

Ethical Considerations in Predictive Analytics

Amy Hawn Nelson
AISP

Katy Collins
Allegheny County DHS

Ken Steif
Urban Spatial



AD
LOVE BOOKS
fiction

CROSS COUNTRY
ELEMENTARY
SCHOOL
MS HAWN
GRADE 2
2002 2003

Biased inputs + biased algorithms = biased outcomes

Communities of color experience staggering disparities and negative outcomes in Ramsey County, particularly in the areas of school suspensions and discipline, policing and arrests, out of home placements in criminal justice facilities, and removals through the child welfare system. **Yet the data-sharing JPA makes no commitment to reducing systems' disproportionate harm on communities of color, and instead would scrutinize individual children to reduce "delinquent behavior by youth in communities and in schools."** Using data from these institutions to predict individual children's behavior will likely serve to further magnify those disparities in scores that over-identify children of color as "risks." **Communities fear the predictions will essentially operate like racial profiling of children predicted to engage in crime.**

Source: Policy Brief, Data Sharing Joint Powers Agreement Response, INEquality/Stop the Cradle to Prison Algorithm Coalition, 2018

KEY CONCERNS

Any analysis to predict interaction with criminal justice systems will likely draw upon data that reflect bias in decision making about children.

The JPA makes no commitment to reducing systems' disproportionate harm on communities of color, and instead would scrutinize individual children to reduce "delinquent behavior by youth in communities and in schools."

Assigning risk scores to predict stigmatize human behavior is not a neutral intervention. Risk becomes interpreted as "threat" when applied to children of color.

We are all at different stages



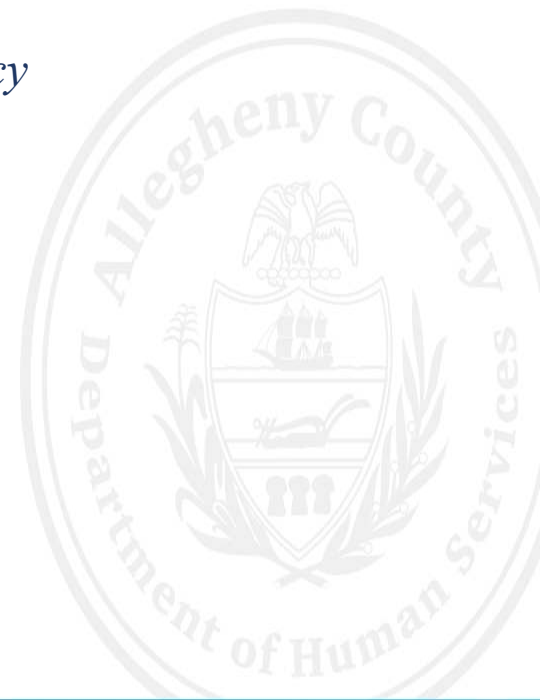


Allegheny County
Department of
Human Services

Using integrated data to support practice – Ethical considerations

Annual Meeting, Actionable Intelligence for Social Policy

June 20, 2019



Integrated Data Systems



Childhood & Education Services

Early Intervention
HeadStart
Homevisiting
Family Support Centers
Child Welfare
Family Court
Pittsburgh Public Schools + 10
additional School Districts



Basic Needs

Homeless
Housing Supports
Public Benefits
Public Housing
Employment/Unemployment
Transportation (for medically fragile)
Aging services & supports



Physical & Behavioral Health

Mental Health Services (Medicaid & Uninsured)
Substance Use Services (Medicaid & Uninsured)
Physical Health Services (Medicaid)
UPMC Health Plan (Commercial)
Intellectual Disabilities



Juvenile & Criminal Justice

Juvenile Probation
Delinquency
Pittsburgh Bureau of Police
Criminal Court
Allegheny County Jail
911 Dispatches



Vital Records

Birth Records
Autopsy Records

Improving Key Decisions with Predictive Risk Modeling



Preventing Homelessness

Improving Response to Homelessness

Improving Child Protection

Preventing Child Abuse & Neglect



Process Non-Negotiables

- Commitment to Implement
- Competitive Procurement (modeling, intervention & evaluation)
- Ethical Review (independent for most challenging approaches)
- Model Fairness & Discrimination Review
- Model validation
- Stakeholder Input
- Community Engagement
- Willingness to Modify
- Evaluation
- Commitment to Improve
- Transparency

Improving child protection

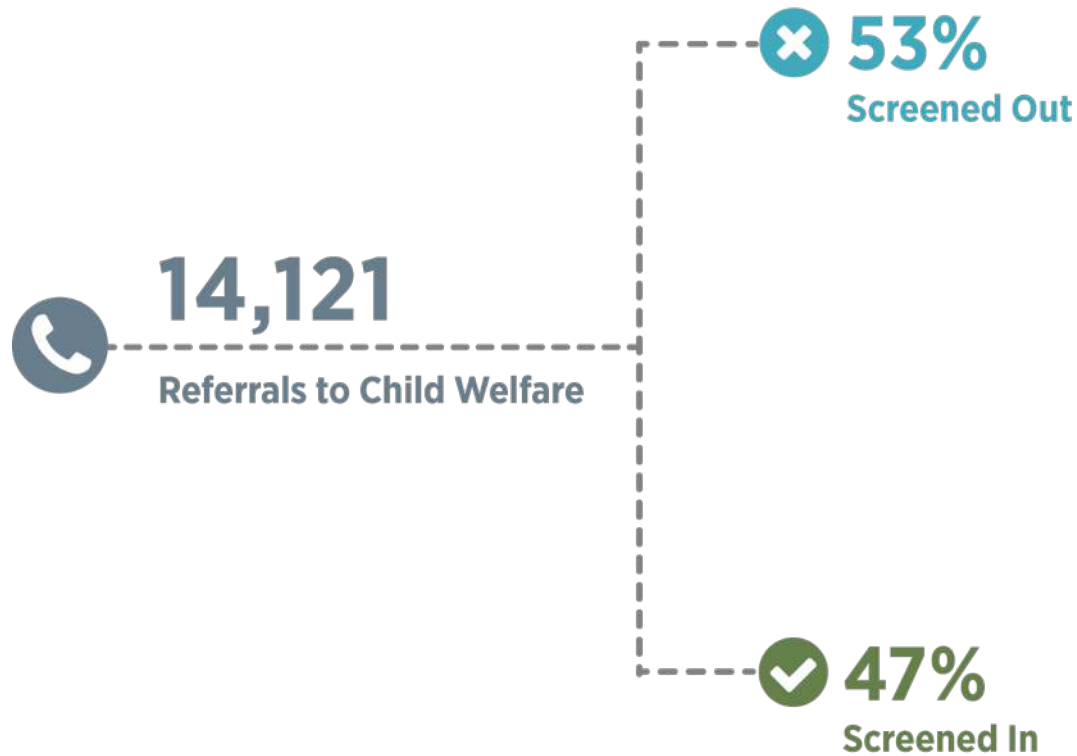
A report of child abuse is made every 10 seconds in the US, involving 6.6 million children per year

