Responsible AI to Benefit Society

A Survey of Use Cases and Lessons Learned

Kit Rodolfa

IDHS
Illinois Department of Human Services
FEEDING AMERICA
Department of Labor – United States of America
VANCOUVER PUBLIC SCHOOLS
VPS
NOBLE
WAKE COUNTY PUBLIC SCHOOL SYSTEM
Rijkswaterstaat
Ministerie van Infrastructuur en Milieu
Perspectives Charter Schools
KIPP: New Jersey
KIPP: Chicago
COLLEGE PREP PUBLIC SCHOOLS
INFONAVIT
Hogares con valor
cclba
city of CINCINNATI
PERFORMANCE & DATA ANALYTICS
POLICE
CHARLOTTE-MECKLENBURG
THE WHITE HOUSE
WASHINGTON
Montgomery County Public Schools
CHICAGO PUBLIC SCHOOLS
CPS
Syracuse
TULSA
PUBLIC SCHOOLS
CHICAGO ALLIANCE TO END HOMELESSNESS
INSTITUTE FOR HOUSING STUDIES
City of Memphis
WORLD BANK GROUP
ChapinHall at the University of Chicago
MuskingumValley
Educational Service Center
CITY OF CHICAGO
INCORPORATED 4TH MARCH 1837
cta
METRO POLICE
NASHVILLE DAVIDSON COUNTY
SEAL OF COOK COUNTY ILLINOIS
BERKELEY LAB
Lawrence Berkeley National Laboratory
Get Covered Illinois
The Official Health Marketplace
CHICAGO HARRIS
PUBLIC POLICY | THE UNIVERSITY OF CHICAGO
AGENTIS energy
AUSTRALIAN CONSERVATION FOUNDATION
spire
DigitalGlobe
WORLD ECONOMIC FORUM
planet
New Vision for the Ocean
Pecan Street Inc.
Arlington Public Schools
SUNLIGHT FOUNDATION
THE CASE F
Health Leads
MÉXICO
PRESIDENCIA DE LA REPÚBLICA
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
JOSÉ DE MELLO · SAÚDE
Wiki Energy
mesa
PUBLIC SCHOOLS
HEALTHY CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH
Ushahidi
careerbuilder
NorthShore
ENROLL AMERICA
Nurse-Family Partnership
Helping First-Time Parents Succeed
SKILLS FOR CHICAGOLAND'S FUTURE
DIVVY
CONSERVATION INTERNATIONAL
CDOT
CHICAGO DEPARTMENT OF TRANSPORTATION
City of Milwaukee
JOHNSON COUNTY
KANSAS
NEW YORK STATE OF OPPORTUNITY
Department of Environmental Conservation
City of Rotterdam
SANERGY
SEDESOL
SECRETARÍA DE DESARROLLO SOCIAL

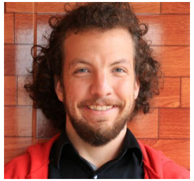
More details on projects at http://dssgfellowship.org/projects













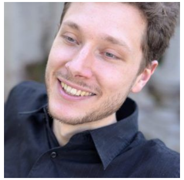










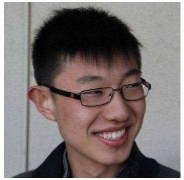




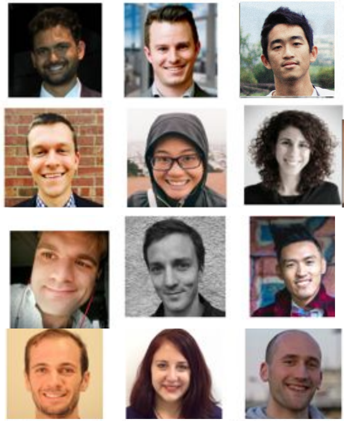






~275 DSSG Fellows



OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

Preventing and Reducing Water Mains breaks (Syracuse, NY)

Using Machine Learning to Predict and Prevent Water Mains Breaks. Kumar et al KDD 2015

240,000 main breaks/yr in US

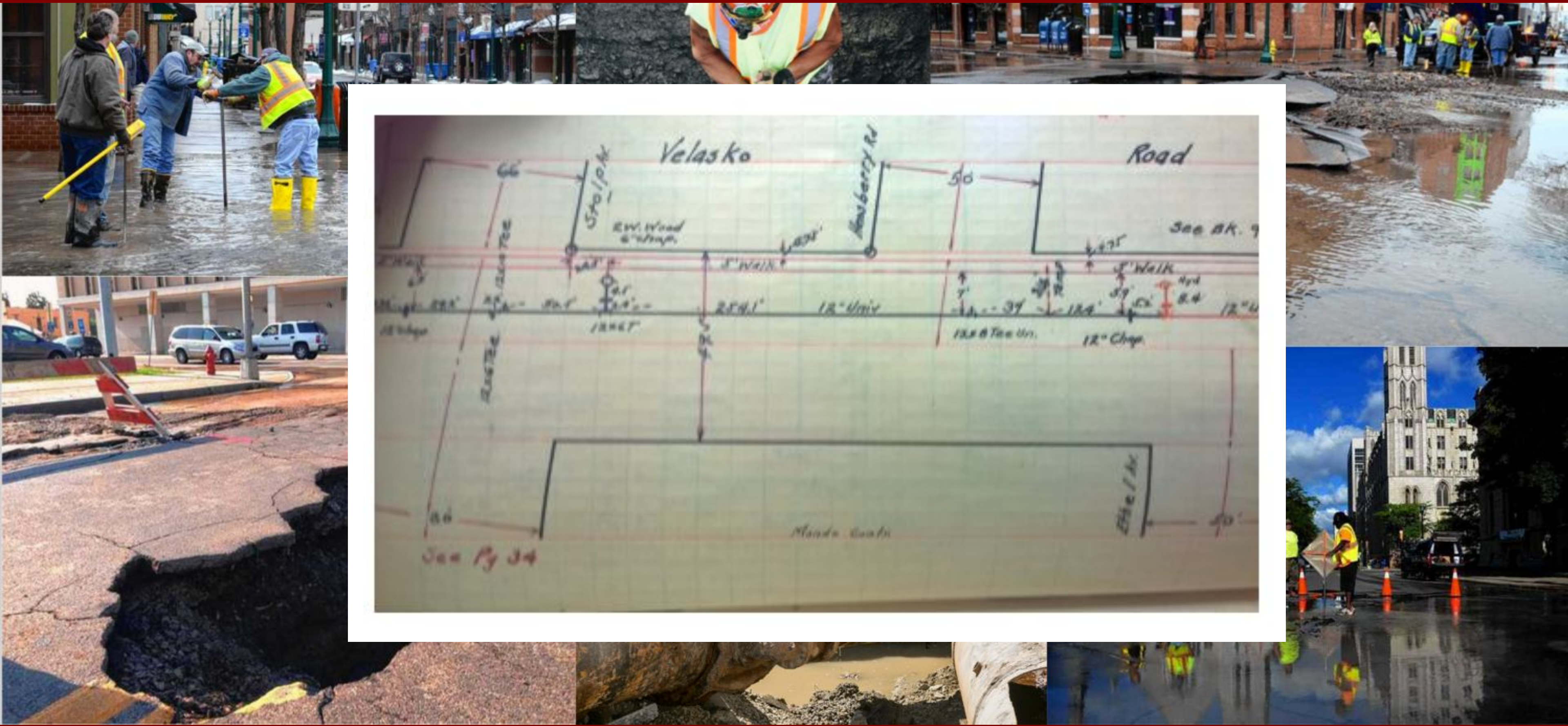
$13 billion in 2010 to repair

Expected $30 billion by 2040

180 breaks/yr in Syracuse, NY

Preventing and Reducing Water Mains breaks (Syracuse, NY)

Using Machine Learning to Predict and Prevent Water Mains Breaks. Kumar et al KDD 2015



Kit Rodolfa | krodolfa@cmu.edu

Carnegie Mellon University

Preventing and Reducing Water Mains breaks (Syracuse, NY)

Using Machine Learning to Predict and Prevent Water Mains Breaks. Kumar et al KDD 2015

64% of blocks in the top 1% of predictions were correctly predicted for last year

LESSON 1

Be diligent in finding relevant data

Most data in real contexts is spread throughout the organization and much relevant data may not even be digitized

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

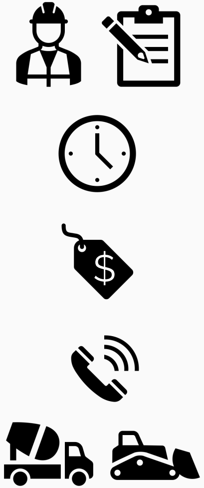
Manual Labeling Effort: Time Consuming & Costly

