Aligning a household-level service array via geospatial and counterfactual modeling: study protocol for a jurisdiction-wide child maltreatment prevention effort

Jamaal Green

University of Pennsylvania, Philadelphia, PA, US

Counterbalance AI, Tualatin, OR, US

Brian Glass & Jordan Purdy

Counterbalance AI, Tualatin, OR, US

Dyann Daley

Predict-Align-Prevent, Grapevine, TX, US

Abstract

Background: Child maltreatment is associated with multiple negative outcomes at the individual and societal level. Children suffering from maltreatment are at greater risk of a host of negative outcomes (e.g., psychological disorders, substance use, violent delinquency, suicidality, educational outcomes).

Objective: In order to prevent and to ameliorate child maltreatment a combination of geospatial smoothing via a risk terrain modeling framework and counterfactual modeling are proffered here to identify risky areas and to determine optimal (re-)allocation of services to maximally improve maltreatment outcomes.

Methods: A three stage process is proposed which can iteratively be applied within a collaborating jurisdiction to enable responsive and sustained achievement of identified child welfare outcomes. This process makes use of two analytic approaches: geospatial smooth- ing via a risk terrain framework and counterfactual modeling. Risk terrain modeling (RTM) is a spatial analytic approach that uses spatial machine learning methods to estimate the risk of maltreatment based on prior cases of maltreatment and risk factors of the built environment. Using prior validated cases of maltreatment, violent crime data and built environment data we estimate a series of machine learning models to geospatially smooth the historically identified places at increased risk of child maltreatment. Areas identified as higher risk receive extensive services associated with preventing or limiting child maltreatment such as pre/postnatal care, subsidized daycare and parental counseling. We make use of counterfactual explanation modeling to optimally align service allocation to maximally improve maltreatment outcomes for future service allocations within a collaborating jurisdiction. This technique leverages a statistical model associating household-level information with maltreatment outcomes in order to explore combinations of services which would be predicted to achieve optimal and practical recommendations for future service allocation efforts.

Results: The household level counterfactual recommendations described in the previous section will be aggregated into actionable recommendations. The primary actionable recommendation will directly inform future iterations of Stage 1 by providing a model-informed approach for

determining eligible regions. There are several methods and strategies for aggregating the resulting household-level counterfactual service arrays. The first step is to select, for each household observation, which of the 2^n service array permutations to select. The second step is to aggregate the household-level counterfactual service arrays into spatial subregions. These subregions could be the fishnet grids specified in Stage 1, or any other chosen subregion definition such as ZIP Code, census tract, or school district. The simplest method is to, for each service type, sum the number of households in the subregion that were determined to benefit from the service. Constraints can be introduced to this logic, such as service availability and cost. Algorithmic fairness is also a potential consideration during aggregation, with possibilities for both measuring and balancing metrics such as "recourse fairness".

Conclusion: This protocol sets forth a novel approach for the allocation of supportive services for families at risk of child maltreatment through geospatial smoothing via a RTM framework and the maximization of service impact through CEM. Child maltreatment is an unfortunate, and ubiquitous, issue in the United States. This proposal builds on jurisdiction-wide public health strategies in order to allocate services in a data-informed fashion and further align future iterations of the allocation strategy using outcomes-based counterfactual modeling at the household level. The flexibility of the proposed methodology enables its application regardless of the collaborating jurisdiction’s preferences and constraints.

Introduction

In 2022, the United States Children's Bureau reported 558,899 cases of child maltreatment and 1,990 fatalities[1]. Although the rates of child sexual and physical abuse have declined in the U.S. since 1990[2], the incidence of child maltreatment fatalities has risen since 2008. Child maltreatment is underreported[3], and actual fatalities are estimated to be two to three times higher than reported due to inconsistencies in definitions and reporting standards across states[4]. Child maltreatment is linked to the development of medical illness and psychiatric disorders, as well as poor education, employment, economic, interpersonal, and community outcomes[5–10,11(p2024),12,13].

The estimated cost per victim of nonfatal child maltreatment is $830,928 (2015 USD). The average lifetime cost associated with each child maltreatment fatality is estimated at $16.6 million (2015 USD). In 2015, the economic burden of child maltreatment in the United States was estimated at $428 billion for substantiated cases, and $2 trillion for annually investigated incidents[14]. This economic burden is significant and becomes even more substantial when considered in the broader context of impact on children, family, and communities.

Given the immense personal and societal costs of child maltreatment, it is important to design programs and policies to not only alleviate but prevent abuse in the first place. Young children (ages 0-3) are particularly exposed to the risks of maltreatment[15] and many children who die due to maltreatment are not known to child protection agencies[16,17]. Given these sobering facts, identifying places and families at increased risk of maltreatment may allow authorities to proactively intervene before maltreatment may occur. Policies such as pre and post-natal counseling, subsidized childcare, and parental counseling may all have protective effects with respect to potential maltreatment if agencies can successfully identify families at risk[18].

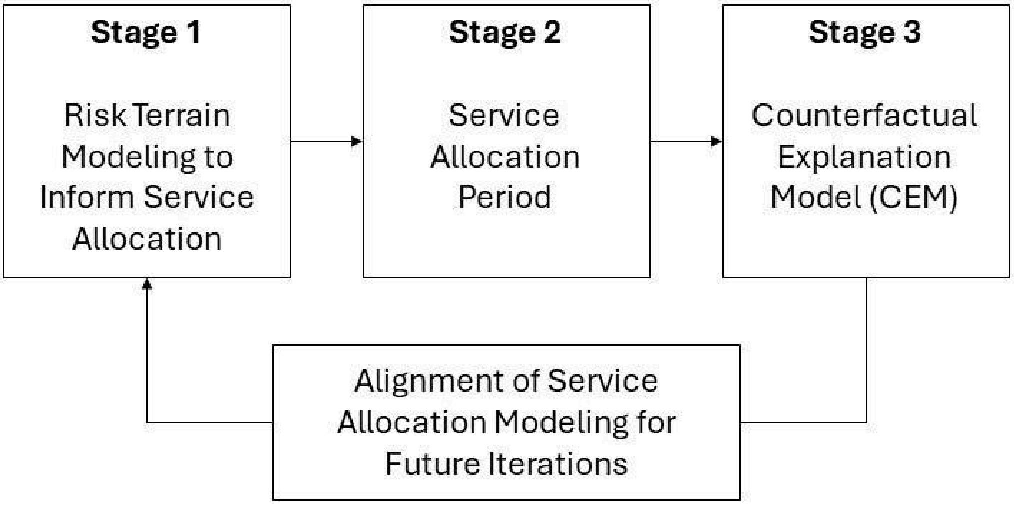
Beyond the general system-wide challenges in child welfare, maltreatment outcomes and preventive service availability demonstrate severe geographical inequities[19]. These service inequities span supportive and preventative services, including child welfare involvement itself[20]. Spatial modeling approaches can assist with identifying underserved areas along with areas with disproportionately poor outcomes[21].