37% of children in the US will experience a child abuse investigation at some point in their childhood

We are not the police. We don't have resources to respond to every report

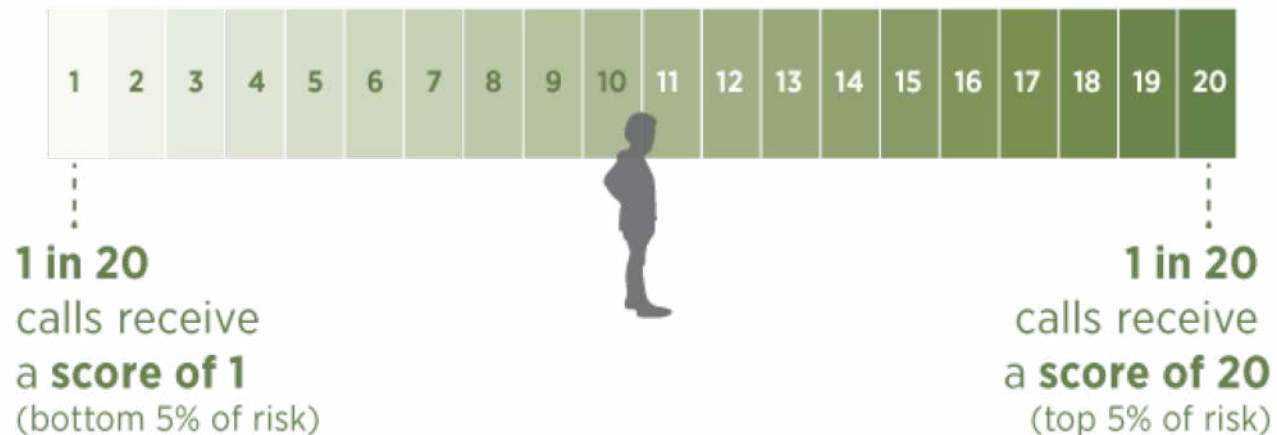
Consequences are tremendous

Improving Hotline Decision-Making



Developing a Screening Score

- The **screening score** is from 1 to 20
- The **higher the score, the higher the chance of the future event** (e.g., abuse, placement, re-referral) according to the data



Researchers built a screening model based on information that we already collect

They identified more than 100 factors that predict future referral or placement

To test if the model might improve the accuracy of screening decisions, we scored thousands of historical maltreatment calls and then followed the children in subsequent referrals to see how often the model was correct...

The Results: Out-of-Home Placements



1 in 100 children
who received a score of 1 were placed
out-of-home within 2 years of the call

The Results: Out-of-Home Placements



A group of six diverse young children are sitting on a dark-colored bench outdoors. They are all smiling and appear to be in conversation. From left to right: a boy in a striped polo shirt, a girl in a floral sleeveless top, a girl with red glasses in a pink t-shirt, a boy in a plaid shirt, a girl in a striped shirt, and a girl in a white sweater. The background shows a blurred outdoor setting with a building and trees.

Under previous practice:

27% of highest risk cases
were screened out

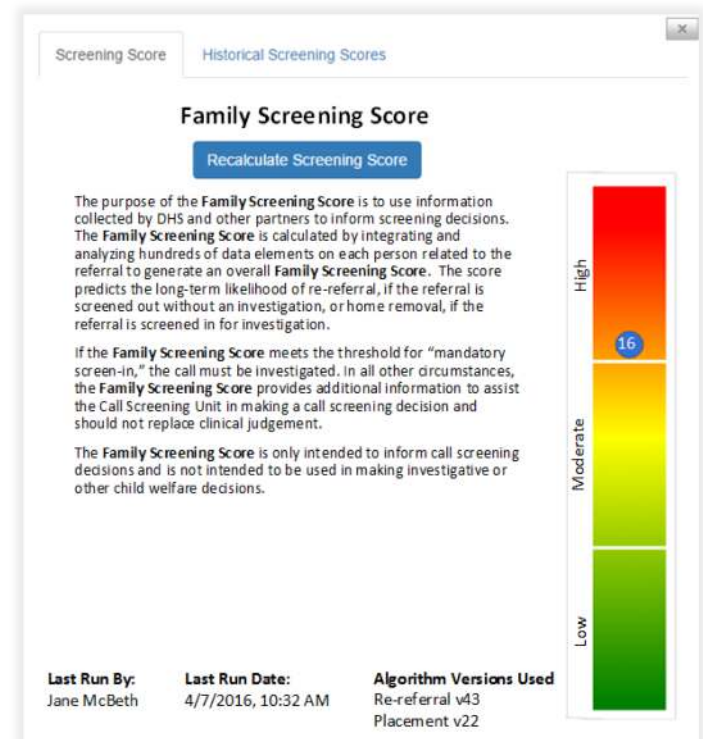
48% of lowest risk cases
were screened in