2400 blocks

Only 2018

blocklot	roof_damage	roof_damage_score
1451 024	High	50-99

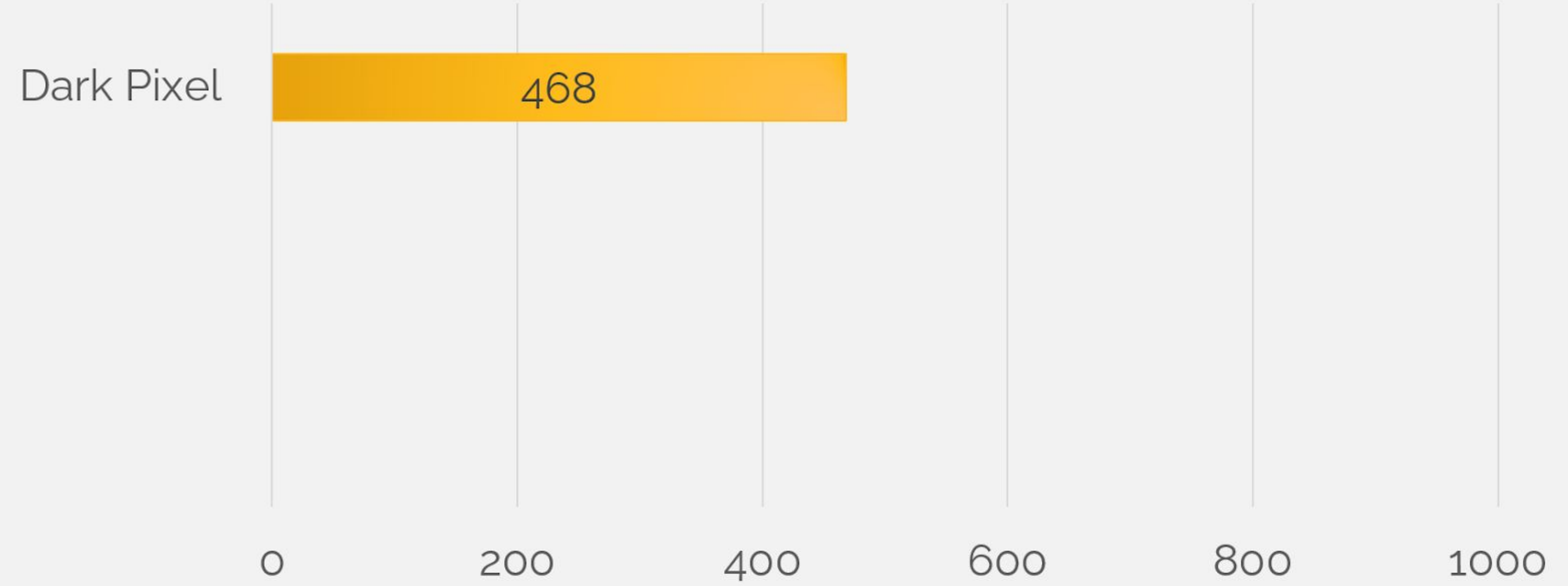
LIST OF 1000 ROWHOMES FOR INSPECTION



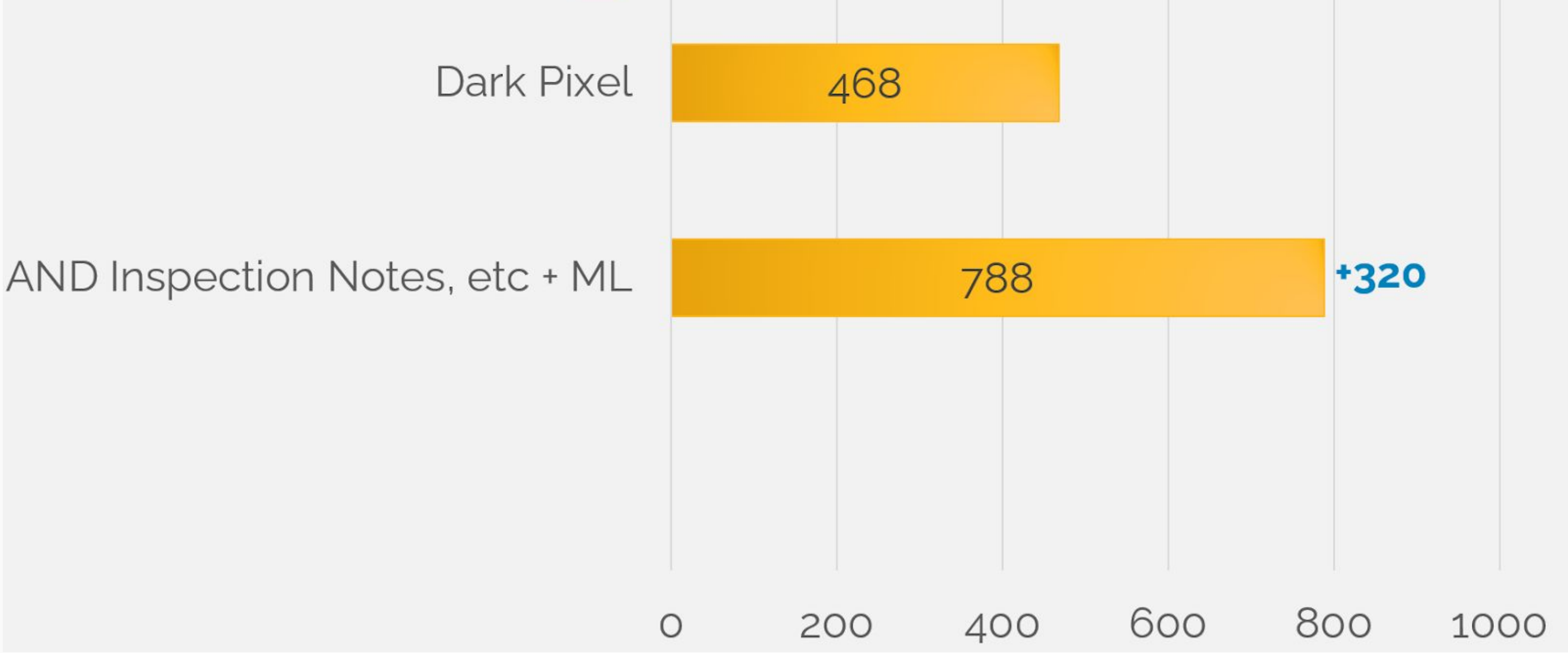
Address	Roof Damage score
14 Columbia Blvd.	0.9
3939 Calvin St.	0.7
1656 Benson St.	0.4
374 Margaret St.	0.1

DARK PIXEL

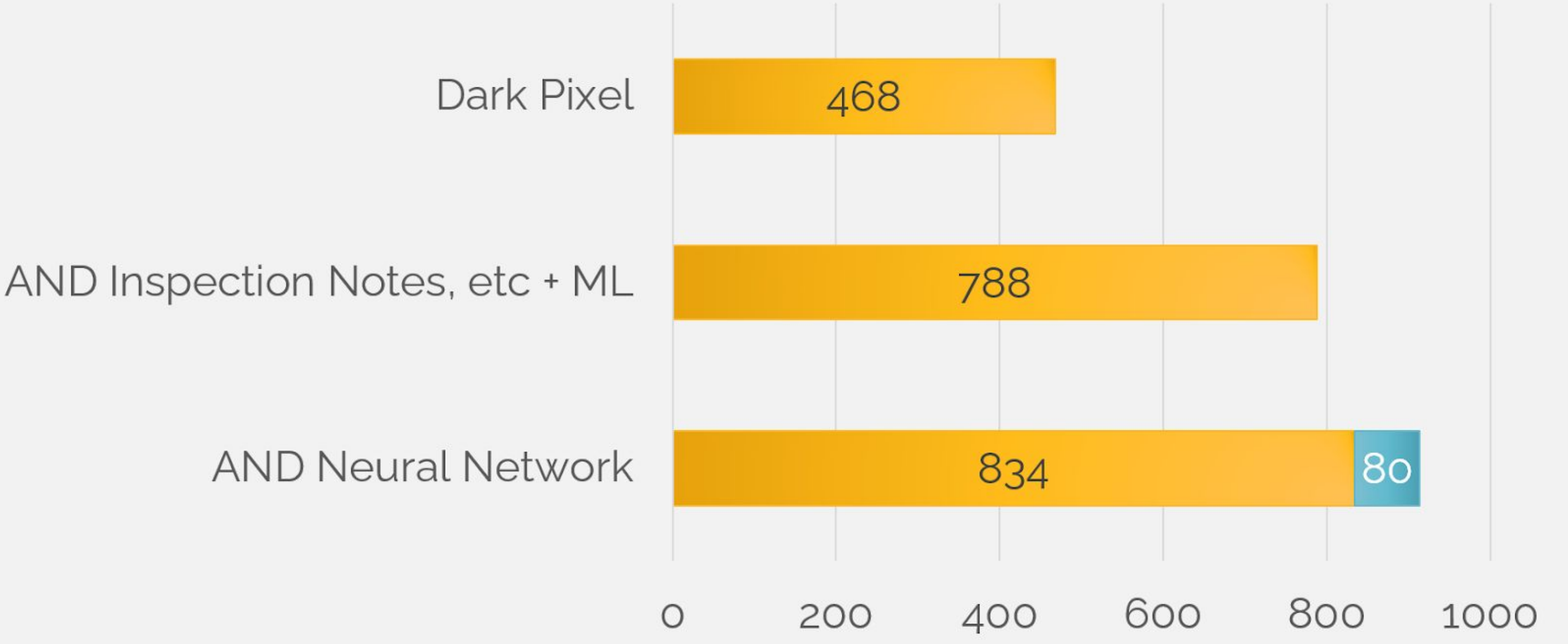
What percentage of the pixels are darker than ?



MODEL PERFORMANCE



MODEL PERFORMANCE



MODEL PERFORMANCE

LESSON 2

Models will benefit from a range of data types and sources

Even with a seemingly straightforward vision problem, administrative data helped the model learn the nuance of the task

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

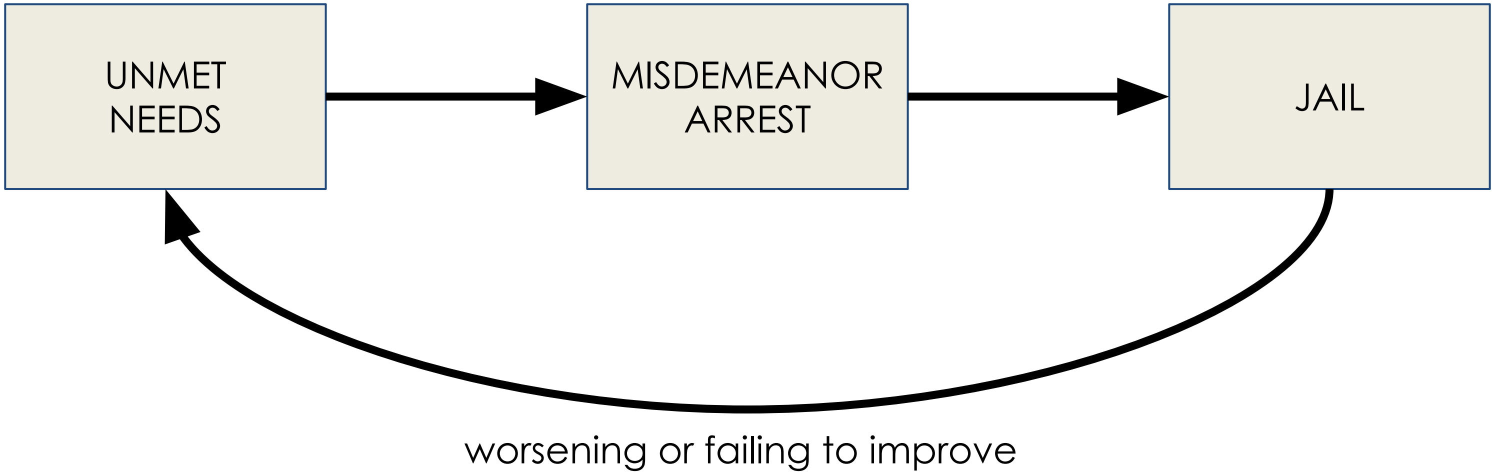
Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