Therefore, the aim of this protocol is to address three enduring issues: targeting of primary prevention resources, expansion of prevention programs to span risk factors, and identification of the bundle of programs that effectively prevents child maltreatment for a given region or household. To accomplish this, the proposed study protocol seeks to identify areas with young children at increased risk of experiencing child maltreatment, and to target these areas with extensive support programs with the goal of reducing and preventing child abuse.

The study is composed of three stages which can iteratively be applied within a collaborating jurisdiction to enable responsive and sustained achievement of the aforementioned goal. Stage 1 of the study is the development of a smoothed risk surface of child abuse and maltreatment using spatial machine learning approaches within a “risk terrain modeling” framework[22]. This stage also includes working with the collaborating jurisdiction to define child maltreatment, which can vary greatly depending on policy requirements, societal norms, and data availability[23–25]. Stage 2 of the study will be the resulting service allocation process for areas of increased estimated risk identified in Stage 1. Finally, Stage 3 uses a counterfactual explanation model to determine optimally aligned service allocation to maximally improve maltreatment outcomes for future service allocation.

We propose the use of spatial risk modeling to improve targeting of primary prevention resources, coordination of the delivery of voluntary prevention programs that span individual and contextual risks using targeted universalism, and identification of the most cost-effective bundle of primary prevention programs using counterfactual modeling. It is important to note that this work is not concerned with simply validating whether individual support programs or their various combinations are effective, in a general sense, across a collaborating jurisdiction. This work seeks to expand beyond the utility of classic program evaluation models by identifying household-specific combinations of available support programs that optimally improve household-specific outcomes.

Figure 1. The three stages of the proposed study design.


Stage 1
Risk Terrain Modeling to Inform Service Allocation
Stage 2
Service Allocation Period
Stage 3
Counterfactual Explanation Model (CEM)
Alignment of Service Allocation Modeling for Future Iterations

Background

General Study Design

The study protocol is divided into three stages: Stage 1) Geospatial service recommendations, Stage 2) Execution of service array in the collaborating jurisdiction, and Stage 3) Counterfactual modeling of service array impact (see Figure 1). The results from Stage 3 can then be used in an iterative fashion to inform the geospatial service recommendations moving forward. In Stage 1, geospatial information about historical child maltreatment events are leveraged in a risk terrain modeling framework to produce a smoothed outcome surface of child maltreatment. The top areas of risk identified by the RTM (e.g., the top quartile of "neighborhoods") will give us a recommended set of locations to receive expanded child welfare related services. In Stage 2, the services are offered and allocated within the collaborating jurisdiction guided by the geospatial service recommendation map from Stage 1. After a period of service delivery of sufficient duration to allow for a long-arc child maltreatment outcome window (e.g., 5 years total, with 3 years of service delivery and a 2 year outcome window), Stage 3 involves the development of a Counterfactual Explanation Model (CEM) to develop actionable jurisdiction-wide recommendations based on the impact of the service array on child maltreatment in the jurisdiction. Rather than provide a simplified pre-post evaluation of service effectiveness, CEM offers a tailored approach to resolve optimal service arrays at the household level, thus further aligning the geospatial service recommendations offered in Stage 1. Several details of the protocol, including the specific services offered, duration of Stage 2, and the final format of the actionable recommendations will be determined during the design scoping phase, which will involve feedback and specifications provided by the collaborating jurisdiction. The current study protocol seeks to propose a design framework robust to the various configurations and challenges that may arise during the final determination of the study details, which will be highly dependent on the governing body in the collaborating jurisdiction.

Risk Terrain Modeling: Theoretical Background and Use Cases

Risk Terrain Modeling (RTM) is a spatial analytic approach developed by Caplan et al for the purpose of estimating spatial risk with respect to environmental determinants of crime[26]. While RTM comes out of the environmental criminology literature, the basic approach is an application of existing spatial analytic approaches. The ultimate goal is the identification of risky places as opposed to individuals. The environmental criminological literature has long argued that certain environmental features can increase, or decrease, the risk of certain crimes occurring. Brantingham and Brantingham[27] theorized there are two major types of features in the built environment that influence crime: crime attractors and crime generators. Crime generators are areas that attract a large number of people and offer increased opportunities for crime. Stadium or entertainment districts are two examples of crime-generating spaces. Crime attractors, conversely, are places that offer opportunities for crime and are attractive to potential offenders because of the opportunity to engage in certain types of crime. Open-air drug markets would be an example of such a place. RTM, then, is an application of the basic logic set forth by Brantingham and Brantingham (1995) through the use of spatial predictive modeling.

The classic use cases for RTM are, of course, for estimating neighborhood risk profiles with respect to crime of varied kinds. Violent crimes, such as shootings or robberies, are often studied[28]. Recently, RTM has also been used to identify places at higher risk of child abuse and maltreatment[22,29].

For this protocol, the smoother produced through RTM is used to identify areas of heightened risk of child maltreatment to steer the initial round of service allocation decisions.

Service Allocation Period

The service allocation period will rely on the geospatial analysis from Stage 1, and involves the actual service delivery of various prevention programs to households in the jurisdiction. The jurisdiction-wide service delivery expansion effort is influenced by prior work in public health strategies for the deployment of large scale service delivery efforts. See "Stage 2 - Service Allocation" for further details.

Counterfactual Explanation Models: Theoretical Background and Use Case Examples

CEMs are statistical models of real world phenomena which can produce actionable recommendations in the interest of maximizing or minimizing the probability that some alternative outcome would have occurred. CEMs have generated considerable recent interest due to 1) their ability to resolve interpretations from "black box" models[30,31], 2) their structural alignment with both scientific and everyday reasoning about the world[32], and 3) their ability to offer deployable and testable courses of action for policy makers and practitioners. CEM methodologies can produce counterfactual courses of action for a single observational input (e.g., how a loan applicant can improve their standing, how a medical course of action can improve the survival of an ICU patient, or how a given service combination can improve the child welfare outlook of a single household).