Implementation

- Live since August 2016
- Fixed bugs in November 2016
- Major changes to model, business processes & policies, November 2018

So far:

- Viewed in 100% of cases
- Caseworkers not as impressed as the New York Times
- No increase in investigations but an increase in new cases
- Not replacing clinical judgement: Concurrence with the score: ~28% of low risk cases being screened in; ~61% of high risk cases screened in



The screenshot displays a web interface for the Family Screening Score. At the top, there are two tabs: "Screening Score" (selected) and "Historical Screening Scores". Below the tabs is the title "Family Screening Score" and a blue button labeled "Recalculate Screening Score".

The main content area contains three paragraphs of text explaining the score's purpose and use. To the right of the text is a vertical color scale legend with three levels: "High" (red), "Moderate" (yellow), and "Low" (green). A blue circle with the number "16" is positioned on the "High" level of the scale.

At the bottom of the interface, there are three columns of information:

- Last Run By:** Jane McBeth
- Last Run Date:** 4/7/2016, 10:32 AM
- Algorithm Versions Used:** Re-referral v43, Placement v22

Impact Evaluation

“Implementation of the AFST saw no adverse consequences and increased the accurate identification of children who needed further intervention services, without increasing the workload on investigators.”

Impact Evaluation

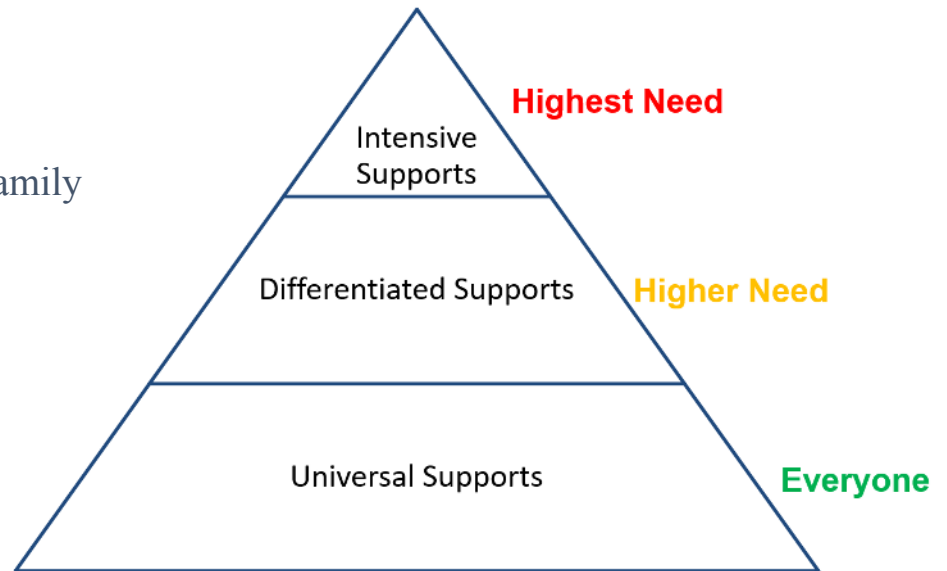
- Increased the identification of children determined to be in need of further child welfare intervention.
- Led to reductions in disparities of case opening rates between black and white children.
- Did not lead to increases in the number of children screened-in for investigation.
- No evidence that the AFST resulted in greater screening consistency.

Preventing Child Abuse and Neglect

- In over half of the cases where a child died or nearly died as a result of abuse & neglect, there had not been a child welfare referral prior to the critical incident...meaning we had no opportunity to support the family.

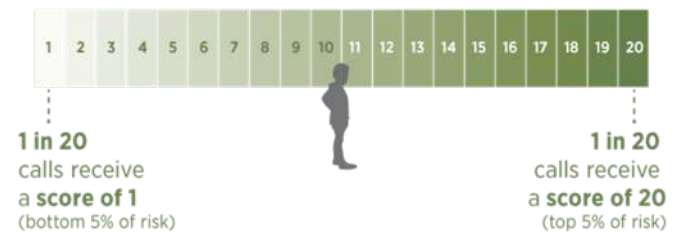
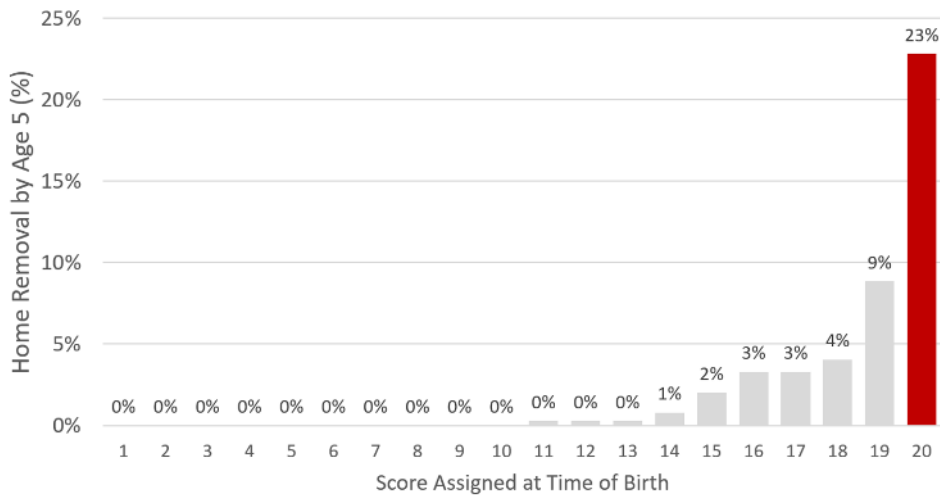
What if we could...?

