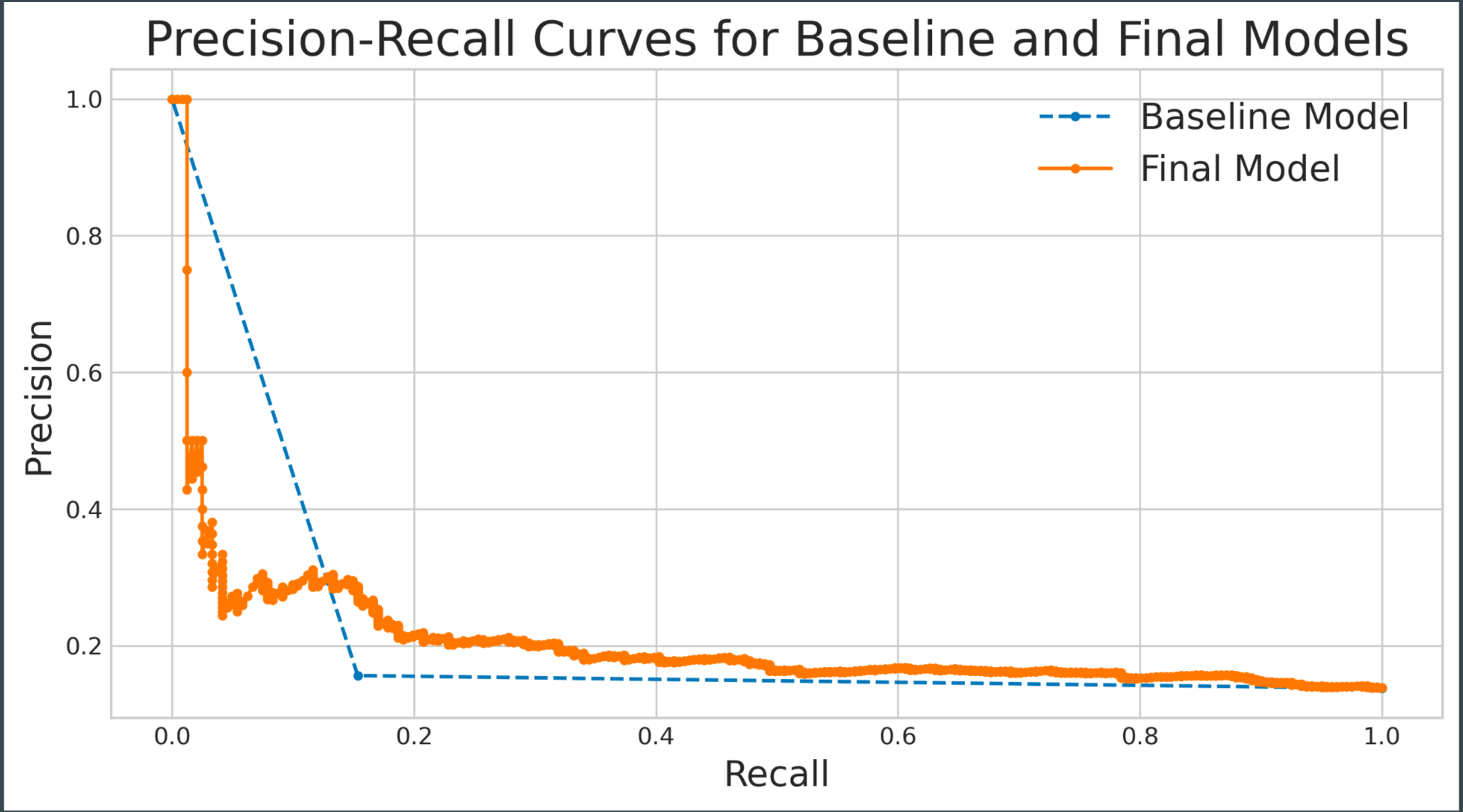


Key Performance Indicators:

	Best Model	*Baseline*
Recall	0.413	0.145
F1	0.242	0.137
PR-AUC	0.172	0.192
SPD	*	
Eq. Odds	0.0	