The classic use case example for CEMs is the financial loan scenario: consider a statistical model which predicts whether a lendee will repay a loan within a given timeframe. By perturbing the input factors in the statistical model, the lending institution can discover a set of "solutions" which would be predicted to convert an unsuccessful loan applicant into a successful loan applicant, thus offering recommendations to rejected applicants. Another recent example from the field of healthcare is the proposal for CEM in an intensive care unit[33]. In the proposal, a CEM built on historical data offers treatment recommendations to maximize patient survivability. If a patient is predicted to have low chances of survival upon ICU intake, the CEM offers a treatment array sequence to maximally impact the chance of survival. Thus, a CEM leveraged a statistical model linking treatment combinations to health outcomes in order to provide actionable individual-level treatment array recommendations.

"Counterfactual" Analysis in Child Welfare

Statistical approaches to investigate the decision making in the area of child welfare often deploy methodology which uses the term "counterfactual". One such approach is in counterfactual causal modeling, which aims to determine hypothetical outcomes under different child welfare decision policies[34,35]. Another approach uses an economic formulation of "counterfactual" analysis to investigate alternative hypothetical mechanisms for assigning child welfare cases to government workers[36]. "Counterfactual" modeling has also been used for evaluating the effectiveness of a substance use intervention, by generating simulated treatment and control groups[37]. While these approaches are counterfactual in the general sense, they differ from the approach specified in the current work. Specifically, the current work seeks to model service delivery information and outcomes in order to optimize service array combination at the family level. It is possible that jurisdictions have internally planned or developed similar methodology, but the authors are unaware of any publicly acknowledged efforts to develop or deploy such a technique.

Characteristics of Counterfactual Explanation Models

A variety of theoretical work has considered the evaluation of a CEM's quality[38]. This prior work generally considers three domains for the evaluation of a given CEM: Validity, Proximity, Diversity, and Actionability[39].

Validity can refer both to the performance (e.g., AUC, accuracy, smoothness of the data manifold) of a CEM's underlying statistical model, as well as the CEM's ability to provide truly counterfactual predictions[39]. Individual predictions are truly counterfactual when the resulting outcome is in a different class than the original outcome (i.e., perturbing the features of a given input vector results in a different predicted binary outcome). Thus, statistical models whose predicted outcomes are invariant to alterations in the input vector values are not valid for use as a CEM.

Proximity refers to the minimally required alteration of an input vector CEM to generate effective counterfactuals (i.e., altered input vectors which result in a different prediction class). CEM with superior proximity generate counterfactual input vectors using minimal perturbations to the original input vector[40–42].

Diversity refers to the size or cardinality of the set of counterfactual options resolved by a CEM[42]. For example, for a given individual input vector, a CEM might recommend a large and varied set of counterfactual feature configurations which all successfully convert the prediction to the more ideal class. While not necessarily a detriment, large sets of "other possible worlds" offers a conundrum to CEM deployment, as many differing and often competing courses of action must be further compared and evaluated. This challenge, known as the Rashomon Effect[43], requires practitioners to select a single available counterfactual option. The development of a framework to compare these options is available in the section "Compute Counterfactual Service Arrays".

Actionability refers to the level of controllability of the input vector features perturbed by the CEM[44]. For example, an ICU treatment model which counterfactually recommends that a patient become taller is not actionable. This proposal seeks to maximize actionability by only perturbing input vector features which relate to the application of specific family services.

Methodology

The study consists of three stages which can iteratively be applied within a collaborating jurisdiction. Stage 1 of the study is the development of a smoothed risk surface of child abuse and maltreatment. Stage 2 of the study will be the resulting service allocation process for areas of heightened risk identified in Stage 1. Finally, Stage 3 uses a counterfactual explanation model to determine optimally aligned service allocation for future iterations of the service array.

Stage 1 - "Risk Terrain Modeling"

The objective of Stage 1 is neither to predict when or where the outcome will occur, nor to identify regions of latent risk, but rather to apply risk terrain modeling to initially identify the regions where the benefits of service allocation might reasonably be expected. To this end, we propose smoothing the historical outcome surface across the jurisdiction via development of a geospatial machine learning algorithm, where regions corresponding to higher values from the trained smoother will inform the initial service allocation encompassed by Stage 2.

Within the jurisdiction of interest, the observational units of this geospatial smoother are regular polygons the size of which will depend upon the collaborating jurisdiction. Henceforth, these observational units are referred to as "fishnet" grid cells. This "fishnet" is a regular grid of polygons (in this case squares) with an area sufficient for block to neighborhood level modeling, but is smaller than standard census geographies. The outcome being smoothed will pertain to child maltreatment, as defined by the collaborating jurisdiction, with the specific form (e.g., rate, count, etc.) likely being determined after initial exploratory data analysis. The feature set will consist of variables based on the built environment, crime, and census data both within the cell and within neighboring cells, as well as the value of the outcome in neighboring cells (i.e., auto-features), where a variety of isotropic neighborhood structures will be considered (e.g., first-order; second-order; etc.).

RTM requires appropriately robust models that describe the environment. Traditionally, this has involved the use of administrative data sources, such as US Census Data, arrest or crime data from local law enforcement, as well as "environmental" data. For this study, our primary datasets will include address level child welfare data on substantiated cases of maltreatment (neglect, physical or sexual abuse etc), local law enforcement arrest data for violent crimes (e.g. assaults, domestic disturbance calls and homicides), socioeconomic data from the US Census and built environment data from jurisdiction open data sources as well as state-level administrative data where possible.

Local socioeconomic environmental variables from the US Census will be represented by the Neighborhood Deprivation Index (NDI)[45]. The NDI is a validated and supported composite measure of neighborhood deprivation as taken from the first principal component from a set of 20 census variables. The final index makes use of 8 of those variables found in the first component: share of males in management and professional occupations, share of crowded housing, share of households in poverty, share of female-headed households with dependents, share of households on public assistance, share of households earning less than $30,000 a year, share of the population with less than a high school diploma, and the share unemployed. Built environment data will include recognized risk factors in child maltreatment including alcohol serving establishments (both bars/restaurants and liquor stores) and cannabis dispensaries (where recreational cannabis is legal), dangerous buildings from local code violations data, transit stops and supportive/protective land uses such as community centers, daycares and houses of worship.