- Identify families who need help earlier
- Ensure that scarce resources are getting to the families who need them most
- Offer voluntary supports that could improve family wellbeing & reduce serious abuse & neglect



Model

- Variation across the population
- 23 times the likelihood of child welfare action (home removal) by age 5
- 10 times more likely to experience infant mortality



Preparation

- Community engagement
- Independent ethical review
- Case reviews with clinicians and peer supports
- Interviews with high need families
- Responding to concerns
- Search for the best engagement approach to “bend the curve”



Concerns

- All services should be universal
- Government has weaponized data against people of color in the past
- No explicit consent
- Efficacy of the intervention
- Sounds deficit based
- Surveillance
- Stigma
- Protecting the score

Some Summary Points

- We collect data from our clients
- We can (and should) use it to improve decision making & the allocation of resources
- But we should do so with care
- The tools currently being used to do this probably aren't that great
- The process is critical
- At some point you have to implement
- Be ready for criticism (and listen to the real critiques)
- Black box tools shouldn't be employed by the public sector
- The government should enforce checks & balances on themselves
- Independent review & evaluation can be critical
- Quality assurance and maintenance is just as important as development

 alleghenycountyanalytics.us

 Kathryn.Collins@alleghenycounty.us

Geospatial risk prediction, child maltreatment & fairness

Ken Steif, Ph.D

Founder, Urban Spatial
Director, Master of Urban Spatial Analytics, Penn

<http://UrbanSpatialAnalysis.com/>



Agenda

1. People vs. place-based prediction in child maltreatment
2. The open source geospatial risk prediction framework.
 - Exploratory Analysis
 - Modeling
 - Validation/fairness metrics
3. Operationalizing fairness in ml models.

Placed-based vs. people-based prediction.

People-based approach: Gather individual & household-level Integrated Data from educational, criminal justice, health & human services and housing systems to estimate a risk score interpreted as *the probability abuse is happening here/now*. Allocate resources at the household/individual level.

Placed-based approach: Gather de-identified, geospatial event data on abuse events and other, typically open datasets describing the environmental characteristics of places to estimate a risk score interpreted as *the geospatial risk for maltreatment in this place* (~1000ft² area). Allocate resources at the community level.

Different **interventions** at different '**costs**' (financial, data privacy & data security)

Hypothesis

The geography of child maltreatment is a function of people/family's **exposure** to a series of geospatial risk & protective factors.

The open source geospatial risk prediction framework

Number of Child Protective Service events by neighborhood

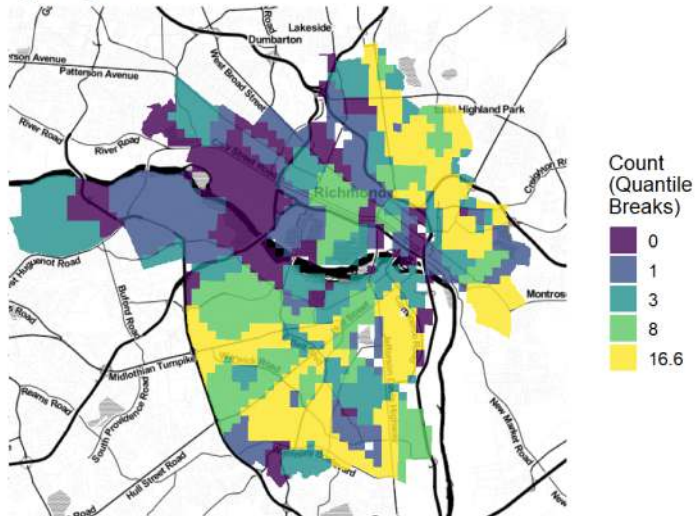
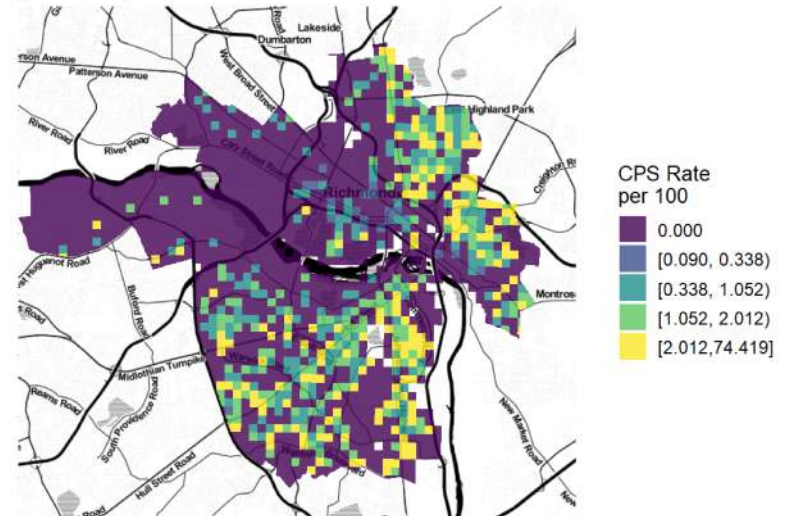


Figure 1.1

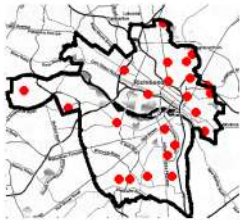
Child Protective Service rate per 100 people



That maltreatment clusters in space suggests that 'Neighborhood Effects' may play a role