** The SPD metrics of predictions mirrored training dataset disparities.*

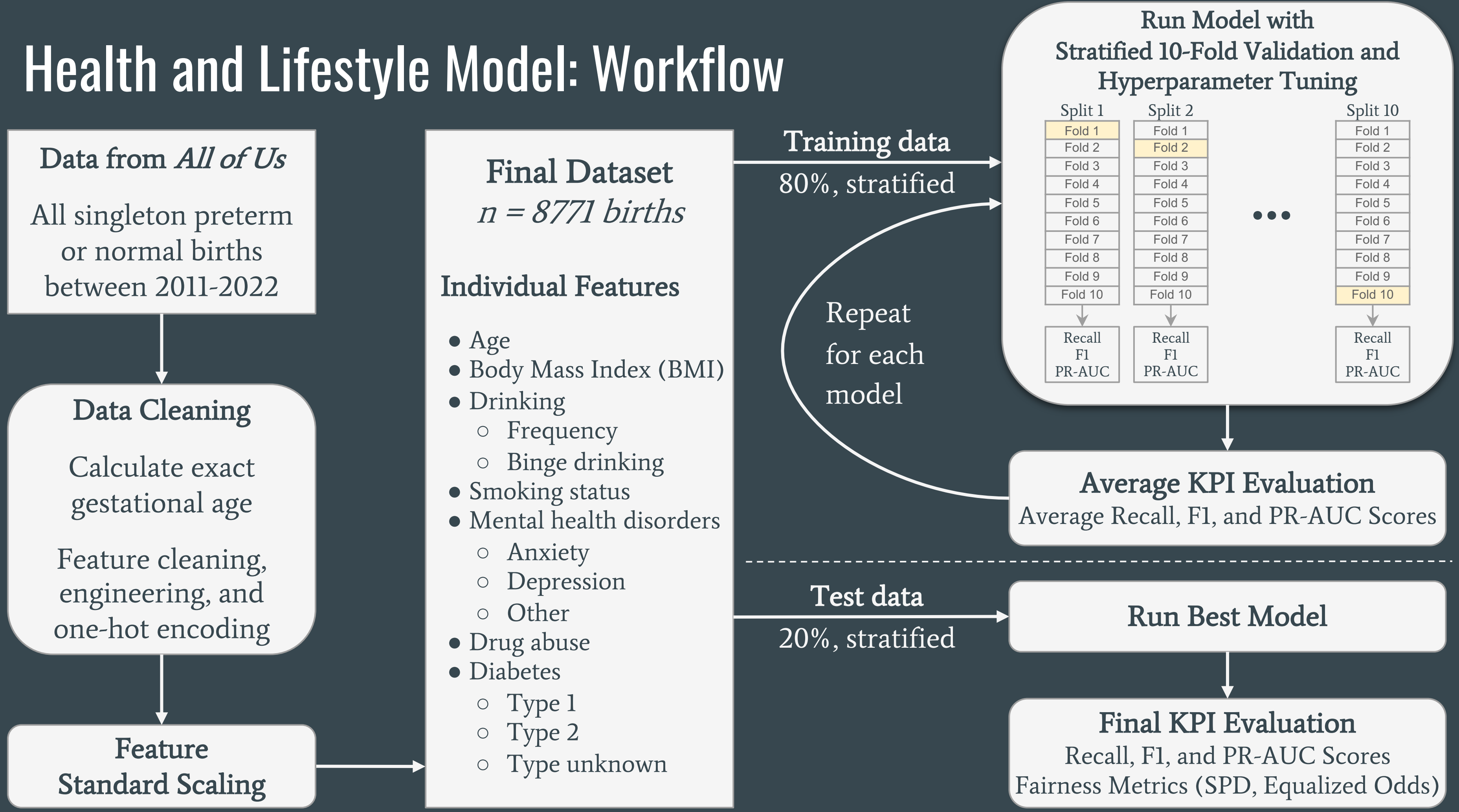
Demographic Models: Results



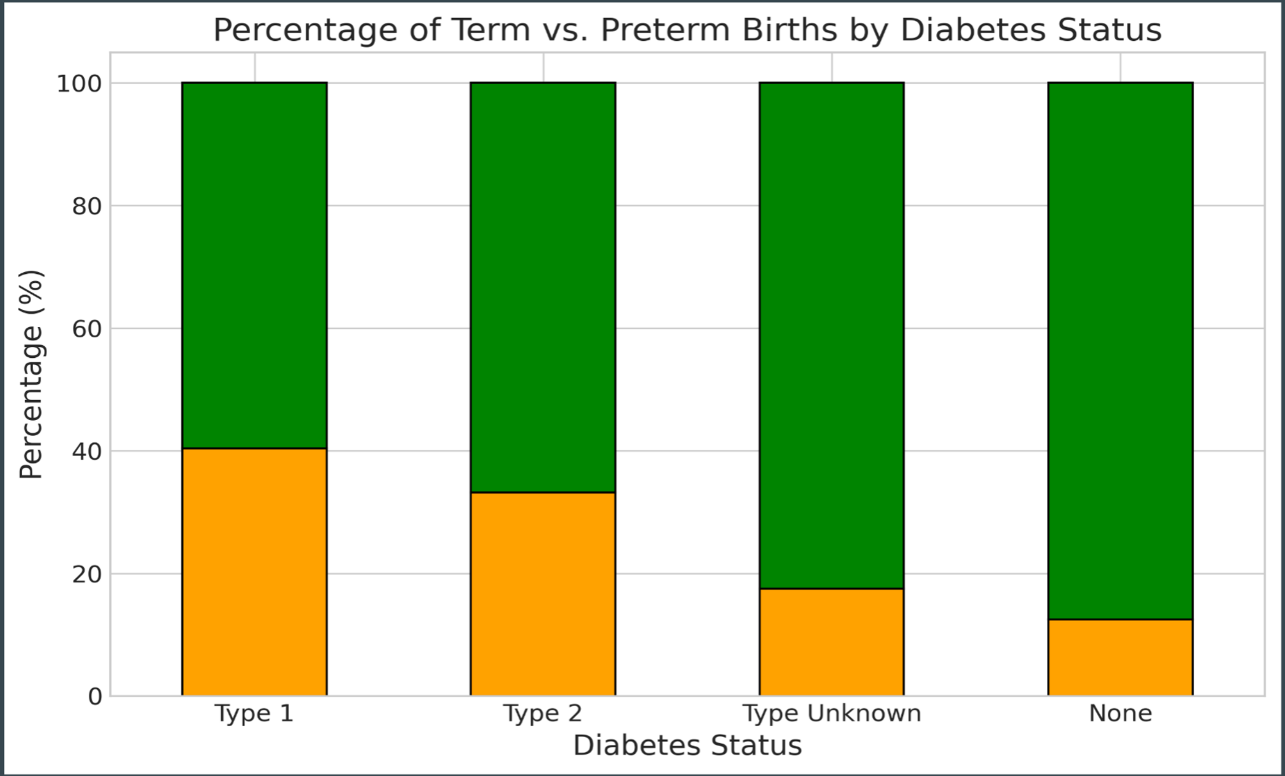
Demographic Models: Statistical Parity Difference (SPD)

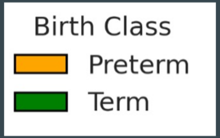
Racial/Ethnic Group (privileged group: White)	**'Ground Truth' Entire Dataset**	**Test Predictions**
Black or African American	-0.071	-0.065