Two approaches to creating the smoothed surface will be considered, with goodness-of-fit measures being employed to identify the better of the two. The first, kernel density estimation (KDE)[26], is expected to be inferior and will serve as the baseline approach. The second, a tree-boosted learner (e.g., XGBoost[46]), is more complicated, but is expected to be superior. For the boosted learner, k-fold cross validation - with the spatial neighborhood structure dictating the partitioning of the folds - will be used. All risk factors will be included in the final model.

Tree-based models like XGBoost are generally robust to issues of multicollinearity and other risks of high dimensional data. Hyperparameter tuning will be led by a grid search based on a range of hyperparameter values [cite of some sort?]. The combination of the grid search and the spatial based k-fold validation should allow for the estimation of a strong predictive model while guarding against overfitting and excessive residual spatial autocorrelation.

Algorithmic Fairness of Geospatial Smoother

Within the paradigm of person-level algorithms and related decision-support tools, there exists a robust literature on algorithmic fairness. Among other things, this literature posits numerous definitions of algorithmic fairness[47,48] for an identified, and potentially multidimensional, protected attribute, along with a number of mitigation procedures for "correcting" the output of such person-level algorithms with respect to the protected attribute[49–51]. Unfortunately, within the paradigm of geospatial algorithms and related decision-support tools, there is a relative dearth of literature[52] pertaining to algorithmic fairness.

In light of this reality, one approach sometimes used is an assessment of the geospatial algorithm's generalizability across coarse categorizations of poverty (high- versus low-poverty) and/or race (majority white versus majority non-white) based on census data[53]. Such assessments are akin to a type of algorithmic fairness audit, but provide no clear course of action in the face of poor generalizability.

Here we propose a relatively simple approach encompassing the algorithmic fairness of the geospatial smoother and the corresponding identification of fishnet cells for service allocation. Our goals with this procedure are to enable a simple, coarse auditing of the geospatial smoother's outputs across a census-derived protected attribute and to provide a means for the subsequent and corresponding equitable allocation of services across the levels of the protected attribute.

The approach to algorithmic fairness consists of two steps. In step 1, unsupervised machine learning (e.g., clustering) is used to identify “levels” (e.g., clusters) of a protected attribute. Such unsupervised learning would use relevant - as identified by the jurisdiction of interest - features from the census data (e.g., racial percentages; poverty percentages; etc.). If the proposed unsupervised learning “fails” (i.e., is uninformative for identifying groups/levels of a protected attribute), then we will default in step 1 to a poverty- and/or race-based protected attribute consisting of only a few levels. In step 2, the limited number of fishnet cells identified for allocation of services within Stage 2 is based on protected attribute level stratification of the geospatial smoother's output. Such identification of fishnet cells for Stage 2 is consistent with an equitable allocation of services, where the form of equity is ultimately determined by the collaborating jurisdiction. For example, a jurisdiction could elect to allocate each level of the protected attribute an equal number of fishnet cells receiving services, with the specific fishnet cells within each protected attribute level being identified according to higher values of the geospatial smoother.

While this proposed approach to algorithmic fairness related to the geospatial classifier is relatively “simple”, it is important to recognize that the norm is to ignore considerations of algorithmic fairness altogether. Furthermore, while there may understandably be concern for this geospatial smoother inadvertently perpetuating existing biases, either with or without our proposed approach to algorithmic fairness, it is important to recognize that the proposed “intervention” (i.e., Stage 2) consists of providing optional supportive services rather than imposing non-optional punitive measures. Hence, the proposed approach pushes the standard by incorporating algorithmic fairness and inherently mitigates fairness concerns by allocating optional and supportive services.

Stage 2 - "Service Allocation"

The service array provided to the households within the fishnet cells identified in Stage 1 will be finalized with input from the collaborating jurisdiction, although the following list presents a core set of service programs and/or service types, with the target population included for each item. The service array will be selected to impact the various known major risk factors for child maltreatment or adverse childhood experiences.

1. Nurse-family partnership (Pregnant people and young families)[54]
2. Cure Violence (Neighborhood)[55–57]
3. Crime Prevention Through Environmental Design (CPTED; Neighborhood)[58–60]
4. Universal basic income (Household)[61,62]
5. Early Head Start (Child and young families)[63,64]
6. Pregnancy prevention (e.g., Upstream USA; Adults and families)[65,66]
7. Mobile medical clinics (Household)[67,68]
8. Crisis intervention teams (Neighborhood and household)[69]
9. Free college tuition (Municipality / State)[70,71]
10. Stewards of Children: Darkness to Light Training (Children)[72]

Each service has been demonstrated to address one or more of these major risk factors. This comprehensive multi-service jurisdiction-wide approach uses a public health strategy[73,74] to expand preventive services to households which may not already be known to the collaborating jurisdiction's Child Welfare system. The entire set of services chosen by the collaborating jurisdiction will be available to every qualified household of each fishnet cell identified in Stage 1.

Data records will track household-level service delivery and participation, allowing for the quantification of service array information for use in Stage 3. A centralized service allocation database will be maintained by the jurisdiction. In order to properly quantify service delivery in a manner that is quantifiable for the CEM, data must include service delivery period, type and/or subtype of the service delivered by the provider. In the feature specification phase of the CEM construction, the service delivery features will be identified as boolean values. Therefore, it may be necessary to define a single service type as multiple mutually exclusive features. For example, if data investigation reveals a distribution of service delivery periods for a specific service type, further discussion with the jurisdiction and service provider may motivate the differentiation of a given service type into two or multiple features (e.g., Service Type A [One month or less], and Service Type A [More than one month]). Careful analysis of the service delivery tracking database, along with input from the collaborating jurisdiction and service providers, will be required for appropriate specification of the service type features used in the CEM.

Stage 3 - "Service Alignment Recommendation via Counterfactual Explanation Model"

Maltreatment Classifier Model Construction and Selection

The first step of Stage 3 is to develop a statistical model to associate available and quantifiable features with the selected child welfare maltreatment outcome of interest. Model specification and selection will consider the four characteristics of CEMs identified above. While data availability details will not be finalized until Stages 1 and 2 are executed in the collaborating jurisdiction, prior work suggests it is reasonable to anticipate that machine learning models can be trained to associate quantifiable household factors with child maltreatment outcomes with high performance, high algorithmic fairness, and high computational efficiency[51,75,76]. One novel contribution of this study protocol will be to extend the use of such models to inform the alignment of jurisdiction-wide service deployment strategies by maximizing the potential for positive outcomes at the household level.