Feature engineering

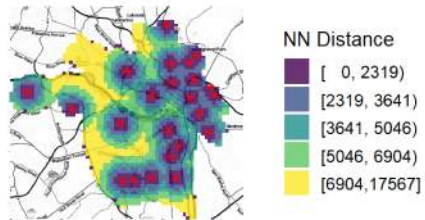
Community center locations



Community center count by fishnet



Community center euclidean distance



Community center average nearest neighbor distance

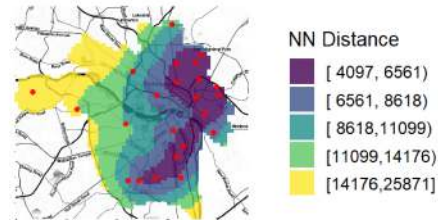
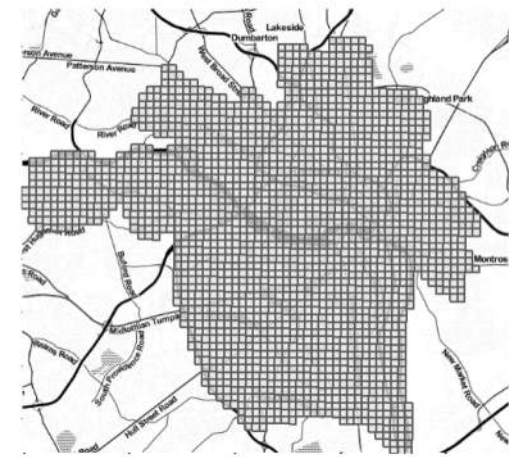


Figure 4.3

Fishnet - Richmond, VA



To quantify 'exposure', features are created by relating points in space to a lattice grid covering the City - the 'Fishnet'.

Modeling & Validation

“Borrow the observed maltreatment experience and test how **generalizable** that experience is to other places where maltreatment has yet to be reported.”

This means that a ‘good’ prediction is not just accurate (although that is part of it). A good prediction identifies places that may be at risk for maltreatment despite a lack of reporting, and does so equally well across the City. This is tested with spatial cross-validation.

LOGOCV Neighborhoods

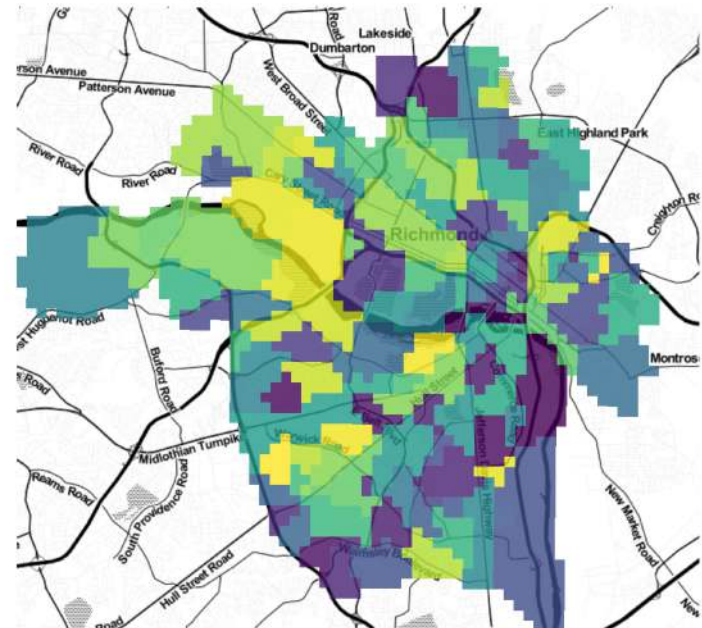


Figure 4.4

Modeling & Validation

Model Name	R2_mean	R2_sd	MAE_mean	MAE_sd	RMSE_mean	RMSE_sd	logdev_mean	logdev_sd
GLm - Poisson	0.522	0.401	0.560	0.767	0.927	1.336	0.685	0.240
Meta-Model	0.513	0.393	0.533	0.746	0.900	1.308	0.697	0.227
Random Forest	0.496	0.398	0.547	0.739	0.888	1.347	0.666	0.226
Spatial Durbin - sqrt	0.835	NaN	0.486	NaN	1.278	NaN	0.706	NaN

Ultimately, three machine learning models are estimated and combined into a fourth meta model or ensemble. We derive a host of 'goodness of fit' metrics, each calculated by way of spatial cross-validation. The model error is on average, one half of one maltreatment event.

Modeling & Validation

Ultimately, does the model help us make more useful resource allocation decisions relative to the business as usual approach? The risk prediction model is far more useful relative to the hotspot map. Here we test on hold out maltreatment events.

Risk categories from KDE

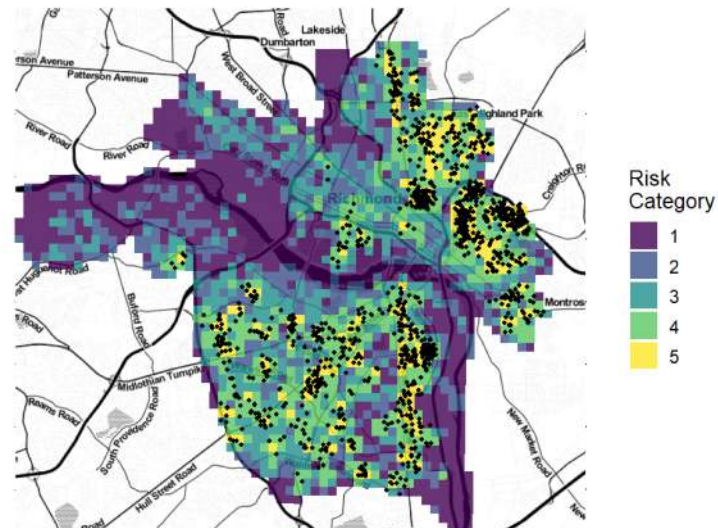
Privacy Controls: Maltreatment events in grid cells with 1 point are masked; Remaining event locations are offset at random.