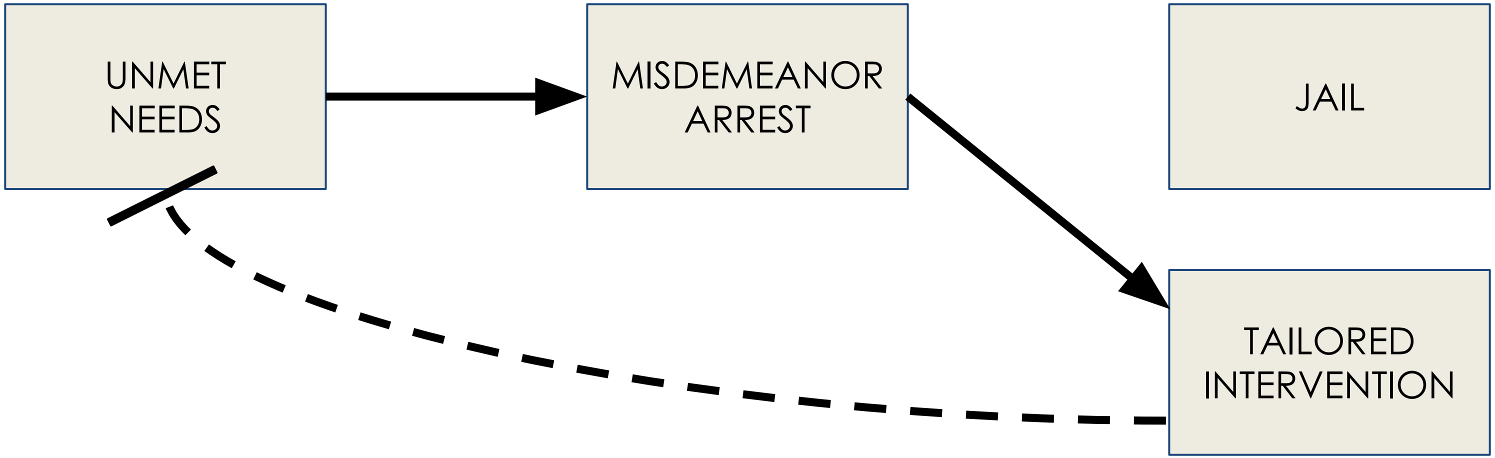
BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