Observational Unit and Data Universe

The granularity of the statistical model will be the household level. All households in the collaborating jurisdiction with at least one child member will comprise the data universe. The data set used to train and validate the statistical model will consist of the union of three mutually inclusive subsets of the data universe: Set 1) Households eligible to receive services as identified in Stage 1 of the protocol, Set 2) Households that actually received services in Stage 2 of the protocol (some of which may not have been in Set 1), and Set 3) Households found in the Statewide Automated Child Welfare Information System (SACWIS)[77]. Thus, the union of these three sets encompasses all households who either were eligible to receive services and/or are known to have the necessary SACWIS records to define an outcome. Households in the jurisdiction universe who fall out of this union set will be available as a counterfactual inference set. That is, despite not having a defined services record, their features will be available to construct an input vector for the statistical model, allowing for a predicted outcome and in turn an optimized counterfactual service array.

Sources of Model Features

A feature engineering stage will attach available quantifiable information to each household observation. These features will be comprised of information drawn from four general sources: Source 1) Household level data available from consumer and market data, Source 2) local built environment data attached to nearby households, Source 3) Governmental reporting and services data, and Source 4) the household's actual service array experience during Stage 2. During counterfactual optimization, Sources 1 and 2 remain static while Source 4 (the service array) is perturbed in order to minimize (or maximize) the predicted probability of the chosen child welfare outcome of interest. In this way, maximum actionability is preserved, since no household level features beyond the controllable service array are perturbed as part of the counterfactual modeling. Importantly, administrative data from the SACWIS system will not be included as model features. This allows for the counterfactual model to be deployable for households which have not been involved with the collaborating jurisdiction's child welfare system. Governmental reporting and services data (Source 4) can act as important predictors as well as inform a potential post-hoc outcomes analysis (see "Stage 2 - Service Allocation"). The list below presents potential data elements for Source 4.

1. Healthcare / Hospital
    a. Preterm Birth
    b. Very Low Birthweight
    c. Failure to Thrive
    d. Neonatal Abstinence Syndrome
    e. Pediatric Physical / Developmental Disabilities
    f. Teen Birth
    g. Preventable Hospital Admissions
    h. Prevention Quality Indicators (Hospital Visits)
    i. Pediatric Quality Indicators (Area Level)
    j. Psychiatric Hospitalizations

k. Injuries from Violence

l. Pediatric Lead Poisoning

m. Maternal Morbidity / Mortality

n. Substance Abuse

o. Substance Related Overdoses

p. Accidental Deaths

q. Sexual Violence / Rape

r. Prenatal Care Visits

2. Dept. of Health / Medical Examiner

   a. Infant / Child Deaths

   b. Premature Deaths

   c. Excess Mortality

   d. Marriage / Divorce

3. Crime / Emergency Services

   a. Arrests

   b. Incarcerations

   c. Domestic Disturbance / Violence

   d. Assault

   e. Gunshot / Shooting / Stabbings

   f. Intoxication

   g. Drug Abuse / Manufacturing

   h. Animal Abuse / Control

   i. Sexual Assault / Rape

    j. Arson
    k. DUI
    l. Homicide
    m. Harassment
    n. Juvenile Kidnapping
    o. Reckless Driving
    p. Suicide
    q. Welfare Checks

4. Education / Employment
    a. Graduation Rates
    b. Truancy Rates
    c. Higher Education / Trade School Matriculation
    d. Kindergarten Readiness
    e. 3rd Grade Reading Level
    f. Unemployment
    g. Poverty Indices

Model Outcome

The outcome will be defined via a joint venture with the collaborating jurisdiction's governing body and will relate to child maltreatment at the household level. Defining child maltreatment is a classic family resemblance challenge, with no universally agreed upon definition having emerged from over a century of research and governmental interest in the phenomenon[25]. Prior work has considered various maltreatment related outcomes that vary by intensity and prevalence, such as reports/referrals of child maltreatment (i.e., community members reporting alleged child abuse or neglect to a child welfare agency), whether a Child Protective Services (CPS) investigation occured for an alleged child victim in the household, whether a CPS investigation determined substantiated maltreatment in the home, and whether a child in the home was removed and placed into some form of substitute care as a ward of the state[23,24]. Outcome windows can also vary in length, although typical durations include six months, one year, and two years[51,78–80]. Outcome availability will differ between households depending on the three data universe subsets listed above, although it will be possible to both predict and define an outcome for each observation. For example, a household who was eligible for services and received services, but was not found in the SACWIS system (i.e., a member of Sets 1 and 2, but not Set 3), will be defined as "maltreatment absent", that is, there was no detected household maltreatment. However, it will still be possible to calculate a predicted probability of "maltreatment present", thus allowing for defining a counterfactual service array that could have further decreased any potential for maltreatment present in the household.

Data Governance and Ethical Concerns

The execution of this protocol will require input from the collaborating jursidiction regarding data governance and ethical considerations. The protocol should comply with the data governance requirements of the jurisdiction, including data security, data use agreements, data privacy, and the masking of personally identifiable information. A central ethical concern in the area of predictive analytics and child welfare is the stigmatization of individuals, households, and neighborhoods via automated labels such as "high risk"[81]. In order to minimize stigmatization, the use of technical and statistical terminology should be separated from the communication and design concepts that govern any public-facing or user-facing material, dashboards, or deployable implementation that results from this protocol[82,83]. For example, one practical deliverable from this protocol could be a family-specific report which can guide a human services caseworker or supervisor to additional services which might promote more positive outcomes for a family. Framed as a guide for improving the family's outcomes, the material should avoid terminology such as "high risk". In summary, the protocol must adhere to the specific data governance and ethical considerations of the collaborating jurisdiction.

Model Selection Criteria

Best practice will be followed in order to construct and select between statistical model specifications which best satisfy the requirements for an acceptable associative model for use in the CEM stage[84–86]. This best practice includes 1) exploring a range of diverse ML classifier types (e.g., XGboost, support vector machines, neural networks), 2) comparing classifier performance on a validation set (e.g., area under the curve (AUC), accuracy, specificity, sensitivity), 3) comparing the performance-fairness tradeoff for each model, with the aim of achieving large gains in algorithmic fairness at minimal cost to performance, 4) computational demands (as the CEM stage would ideally utilize an exhaustive search over thousands of perturbed input vectors for each observation), and 5) the ability of the statistical model to underlie a CEM which produces counterfactual results that are 5a) Valid, 5b) Proximal, and 5c) Diverse (see "Characteristics of Counterfactual Explanation Models" above).