Figure 6.5

Risk categories from meta-model

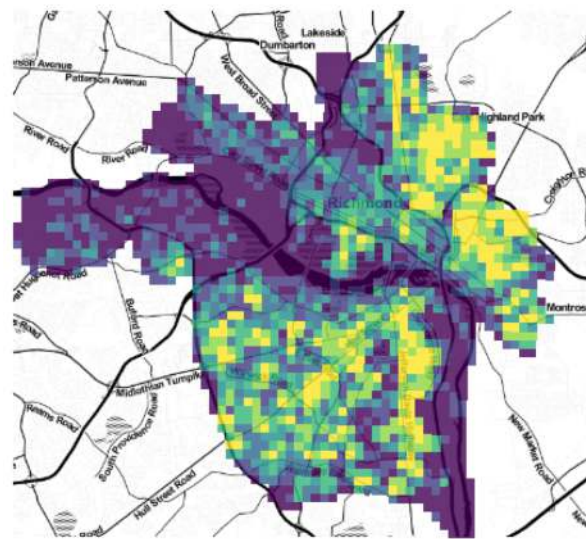
Privacy Controls: Maltreatment events in grid cells with 1 point are masked; Remaining event locations are offset at random.



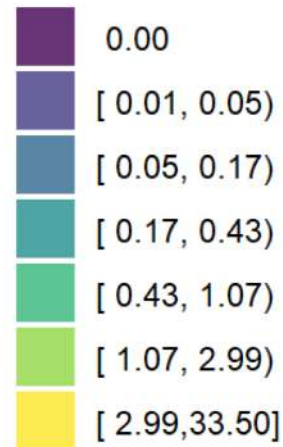
Modeling & Validation

Meta-Model

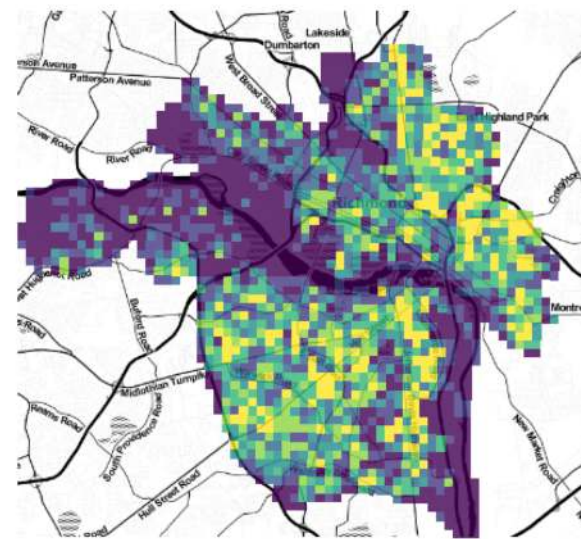
Predicted Maltreatment Count



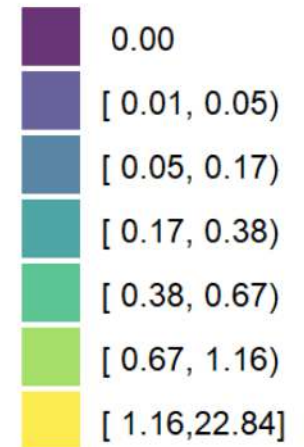
Prediction



MAE



MAE



Above are the risk predictions and errors for the ensemble. These maps are comparable. Why?

Modeling & Validation

Observed maltreatment events are on the x-axis of this plot, with predictions on the y-axis. The model fits well in general; less well in areas with very high observed maltreatment counts.

These are places with high density housing, where maltreatment clusters are recorded, sometimes, at the same address.

The scale of 'neighborhood effects' we use in our current model may not reflect these places where maltreatment feedback effects are hyperlocal.

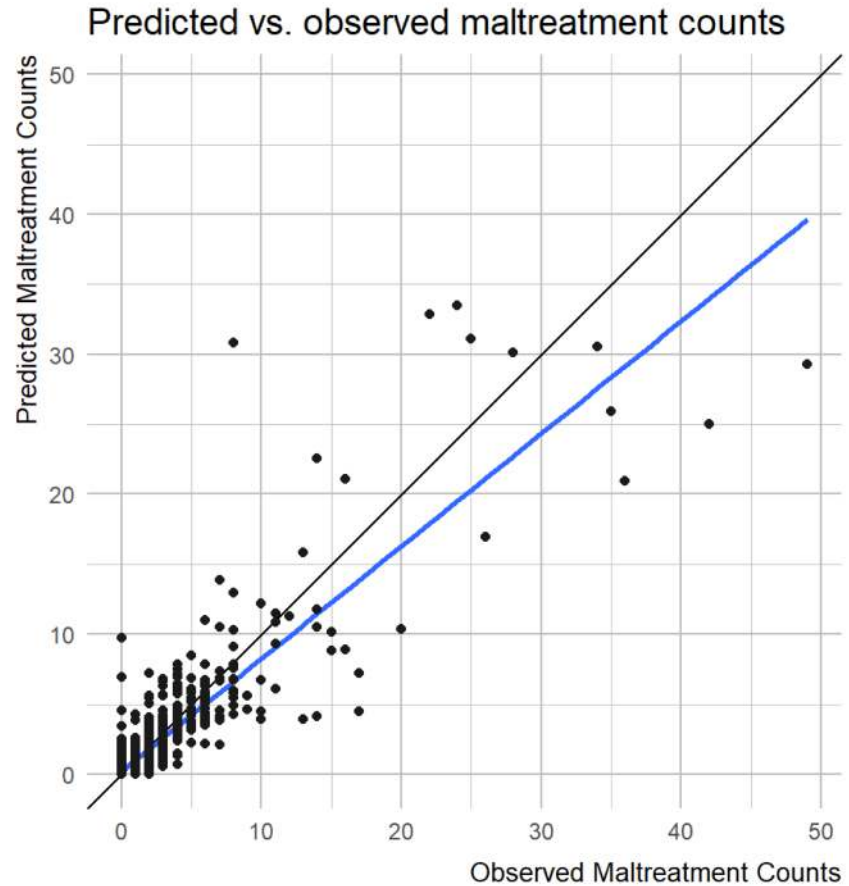


Figure 6.1

Modeling & Validation

How do we test for generalizability (aka fairness)?