Cycle of Incarceration

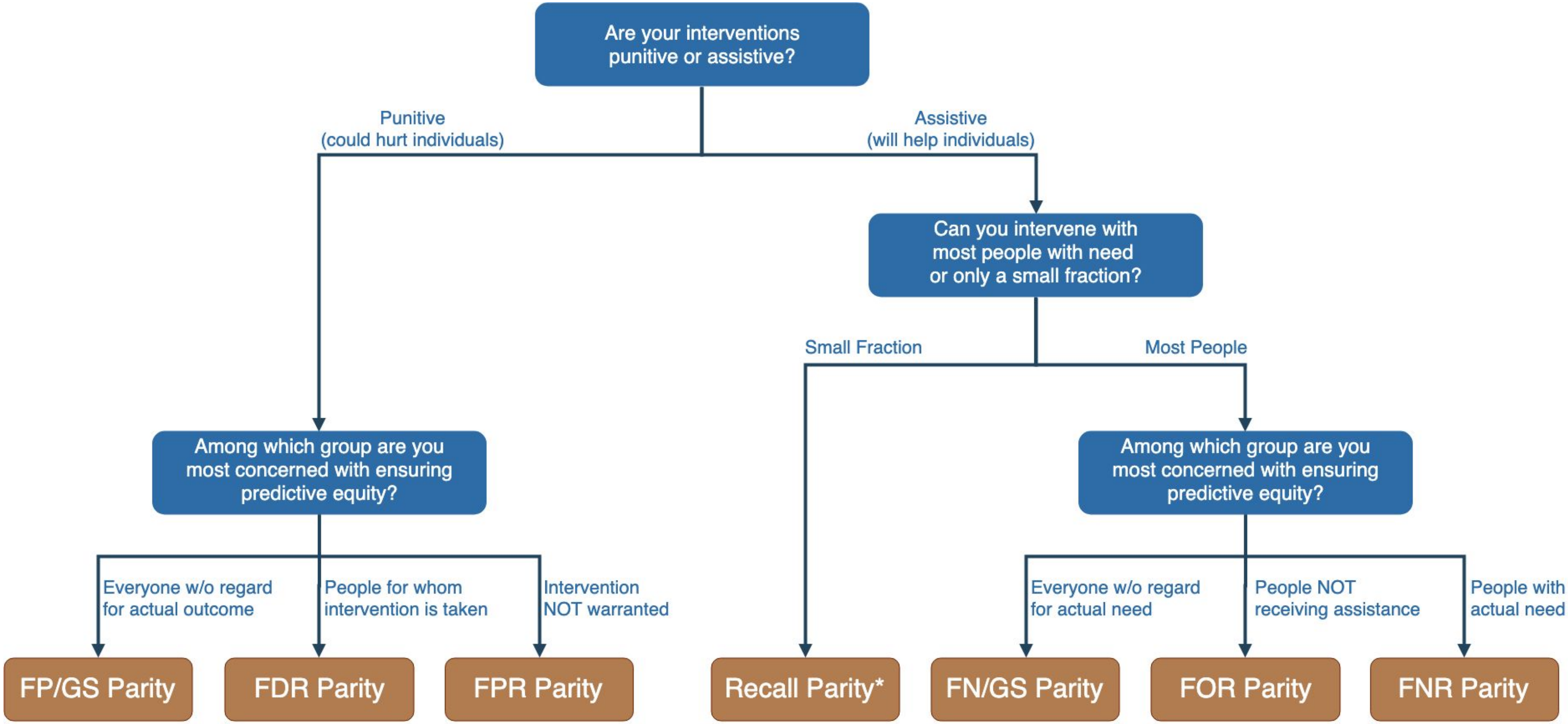


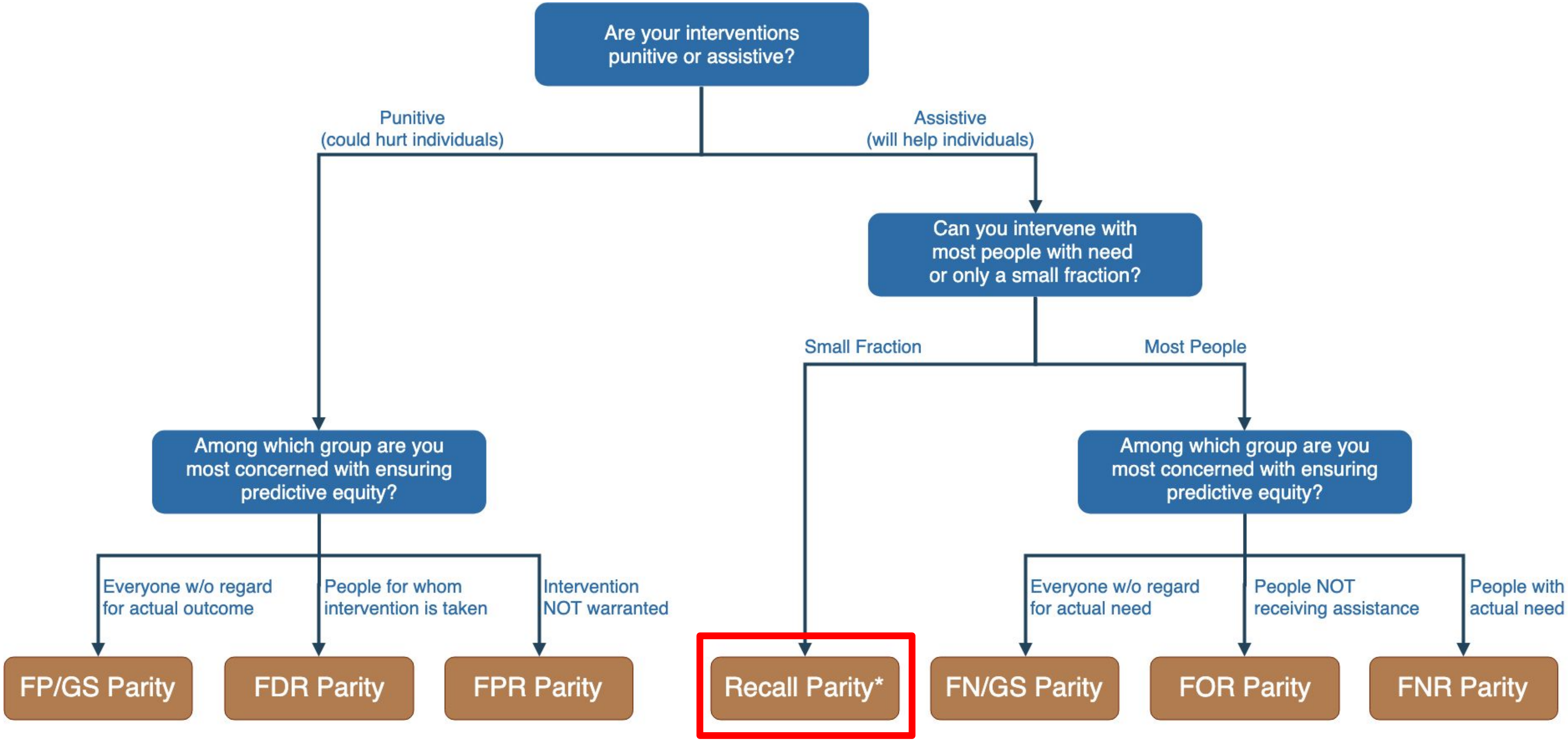
Breaking the Cycle

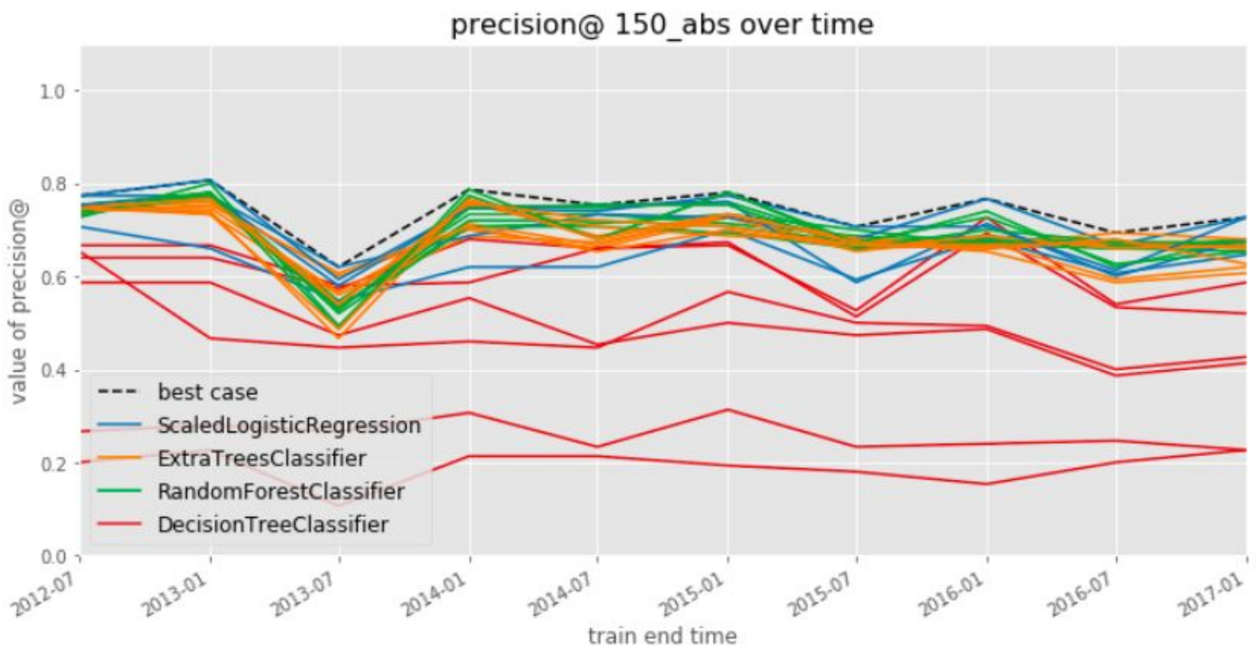


Three Key Questions

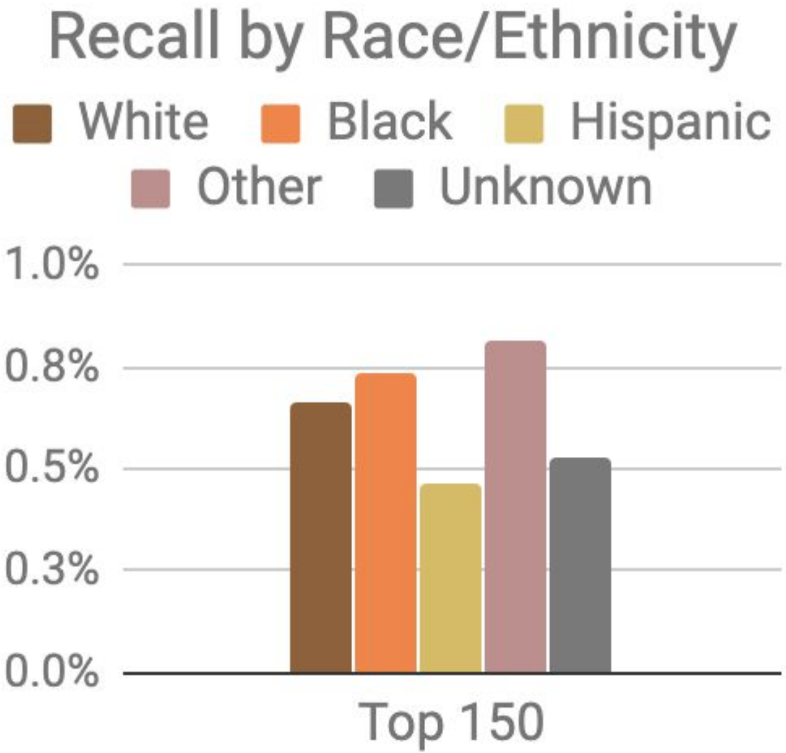
- How do we **define equity** in a given policy context?
- How can we **improve equity** of ML models and implementation?
- How do policy makers **balance trade-offs** between equity, efficiency, resources?


Are your interventions punitive or assistive?
Punitive (could hurt individuals)
Assistive (will help individuals)
Can you intervene with most people with need or only a small fraction?
Small Fraction
Most People
Among which group are you most concerned with ensuring predictive equity?
Among which group are you most concerned with ensuring predictive equity?
Everyone w/o regard for actual outcome
People for whom intervention is taken
Intervention NOT warranted
Everyone w/o regard for actual need
People NOT receiving assistance
People with actual need
FP/GS Parity
FDR Parity
FPR Parity
Recall Parity*
FN/GS Parity
FOR Parity
FNR Parity


Are your interventions punitive or assistive?
Punitive (could hurt individuals)
Assistive (will help individuals)
Can you intervene with most people with need or only a small fraction?
Small Fraction
Most People
Among which group are you most concerned with ensuring predictive equity?
Among which group are you most concerned with ensuring predictive equity?
Everyone w/o regard for actual outcome
People for whom intervention is taken
Intervention NOT warranted
Everyone w/o regard for actual need
People NOT receiving assistance
People with actual need
FP/GS Parity
FDR Parity
FPR Parity
Recall Parity*
FN/GS Parity
FOR Parity
FNR Parity

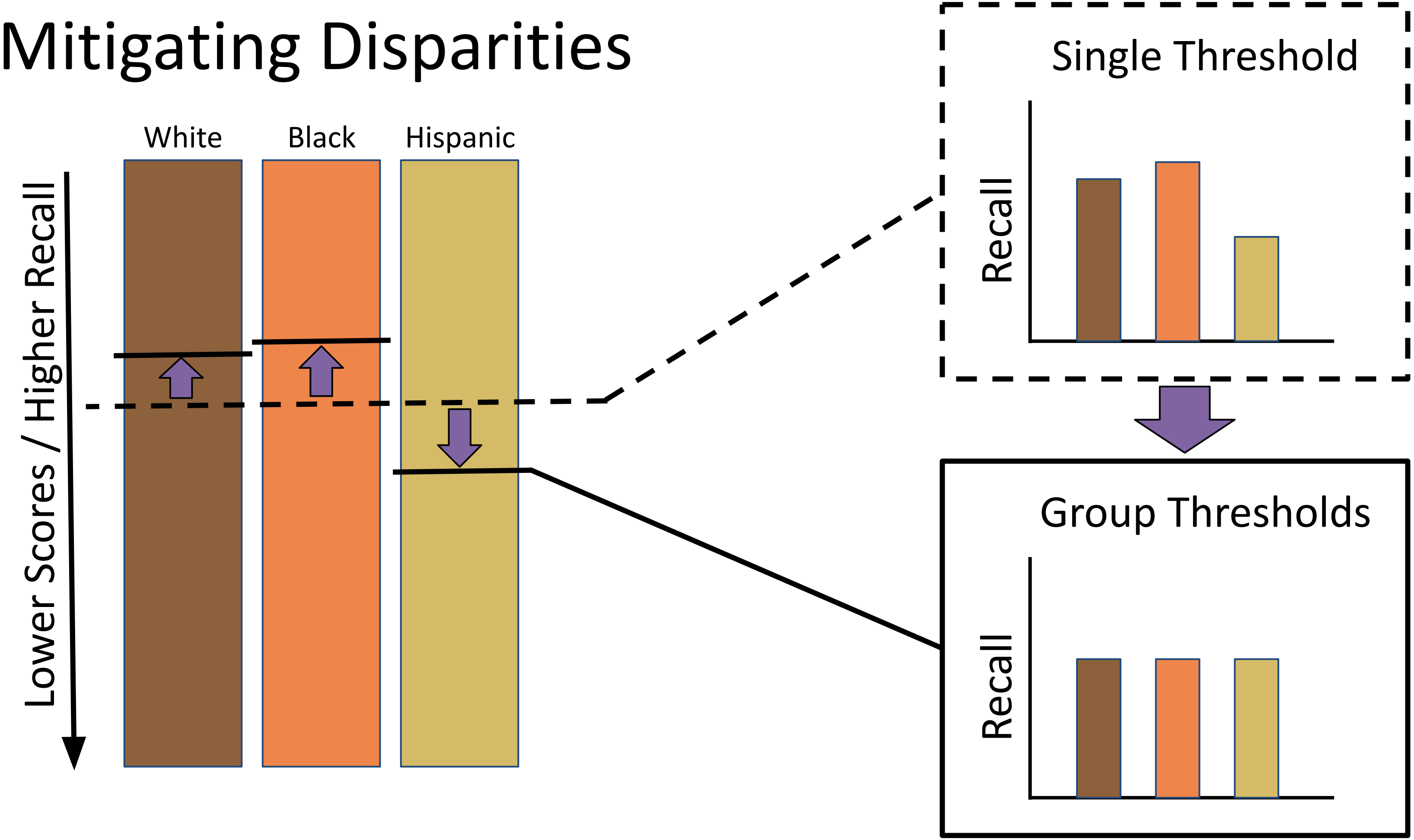


We built and selected a model to choose the 150 highest-risk individuals for intervention...