Potential Algorithmic Fairness Correction

The resulting statistical model will produce predicted probabilities that will be used ordinally within an observation's counterfactual array (see below). However, it is possible that the final deployed CEM will require between-observation comparison at the aggregation stage in order to meet the use case requirements selected in collaboration with the jurisdiction. In the case that predicted probabilities will be directly compared between observations, and therefore between households at different levels of a protected attribute, algorithmic fairness corrections will be considered. The construction of such correctional procedures will follow the guidance of the well-established literature referenced in "Algorithmic Fairness of Geospatial Smoother" above[51]. In particular, the collaborating jurisdiction will determine, ideally with input from corresponding stakeholders, both the protected attribute of interest and the relevant definition of algorithmic fairness. Once these decisions are made, one or more appropriate correction procedures from the peer-reviewed literature can be identified and implemented to improve the fairness of the service-allocation recommender across the levels of the protected attribute.

Compute Counterfactual Service Arrays

The statistical model identified above will become the underlying classification mechanism of the CEM. This CEM will produce service alignment recommendations by computing counterfactual service arrays for each historical observation at the household level. This will be accomplished by optimizing the service array vector for each household observation with respect to certain criteria or constraints, such as minimizing the predicted probability of the maltreatment outcome while also minimizing the number of additional services required. The following computational methodology outlines how this is accomplished: 1) An input vector is constructed for each household observation, which includes both the household characteristic features and the binary service array features (i.e., a 0 signifies the household received Service A, while a 1 signifies the household did not receive Service A), 2) the permutations of the binary service array vector are considered to create a set of (at most) 2^n candidate input vectors (where n is the number of available services), one of which represents the actually received service array, 3) a predicted probability of the maltreatment outcome is computed for each candidate input vector, and 4) each of the 2^n candidate input vectors are quantified in terms of CEM characteristics for use in the aggregation stage. Ideally, a brute force search will be employed in this protocol to exhaustively consider all service array permutations precluding various challenges with CEM deployment associated with the possibility of missing globally optimal regions of the search space[87]. However, if computational demands outweigh the available resources, non-exhaustive optimization algorithms (e.g., mixed-integer programming[88] or genetic algorithm search[89]) will identify counterfactual service arrays which significantly lower the inferred probability of maltreatment.

In step 4 above, each candidate input vector (of the 2^n defined for each household observation) is described by certain characteristics. These characteristics are A) the change in predicted probability of the maltreatment outcome between original input vector and the candidate input vector (e.g., +4%, -10%), B) the proximity of the candidate service array vector to the original service array vector in Hamming distance (i.e., the number of services which differ between the original and the candidate counterfactual under consideration[90], and C) the proximity of the candidate input vector to a known alternative input vector from the newly predicted class, representing a proxy for generalizability confidence (by Gower's distance[91], with 1 representing equality and 0 representing maximal distance in the set). In this way, the full set of counterfactual service array candidates for a given household observation can be sorted, filtered, and evaluated to serve specific purposes in the Recommendation Aggregation stage.

For example, consider a household observation whose maltreatment outcome was positive (with retroactive predicted probability of 85%) and whose actual service array vector was {0, 0, 1, 0, 1, 0} (i.e., in Stage 2, the household received services C and E). And, for clarity, consider three members of its resulting 2^n candidate input vectors. Table 1 describes these candidate input vectors and the resulting CEM characteristics A-D for each candidate.

Identifying a single recommendation from the resulting counterfactual candidates requires developing certain standards or rules. These criteria will be developed in conjunction with the collaborating jurisdiction in order to maximally adhere to their pragmatic considerations such as resource constraints. For this example, consider how these candidates might be quantifiably compared to one another. Candidate 1 reduces the predicted probability by 5%, which is less than the other two candidates, however it requires only 1 additional service and is also very similar to a known observation in which maltreatment was in fact prevented. Candidate 2 reduces the predicted probability by 35%, requires a net change of 1 service (1 subtracted, 2 additional) and has moderate similarity to a known observation in the alternative class. Candidate 3 reduces the predicted probability by 40%, more than the other candidates, but requires 3 additional services, and is minimally similar to a known observation in the alternative class. If the determined considerations called for identifying recommendations that would require minimal additional services and maximize proximity to known observations of the maltreatment prevention class, then Candidate 1 would be selected. If the determined considerations called for identifying recommendations that would maximize the reduction of predicted maltreatment, regardless of the number of additional services and regardless of the confidence provided by proximity to known observations, then Candidate 3 would be selected. A balanced approach might result in the selection of Candidate 2. In this way, the comprehensive quantification of each candidate result works to anticipate the myriad potential weighting scenarios from which one will be chosen as the guiding principle for selecting household level counterfactual recommendations.

Table 1. Example comparison of three counterfactual candidates computed from an actual observation using the Counterfactual Explanation Model (CEM). A walkthrough of this table is available in the section "Compute Counterfactual Service Arrays"

Input Vector Description	Service Array Vector	Predicted Probability of Maltreatment	Change in Predicted Probability of Maltreatment	Change in Services Required	Proximity to Known Member of the No-Maltr. Class
Actual	{0,0,1,0,1,0}	85%	-	-	0.31
CEM Candidate 1	{0,0,1,1,1,0}	80%	-5%	Net 1 (+1)	0.99
CEM Candidate 2	{1,1,0,1,1,0}	50%	-35%	Net 1 (+1, -2)	0.67
CEM Candidate 3	{1,1,1,0,1,1}	45%	-40%	Net 3 (+3)	0.42

Aggregation of Results into Actionable Recommendation

The household level counterfactual recommendations described in the previous section will be aggregated into actionable recommendations. The primary actionable recommendation will directly inform future iterations of Stage 1 by providing a model-informed approach for determining eligible regions. Other secondary recommendations are also possible, if the final project scope calls for other ways to aggregate the results to fit various project goals. The overall conceptual method for generating these recommendations is to aggregate household-level counterfactual service array candidates into an actionable result[92].