There are 2 important sources of potential bias in these models:

- **Reporting bias**: Do the people who report maltreatment systematically over-police certain types of neighborhoods?

- **Selection bias**: Do places generate maltreatment behavior or do people with a propensity to commit maltreatment sort into these places?

The former is less likely in the child maltreatment use case. The latter may be more likely, but this bias pervades all research, including inferential statistics.

We developed custom bias metrics that test how well the model generalizes to different neighborhood typologies, like 'rich vs. poor' and 'minority vs. white'.

Modeling & Validation

We urge you to read Professor Tim Dare's Ethical Evaluation of the model. [Link here.](#)



New Ethics Review of Predict-Align-Prevent's Approach to Place-Based Predictive Analytics for the Prevention of Child Abuse and Neglect

We are pleased to share an ethics review of the Predict-Align-Prevent Program, conducted by Professor Tim Dare of the The University of Auckland, which identifies the ethical considerations raised by the Predict-Align-Prevent (PAP) Program, and makes recommendations to address or mitigate potential risks.

This ethics analysis was made possible by support from Casey Family Programs.



To read the review, please [click here.](#)

Algorithmic fairness: A code-based primer for public-sector data scientists



Urban Spatial Analysis - [Website / Other Work](#)

Ken Steif, Ph.D
Sydney Goldstein, M.C.P.

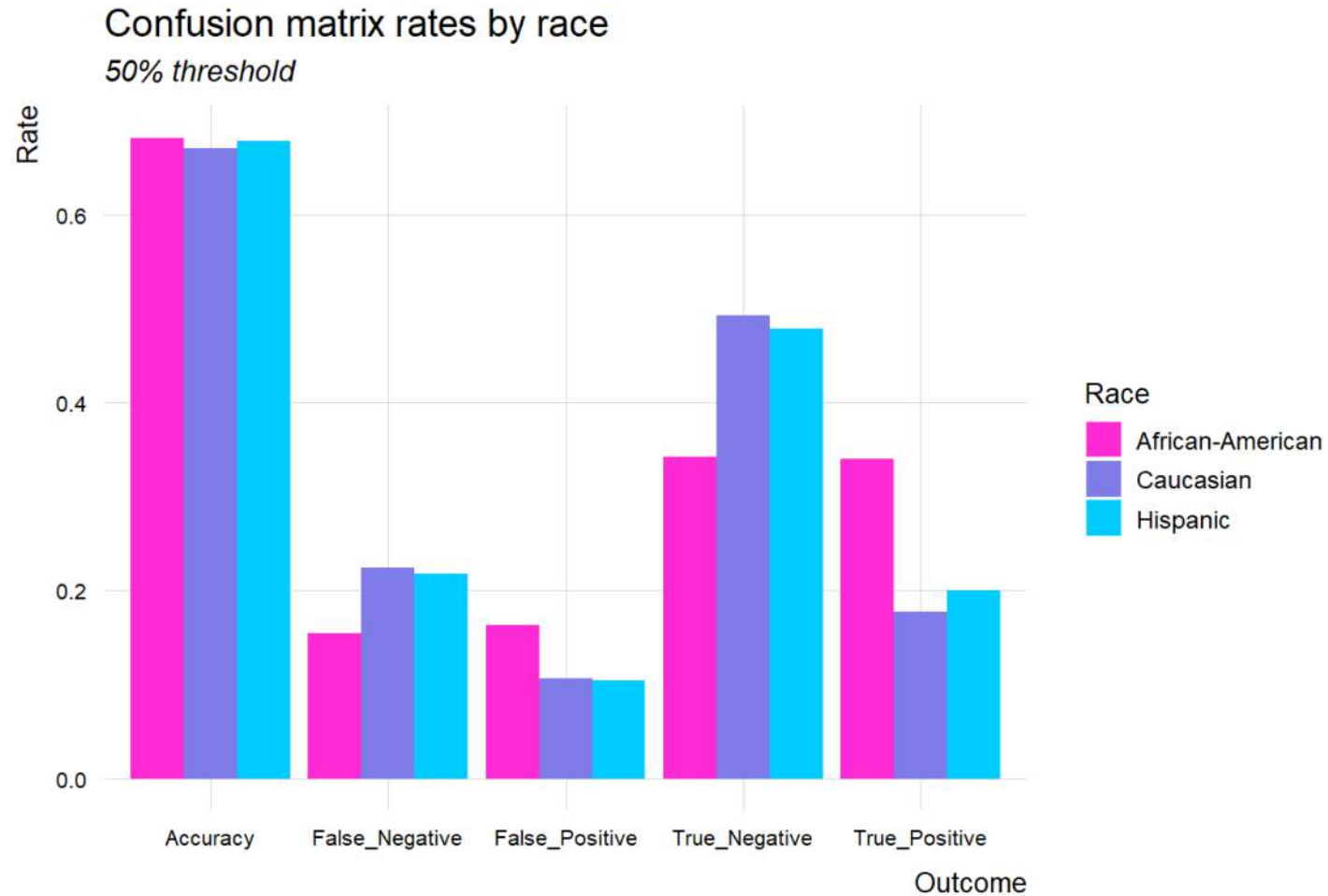
February 19, 2019

Abstract:

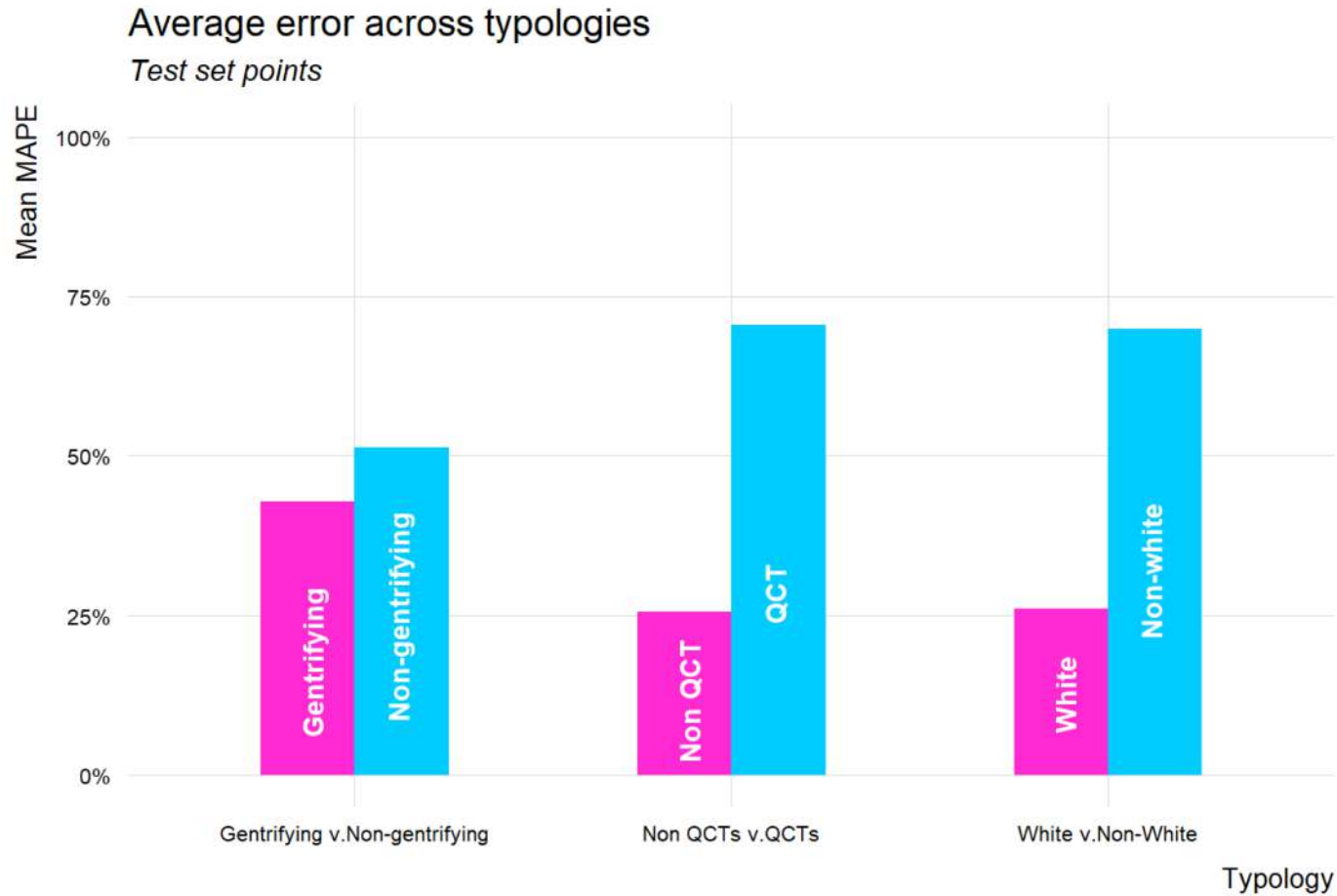
As the number of government algorithms grow, so does the need to evaluate algorithmic fairness. This paper has three goals. First, we ground the notion of algorithmic fairness in the context of disparate impact, arguing that for an algorithm to be fair, its predictions must generalize across different protected groups. Next, two algorithmic use cases are presented with code examples for how to evaluate fairness. Finally, we promote the concept of an open source repository of government algorithmic “scorecards,” allowing stakeholders to compare across algorithms and use cases.



Fairness detection - Person models



Fairness detection - Place models



Fairness correction - People-based models

Frontiers in algorithmic fairness

1. Community driven fairness metrics
2. Learning more about how bias emerges from the data creation (reporting) process.
3. Fairness correction
 - A. Remove bias before or during model estimation by learning the dynamics that make the data bias.
 - B. Remove bias after model estimation by tuning predicted probabilities to minimize across-group error disparities.

More information

This geospatial risk prediction project and all of its source code is open source and can be accessed on [GitHub](#). The full report can be accessed [here](#).

The goal is to refine the code base into an open source R package and a series of educational materials including a book and a classroom curriculum.

The fairness tutorial can be found [here](#).

Finally, there are lots of interesting public sector machine learning use cases on our [website](#). There are also a slew of models/case studies that my graduate students have built for governments around the country [here](#).

Or you can just email me at ksteif@upenn.edu and check out our [other work here](#).

This presentation can be found at:

<https://bit.ly/31Fa72t>



Geospatial risk prediction, child maltreatment & fairness

Ken Steif, Ph.D

Founder, Urban Spatial
Director, Master of Urban Spatial Analytics, Penn

<http://UrbanSpatialAnalysis.com/>



ALGORITHMIC IMPACT ASSESSMENTS:
A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY
ACCOUNTABILITY

Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker

APRIL 2018

KEY ELEMENTS OF A PUBLIC AGENCY ALGORITHMIC IMPACT ASSESSMENT

1. Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities.
2. Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time;
3. Agencies should provide notice to the public disclosing their definition of “automated decision system,” existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired;
4. Agencies should solicit public comments to clarify concerns and answer outstanding questions; and
5. Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct.

Algorithm Impact Assessments: A practical

Questions? Reactions?