- Model was optimized for efficiency, not equity
- Top 150 highest risk reasonably balanced between black and white individuals
- However, hispanic and unknown race/ethnicity groups very underrepresented

Mitigating Disparities



Menu of Options

	Current Scale	Expanded Scale
No Constraint		
Equalize Recall		
Reduce Disparities		

Menu of Options

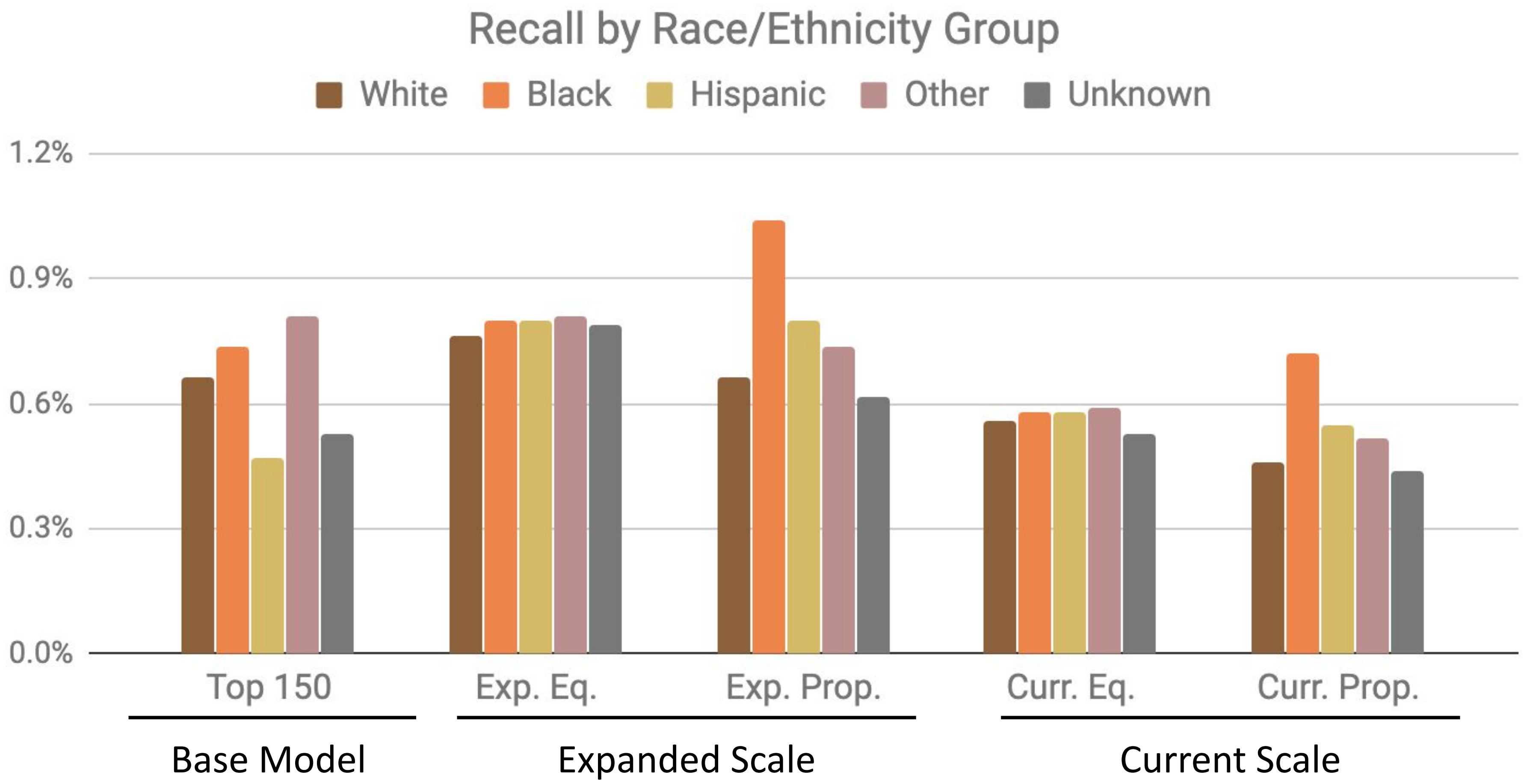
	Current Scale	Expanded Scale
No Constraint	BASE MODEL	
Equalize Recall		
Reduce Disparities		

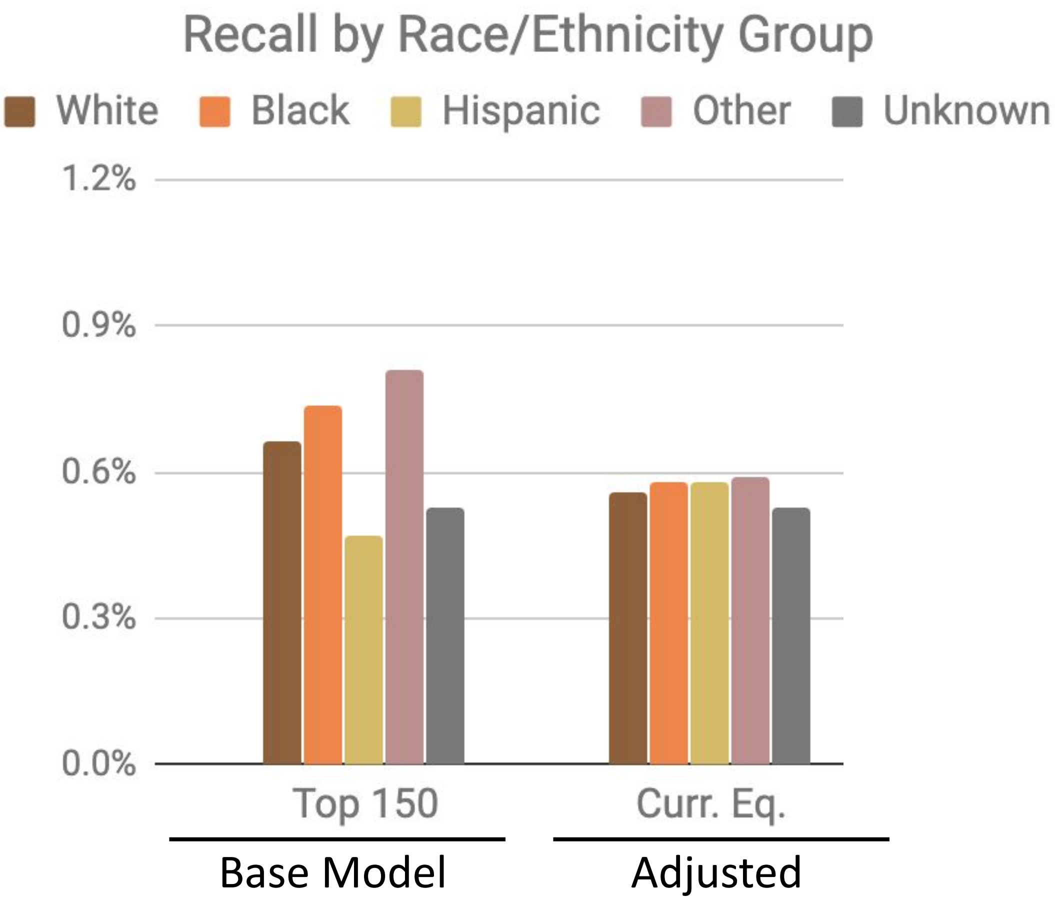
Menu of Options

	Current Scale	Expanded Scale
No Constraint		
Equalize Recall / Reduce Disparities	EXPLICIT EQUITY / EFFICIENCY TRADE-OFF	"COST OF EQUITY"

Menu of Options

	Current Scale	Expanded Scale
No Constraint		
Equalize Recall	IMPROVE OUTCOMES AT SAME RATE ACROSS GROUPS	
Reduce Disparities	IMPROVE OUTCOMES FASTER FOR GROUPS WITH HIGHER UNDERLYING PREVALENCE	


Recall by Race/Ethnicity Group
White
Black
Hispanic
Other
Unknown
1.2%
0.9%
0.6%
0.3%
0.0%
Top 150
Exp. Eq.
Exp. Prop.
Curr. Eq.
Curr. Prop.
Base Model
Expanded Scale
Current Scale


Recall by Race/Ethnicity Group
White
Black
Hispanic
Other
Unknown
1.2%
0.9%
0.6%
0.3%
0.0%
Top 150
Curr. Eq.
Base Model
Adjusted

Little Fairness/Accuracy Trade-Off

Base Model	Adjusted Model
72.7% Precision	70.7% Precision
150 Total Count	150 Total Count

Little Equity/Efficiency Trade-Off at Current Scale

Top 150	Equal Recall	Proportional Recall
72.7% Precision	70.7% Precision	70.7% Precision
150 Total Count	150 Total Count	150 Total Count

LESSON 3

ML Fairness can be achieved (if it is an explicit goal)

In many cases, we've seen little or no trade-off in accuracy when improving fairness, but it needs to be thoughtfully defined, measured, and optimized

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020

Children in at least 4 million U.S. households are exposed to high levels of lead (CDC Report)

Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020

Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

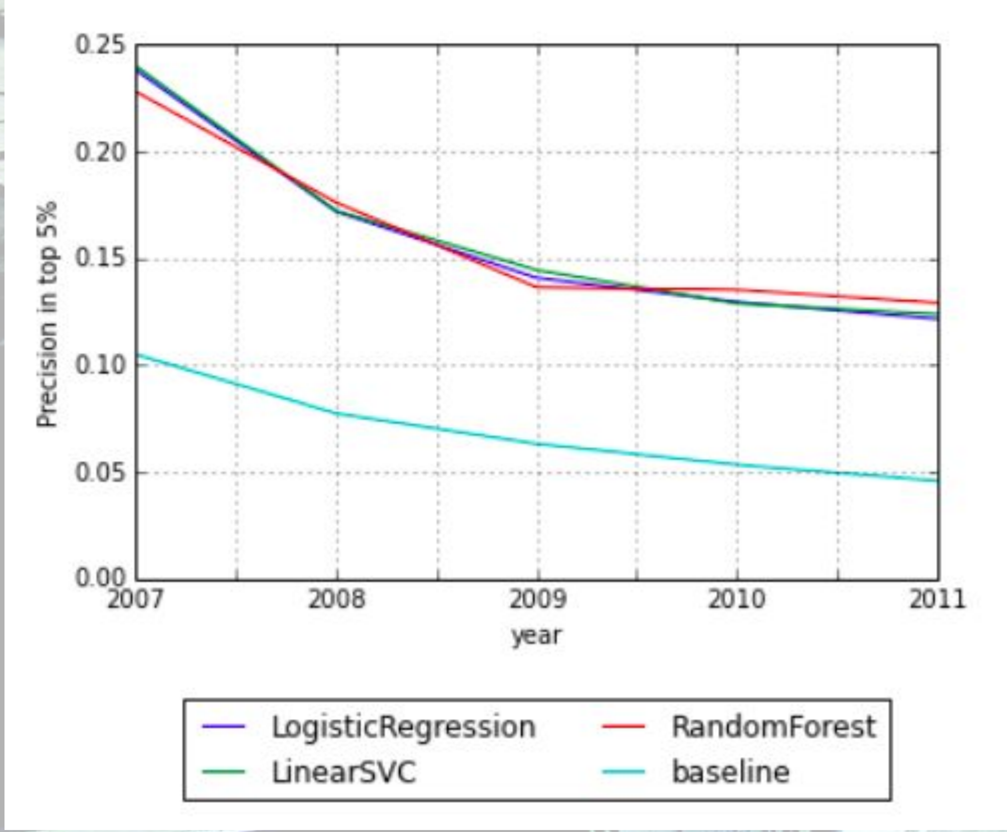
Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020