There are several methods and strategies for aggregating the resulting household-level counterfactual service arrays. The first step is to select, for each household observation, which of the 2^n service array permutations to select. The previous section details how this selection logic can be specified in order to reflect the jurisdiction's overall project goals and constraints. The second step is to aggregate or "roll up" the household-level counterfactual service arrays into spatial subregions. These subregions could be the fishnet grids specified in Stage 1, or any other chosen subregion definition such as ZIP Code, census tract, or school district. Bespoke logic can determine the most pragmatic approach for grouping service recommendations into subregions. The simplest method is to, for each service type, sum the number of households in the subregion that were determined to benefit from the service. Constraints can be introduced to this logic, such as service availability and cost. Algorithmic fairness is also a potential consideration during aggregation, with possibilities for both measuring and balancing metrics such as "recourse fairness"[93]. Finally, these results are organized into an actionable recommendation strategy which can be provided to the jurisdiction to improve their future service delivery.

Conclusion

This protocol sets forth a novel approach for the allocation of supportive services for families at risk of child maltreatment through geospatial smoothing via a RTM framework and the maximization of service impact through CEM. Child maltreatment is an unfortunate, and ubiquitous, issue in the United States. This proposal builds on jurisdiction-wide public health strategies in order to allocate services in a data-informed fashion and further align future iterations of the allocation strategy using outcomes-based counterfactual modeling at the household level. The flexibility of the proposed methodology enables its application regardless of the collaborating jurisdiction's preferences and constraints.

Finally, it is worth noting that while the objective of the presented protocol is to reduce child maltreatment within the collaborating jurisdiction, post-hoc descriptive analyses related to changes in jurisdiction-wide outcomes known to be associated with child maltreatment (e.g., ACES; violent crime, etc.; see "Sources of Model Features" for an expanded list) are possible and may be of interest to both the collaborating jurisdiction and corresponding service providers. For example, the Cure Violence program will likely be interested in whether aggregate measures of violent crime in the collaborating jurisdiction have declined pre- versus post-implementation of a given iteration of the implemented service array. Importantly, however, whether such associated outcomes are causally improved as a result of the implemented service array neither can be nor is intended to be answered by the presented protocol. Such associated outcomes may or may not be causally related to child maltreatment, and therefore reducing child maltreatment may not result in, for example, a reduction in violent crime across the collaborating jurisdiction. Regardless, such post-hoc analyses are possible and can be informative for both the collaborating jurisdiction and corresponding service providers.

While there are challenges in setting up and implementing the proposed three-stage process, including, but not limited to, data collection, privacy, and stigmatization, the potential benefits within a collaborating jurisdiction are numerous. Achievable benefits include, among other things, reduced Child Welfare involvement and reduced ACES, more focused and efficient allocation of services and programs, as well as increased school attendance and graduation rates within the neighborhoods of the collaborating jurisdiction.

References

1. Child Maltreatment 2022. *US Dep Health Hum Serv Adm Child Fam Adm Child Youth Fam Child Bur*. Published online January 2024. https://www.acf.hhs.gov/cb/data-research/child-maltreatment

2. Lucier-Greer M, Short K, Wright EM, O'Neal CW. Trends in the Annual Incidence Rates of Child Sexual Abuse and Child Maltreatment over the Past 25 Years in the United States. *Child Abuse Rev*. 2024;33(2):e2867.

3. Viswanathan M, Rains C, Hart LC, et al. Primary Care Interventions to Prevent Child Maltreatment: An Evidence Review for the US Preventive Services Task Force [Internet]. Published online 2024.

4. Cohen M. A Jumble of Standards: How State and Federal Authorities Have Underestimated Child Maltreatment Fatalities. Published online 2024.

5. Petruccelli K, Davis J, Berman T. Adverse childhood experiences and associated health outcomes: A systematic review and meta-analysis. *Child Abuse Negl*. 2019;97:104127.

6. Tzouvara V, Kupdere P, Wilson K, Matthews L, Simpson A, Foye U. Adverse childhood experiences, mental health, and social functioning: A scoping review of the literature. *Child Abuse Negl*. 2023;139:106092.

7. Carlson P. Impact of adverse childhood experiences on academic achievement of school-aged learners. Published online 2019.

8. Peterson C, Aslam MV, Niolon PH, et al. Economic burden of health conditions associated with adverse childhood experiences among US adults. *JAMA Netw Open*. 2023;6(12):e2346323-e2346323.

9. Brown SM, Rienks S, McCrae JS, Watamura SE. The co-occurrence of adverse childhood experiences among children investigated for child maltreatment: A latent class analysis. *Child Abuse Negl*. 2019;87:18-27.

10. Connell CM, Kim HW, Shipe SL, Pittenger SL, Tebes JK. Effects of community-based wraparound services on child and caregiver outcomes following child protective service involvement. *Child Maltreat*. 2024;29(1):190-201.

11. Brown J, Cohen P, Johnson JG, Salzinger S. A longitudinal analysis of risk factors for child maltreatment: Findings of a 17-year prospective study of officially recorded and self-reported child abuse and neglect. *Child Abuse Negl*. 1998;22(11):1065-1078.

12. Maguire-Jack K, Negash T, Steinman KJ. Child maltreatment prevention strategies and needs. *J Child Fam Stud*. 2018;27(11):3572-3584.

13. Risk and protective factors. *US Cent Dis Control Prev Child Abuse Negl Prev*. Published online May 2024. https://www.cdc.gov/child-abuse-neglect/risk-factors/index.html

14. Peterson C, Florence C, Klevens J. The economic burden of child maltreatment in the United States, 2015. *Child Abuse Negl*. 2018;86:178-183.

15. Kwan R, Abraham S, Low WCJ, et al. Profile of hospitalised maltreated children aged 0 to 3 years and their families. *Singapore Med J*. Published online 2024:10-4103.

16. Garstang J, Eatwell D, Sidebotham P, Taylor J. Common factors in serious case reviews of child maltreatment where there is a medical cause of death: qualitative thematic analysis. *BMJ Open*. 2021;11(8):e048689.

17. Office GA. Child maltreatment: Strengthening national data on child fatalities could aid in prevention. *Rep GAO-11-99 Prep Chairm Comm Ways Means House Represent*. Published online 2011.

18. Jones Harden B, Simons C, Johnson-Motoyama M, Barth R. The child maltreatment prevention landscape: Where are we now, and where should we go? *Ann Am Acad Pol Soc Sci*. 2020;692(1):97-118.

19. Yi Y, Edwards F, Emanuel N, et al. State-level variation in the cumulative prevalence of child welfare system contact, 2015–2019. *Child Youth Serv Rev*. 2023;147:106832.