Asian	0.003	0.000
Middle Eastern or North African	0.063	0.021
Native Hawaiian or Other Pacific Islander	0.003	-0.042
More than one	-0.047	-0.073
None of these	-0.004	0.041
No answer	-0.006	-0.007

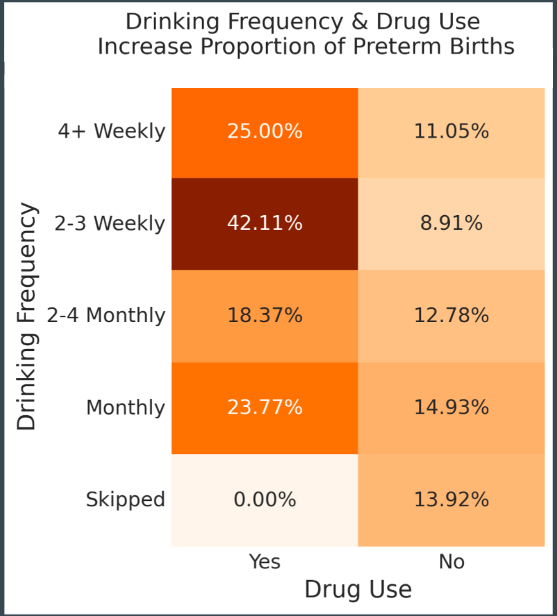
Health and Lifestyle Model: Workflow



Health and Lifestyle Models: Exploratory Data Analysis







Health and Lifestyle Models: Results

Models explored:

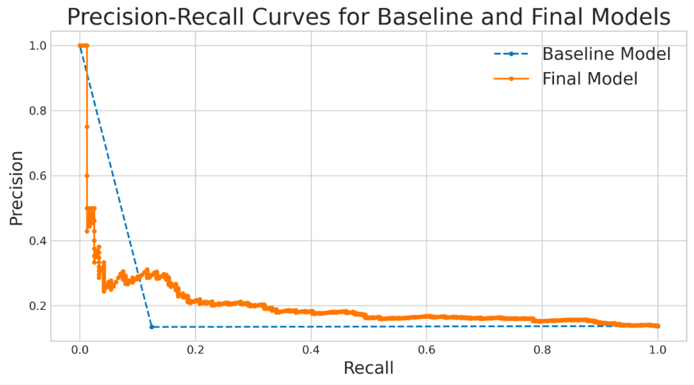
- Logistic regression with class weights (**0:** 1, **1:** 6.33)
- Logistic regression with ROSE (Random Over-Sampling Examples) oversampling
- SVC with class weights (**0:** 1, **1:** 6.33)

Best results: Logistic regression with class weights

Challenges:

- Over 6x more term births than preterm births
- Features not necessarily representative of status during pregnancy

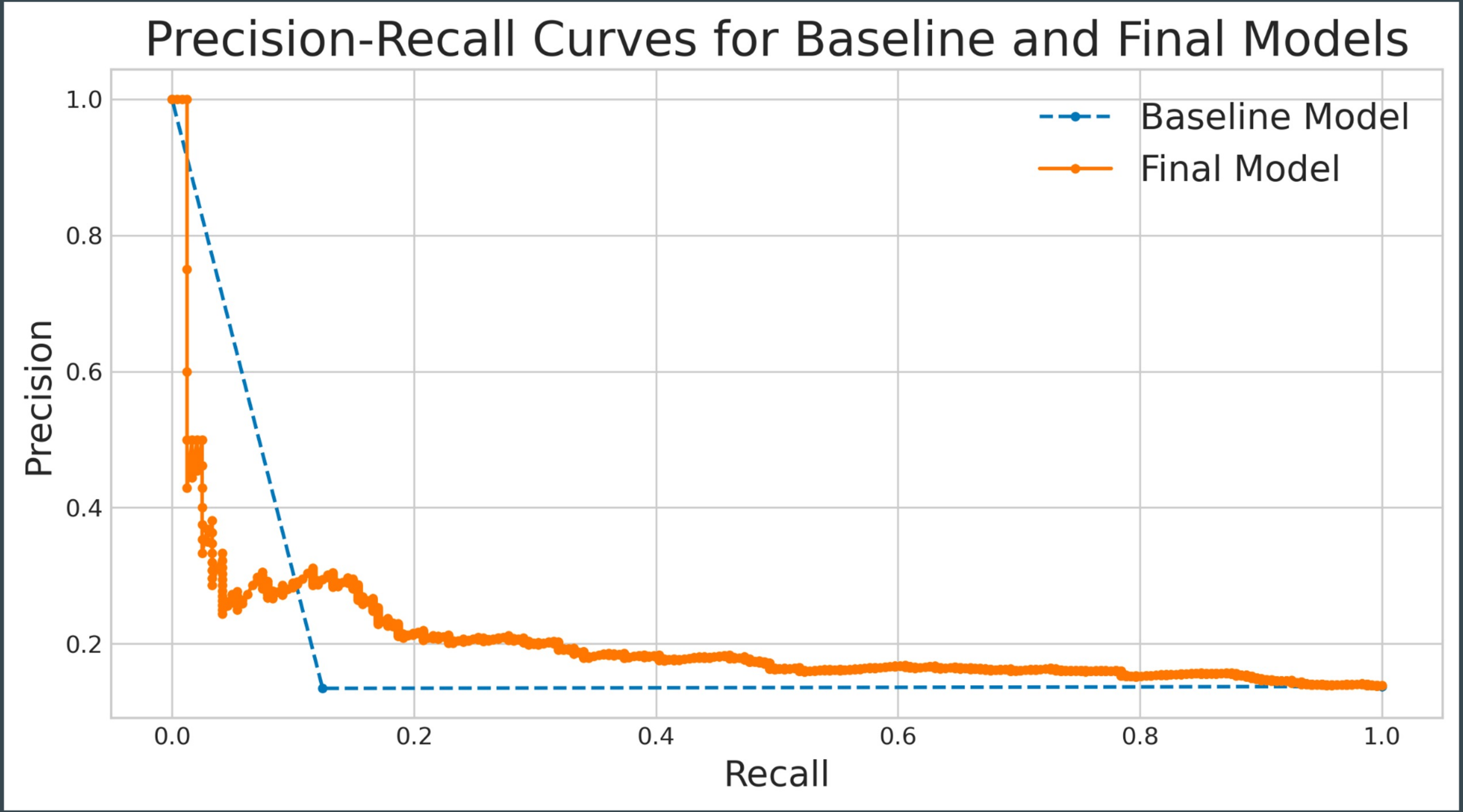
** The SPD metrics of predictions mirrored training dataset disparities.*