- Impaired Attention
- Hearing Loss
- Lower IQ
- Lack of Motor Skills
- Learning Disability
- Memory Problems

Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

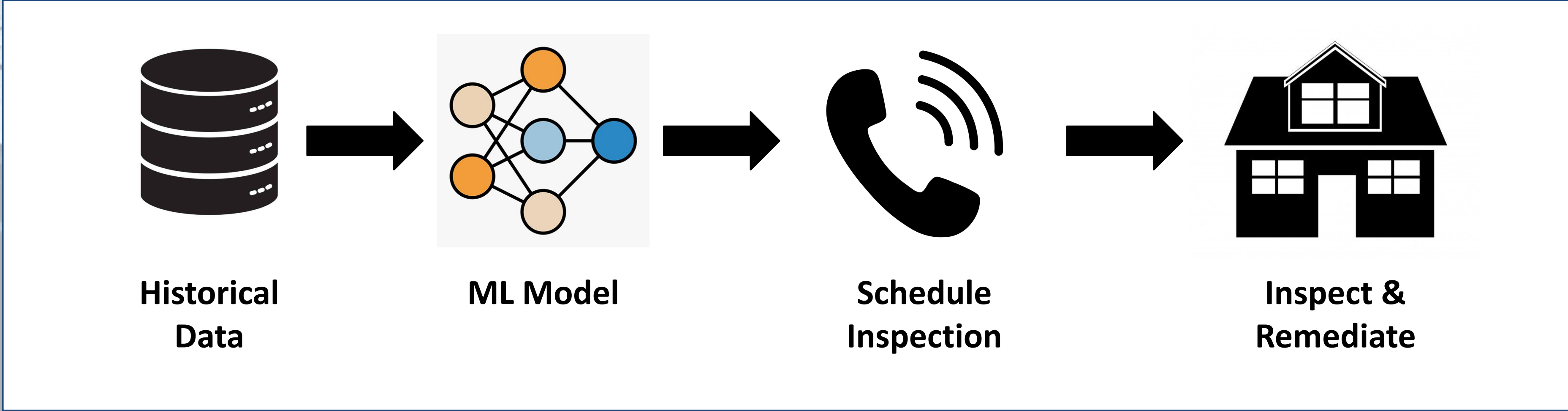
Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020



Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

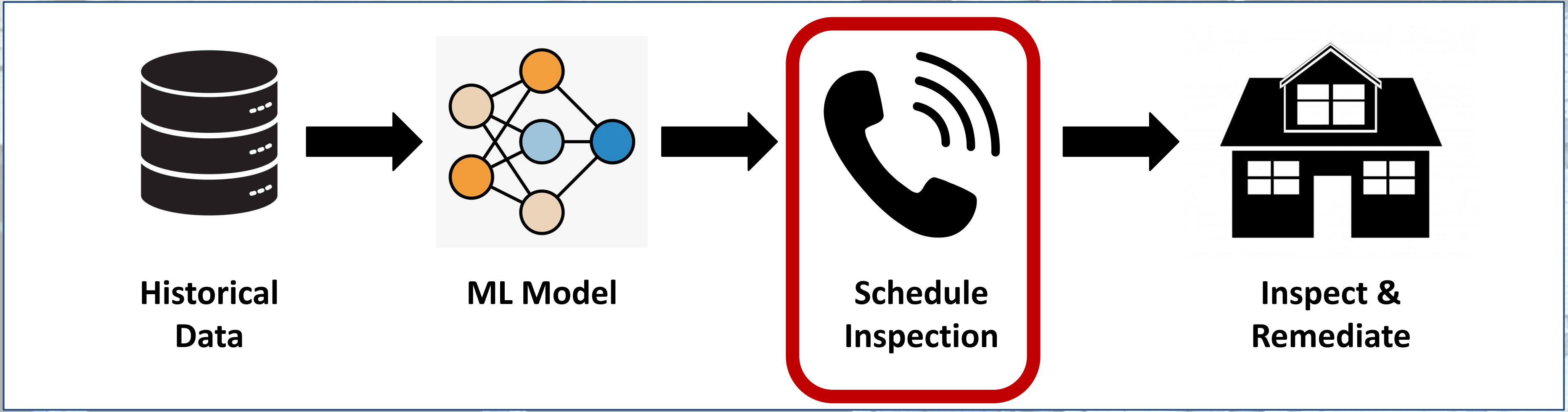
Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020



Preventing Lead Poisoning in Children (Chicago, IL)

Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Potash et al. KDD 2015

Validation of a Machine Learning Prediction Model of Childhood Lead Poisoning. Potash et al. JAMA 2020



LESSON 4

Fairness is a system property

Even with fair model outputs, the implementation matters and it is important to consider how the system as a whole works together to achieve its goals

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

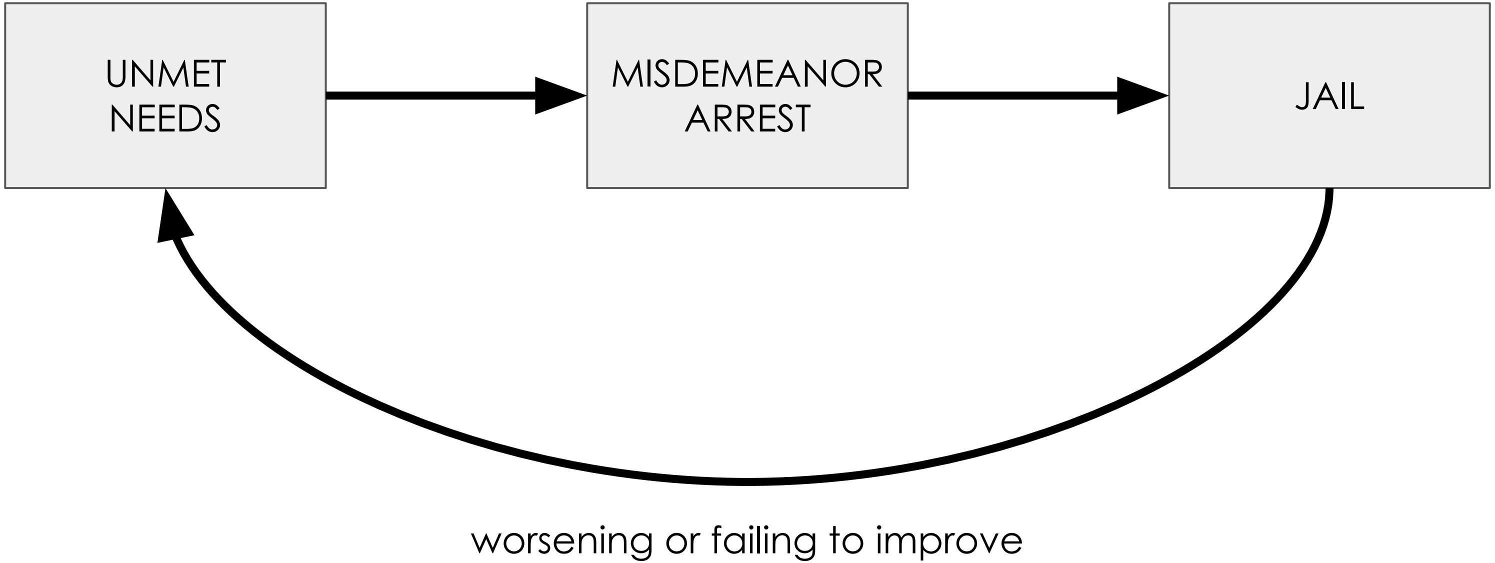
Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

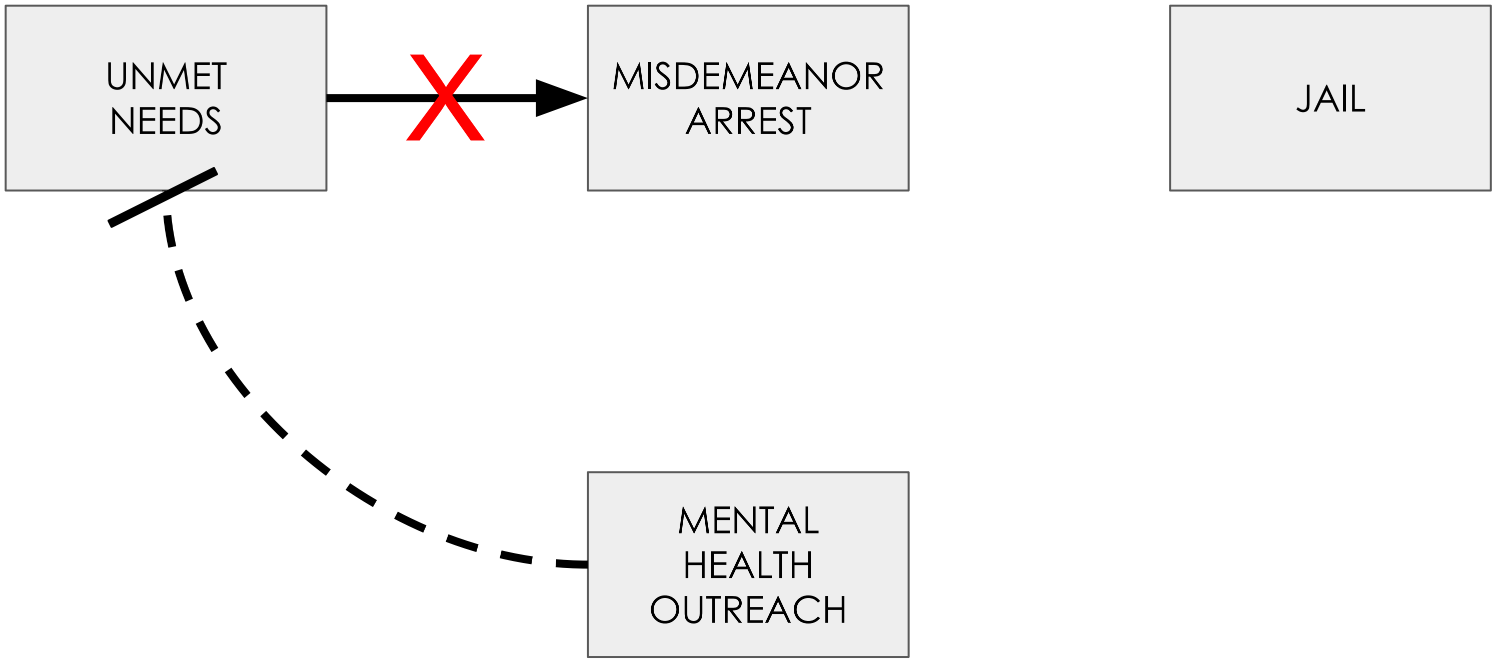
BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