Key Performance Indicators:

	Best Model	*Baseline*
Recall	0.473	0.137
F1	0.247	0.136
PR-AUC	0.197	0.196
SPD	*	
Eq. Odds	0.0	

Health and Lifestyle Models: Results



Health and Lifestyle Models: Statistical Parity Difference (SPD)

Racial/Ethnic Group (privileged group: White)	**'Ground Truth' Entire Dataset**	**Test Predictions**
Black or African American	-0.081	-0.103
Asian	-0.007	-0.053
Middle Eastern or North African	0.042	0.107
Native Hawaiian or Other Pacific Islander	-0.026	-0.226
More than one	-0.057	-0.026
None of these	-0.049	-0.179
No answer	-0.018	-0.032

Conclusions and Next Steps

Our models met, but did not outperform, the baseline models.

- A previous study using EHR data only had similar recall and precision values[5]

Challenges:

- Individual predictions are limited when input data is on the community level
- Highly imbalanced class structure not easily fixed by oversampling methods
- Complexity of preterm birth may require features from multiple domains[2]

Future Work:

- High quality, individual level data collection needed
- Incorporate features from demographics, health records, genetics
- Engineer interaction terms, empowered by a robust dataset

Thanks to...

- Roman Holowinsky, Matt Osborne, Alec Clott, and
  The Erdős Institute for their support during this boot camp
- Evelyn Huszar for her mentorship and advice on this project
- The NIH *All of Us* program for collecting, anonymizing, and allowing us to use the data for this project:

> The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

References

1. Centers for Disease Control and Prevention. (2023, October 24). Preterm birth. Centers for Disease Control and Prevention. https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm.
2. Goldenberg RL, Culhane JF, Iams JD, and Romero R. Epidemiology and causes of preterm birth. The Lancet. 2008; 371(9606):75–84. doi: 10.1016/S0140-6736(08)60074-4.
3. Włodarczyk T, Płotka S, Szczepański T, Rokita P, Sochacki-Wójcicka N, Wójcicki J, et al. Machine Learning Methods for Preterm Birth Prediction: A Review. Electronics. 2021; 10(5): 586. doi: 10.3390/electronics10050586.
4. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol. 2014; 14:42. doi: 10.1186/1471-2288-14-42.
5. Mercer BM, Goldenberg RL, Das A, Moawad AH, Iams JD, Meis PJ, et al. The preterm prediction study: a clinical risk assessment system. Am J Obstet Gynecol. 1996; 174(6):1885-95. doi: 10.1016/s0002-9378(96)70225-9.