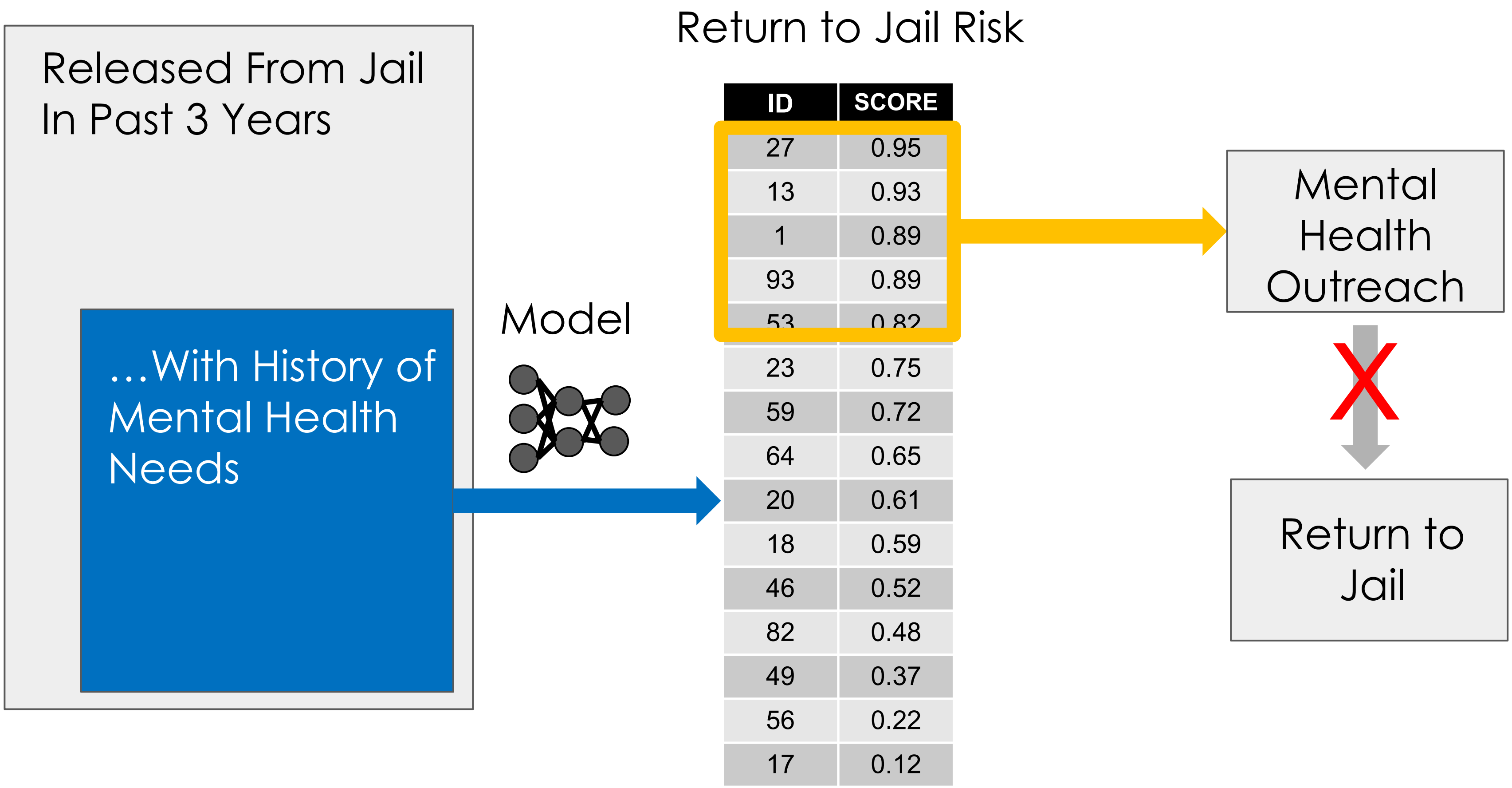
JOHNSON
COUNTY
KANSAS

Cycle of Incarceration

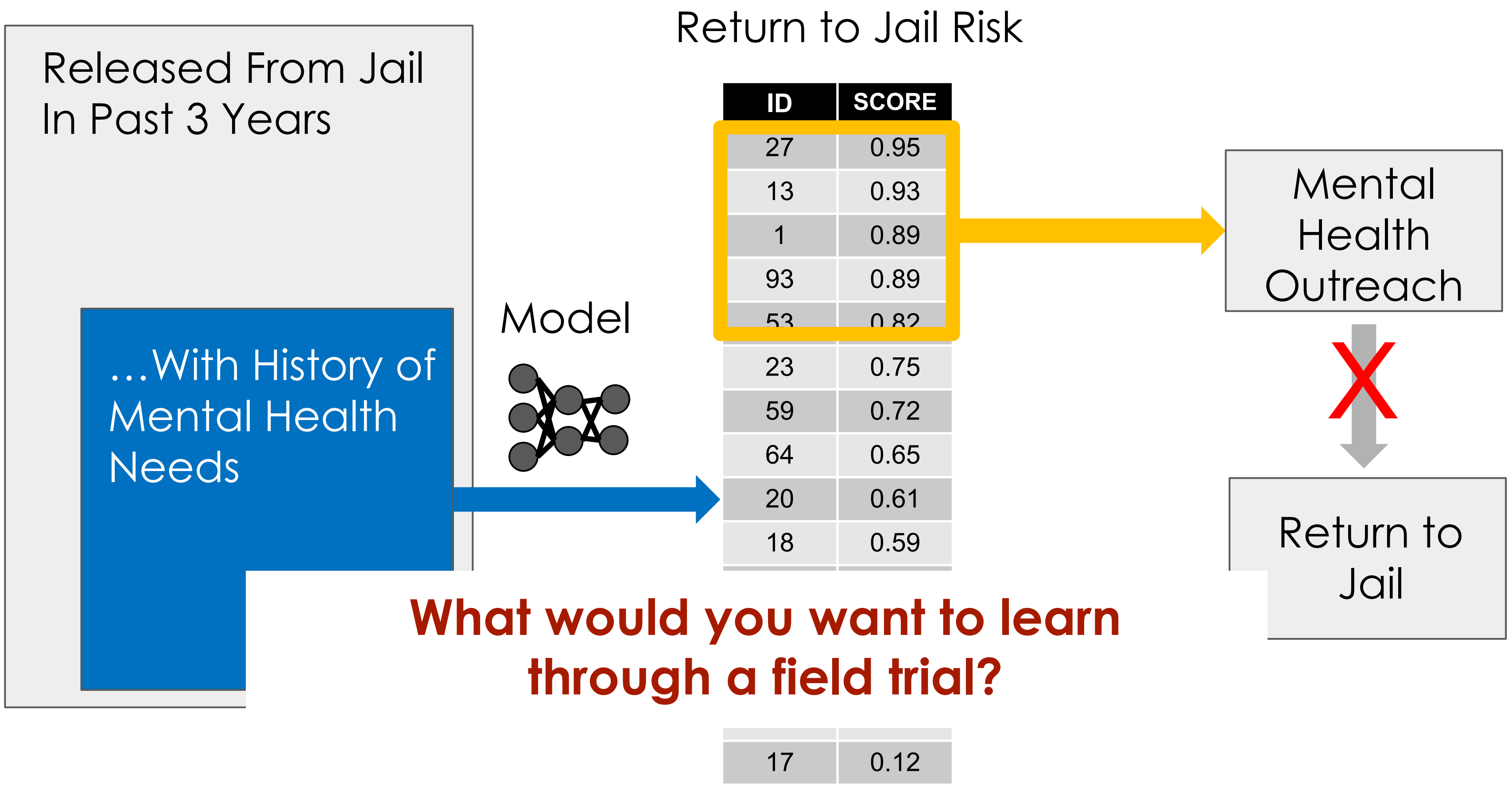


Breaking the Cycle

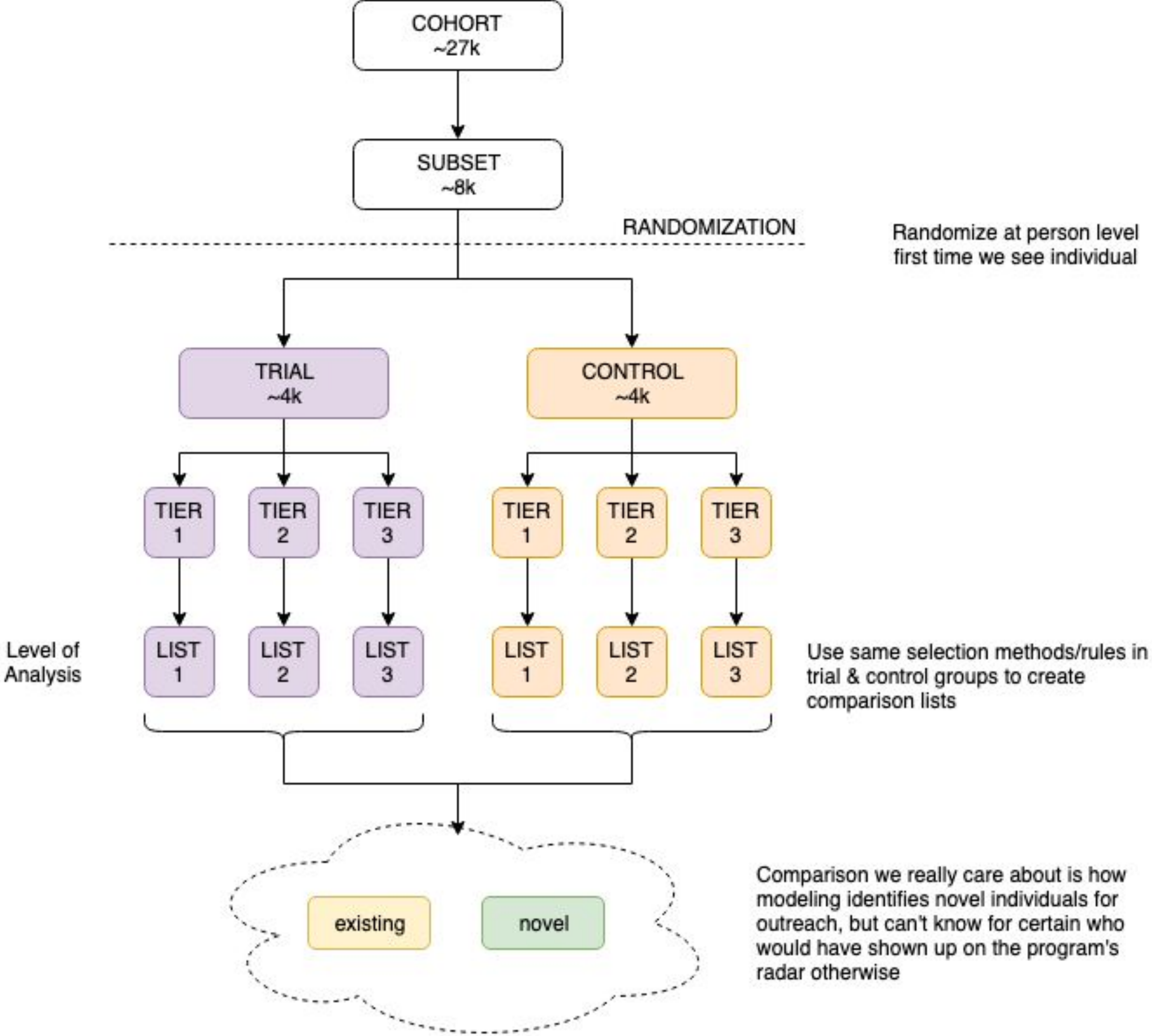




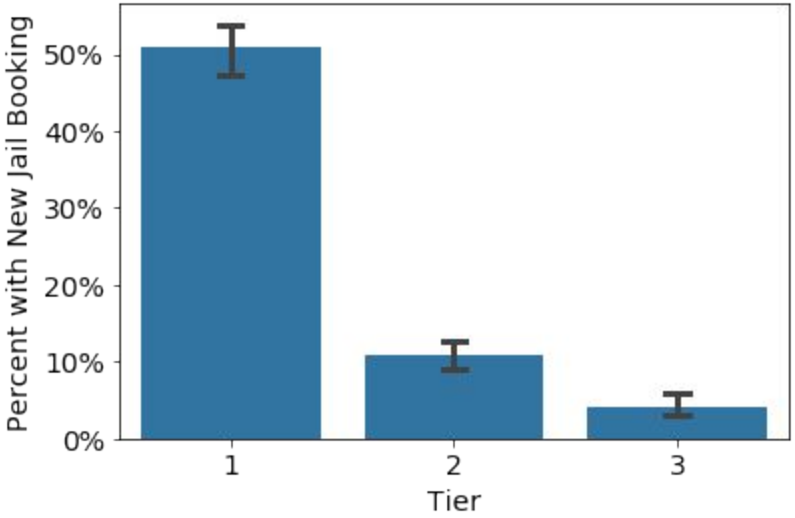
ID	SCORE
27	0.95
13	0.93
1	0.89
93	0.89
53	0.82
23	0.75
59	0.72
64	0.65
20	0.61
18	0.59
46	0.52
82	0.48
49	0.37
56	0.22
17	0.12



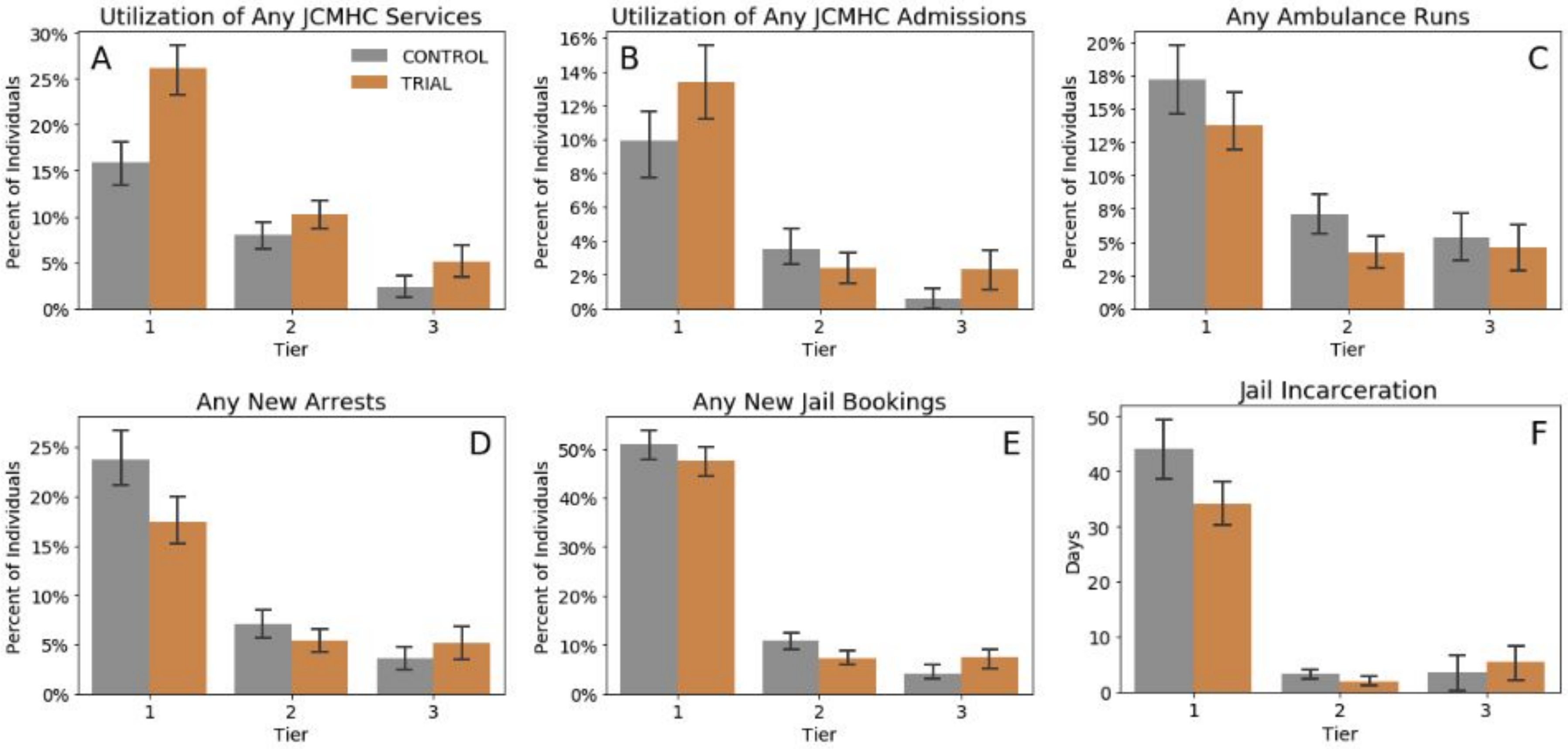
What would you want to learn through a field trial?


COHORT ~27k
SUBSET ~8k
RANDOMIZATION
Randomize at person level first time we see individual
TRIAL ~4k
CONTROL ~4k
TIER 1
TIER 2
TIER 3
TIER 1
TIER 2
TIER 3
Level of Analysis
LIST 1
LIST 2
LIST 3
LIST 1
LIST 2
LIST 3
Use same selection methods/rules in trial & control groups to create comparison lists
existing
novel
Comparison we really care about is how modeling identifies novel individuals for outreach, but can't know for certain who would have shown up on the program's radar otherwise

Q1: Is the Model Predictive?



Q2: Does Model+Intervention Affect Outcomes? For Whom?



LESSON 5

Think beyond A/B tests for field trials

Field validation is critical before deployment, but should go beyond simply asking if the model is predictive – how will it be used? what assumptions should you test?

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

Recap

- Be diligent in finding relevant data
- Models will benefit from a range of data types
- ML Fairness can be achieved (if it’s an explicit goal)...
- … but it is a property of the entire system, not just the model’s predictions
- Think beyond A/B tests for field validation

Kit Rodolfa

Carnegie Mellon University

krodolfa@cmu.edu

datasciencepublicpolicy.org

www.github.com/dssg

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

FUNDACIÓN
Carlos Slim

OFFICE OF THE CITY ATTORNEY
CITY OF LOS ANGELES
FOUNDED 1781
MIKE FEUER

Syracuse
New York

HEALTHY
CHICAGO
CHICAGO DEPARTMENT OF PUBLIC HEALTH

BALTIMORE CITY
DEPARTMENT OF HOUSING &
COMMUNITY DEVELOPMENT

JOHNSON
COUNTY
KANSAS

FUNDACIÓN
Carlos Slim