20. Short A. California's Child Welfare System: Addressing Disproportionalities and Disparities. Published online April 2024. https://lao.ca.gov/Publications/Report/4897

21. Caplan JM, Kennedy LW, Barnum JD, Piza EL. Risk terrain modeling for spatial risk assessment. *Cityscape*. 2015;17(1):7-16.

22. Daley D, Bachmann M, Bachmann BA, Pedigo C, Bui MT, Coffman J. Risk terrain modeling predicts child maltreatment. *Child Abuse Negl*. 2016;62:29-38.

23. Van der Put CE, Assink M, van Solinge NFB. Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse Negl*. 2017;73:71-88.

24. Day E, Tach L, Mihalec-Adkins B. State child welfare policies and the measurement of child maltreatment in the United States. *Child Maltreat*. 2022;27(3):411-422.

25. Valentine DP, Acuff DS, Freeman ML, Andreas T. Defining child maltreatment: A multidisciplinary overview. *Child Welfare*. Published online 1984:497-509.

26. Caplan JM. Mapping the spatial influence of crime correlates: A comparison of operationalization schemes and implications for crime analysis and criminal justice practice. *Cityscape*. Published online 2011:57-83.

27. Brantingham P, Brantingham P. Criminality of place: Crime generators and crime attractors. *Eur J Crim Policy Res*. 1995;3:5-26.

28. Drawve G, Thomas SA, Walker JT. Bringing the physical environment back into neighborhood research: The utility of RTM for developing an aggregate neighborhood risk of crime measure. *J Crim Justice*. 2016;44:21-29.

29. Green JW. The built environment and predicting child maltreatment: An application of random forests to risk terrain modeling. *Prof Geogr*. 2022;74(1):67-78.

30. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv JL Tech*. 2017;31:841.

31. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: A review of machine learning interpretability methods. *Entropy*. 2020;23(1):18.

32. Byrne RM. Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning. In: *IJCAI*. California, CA; 2019:6276-6282.

33. Wang Z, Samsten I, Papapetrou P. Counterfactual explanations for survival prediction of cardiovascular ICU patients. In: *Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15–18, 2021, Proceedings*. Springer; 2021:338-348.

34. Coston A, Mishler A, Kennedy EH, Chouldechova A. Counterfactual risk assessments, evaluation, and fairness. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ; 2020:582-593.

35. Foster EM, McCombs-Thornton K. Child welfare and the challenge of causal inference. *Child Youth Serv Rev*. 2013;35(7):1130-1142.

36. Baron EJ, Lombardo R, Ryan JP, Suh J, Valenzuela-Stookey Q. *Mechanism Reform: An Application to Child Welfare*. National Bureau of Economic Research; 2024.

37. Prosperi M, Guo Y, Sperrin M, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nat Mach Intell*. 2020;2(7):369-375.

38. Artelt A, Vaquet V, Velioglu R, et al. Evaluating robustness of counterfactual explanations. In: *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE; 2021:01-09.

39. Mothilal RK, Sharma A, Tan C. Explaining machine learning classifiers through diverse counterfactual explanations. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ; 2020:607-617.

40. Delaney E, Greene D, Keane MT. Uncertainty estimation and out-of-distribution detection for counterfactual explanations: Pitfalls and solutions. *ArXiv Prepr ArXiv210709734*. Published online 2021.

41. Pawelczyk M, Broelemann K, Kasneci G. On counterfactual explanations under predictive multiplicity. In: *Conference on Uncertainty in Artificial Intelligence*. PMLR; 2020:809-818.

42. Laugel T, Jeyasothy A, Lesot MJ, Marsala C, Detyniecki M. Achieving diversity in counterfactual explanations: a review and discussion. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. ; 2023:1859-1869.

43. Molnar C. *Interpretable Machine Learning*. Lulu. com; 2020.

44. Schulam P, Saria S. Reliable decision support using counterfactual models. *Adv Neural Inf Process Syst*. 2017;30.

45. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83:1041-1062.

46. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. ; 2016:785-794.

47. Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour KL. Prediction-Based Decisions and Fairness: A Catalogue of Choices, Assumptions, and Definitions. *ArXiv Prepr ArXiv181107867v3*. Published online 2020.

48. Sahil Verma, Rubin J. Fairness Definitions Explained. In: *FairWare' 18: IEEE/ACM International Workshop on Software Fairness, Gothenburg, Sweden*. ; 2018.

49. Moritz Hardt, Eric Price, Srebro N. Equality of Opportunity in Supervised Learning. *ArXiv Prepr ArXiv161002413v1*. Published online 2016.

50. Zachary C. Lipton, Alexandra Chouldechova, McAuley J. Does mitigating ML's disparate impact require disparate treatment? *ArXiv Prepr ArXiv171107076v3*. Published online 2019.

51. Purdy J, Glass B. The pursuit of algorithmic fairness: On "correcting" algorithmic unfairness in a child welfare reunification success classifier. *Child Youth Serv Rev*. 2023;145:106777.

52. Flynn C, Guha A, Majumdar S, Srivastava D, Zhou Z. Towards Algorithmic Fairness in Space-Time: Filling in Black Holes. Published online 2022.

53. Steif K, Drawve G, Thomas SA, Walker JT. *Bringing the Physical Environment Back into Neighborhood Research: The Utility of RTM for Developing an Aggregate Neighborhood Risk of Crime Measure*. Vol 44. Elsevier; 2016.

54. Olds D, Donelan-McCall N, O'Brien R, et al. Improving the nurse–family partnership in community practice. *Pediatrics*. 2013;132(Supplement_2):S110-S117.

55. Butts JA, Roman CG, Bostwick L, Porter JR. Cure violence: a public health model to reduce gun violence. *Annu Rev Public Health*. 2015;36(1):39-53.

56. Ransford CL, Volker K, Slutkin G. The Cure Violence Model for Violence Prevention. *Local Led Peacebuilding Glob Case Stud*. Published online 2019:171.

57. Delgado SA, Alsabahi L, Wolff KT, Alexander NM, Cobar PA, Butts JA. The effects of cure violence in the South Bronx and East New York, Brooklyn. Published online 2017.

58. Cozens P. Crime prevention through environmental design. In: *Environmental Criminology and Crime Analysis*. Willan; 2013:175-199.

59. Rupp LA, Zimmerman MA, Sly KW, et al. Community-engaged neighborhood revitalization and empowerment: Busy streets theory in action. *Am J Community Psychol*. 2020;65(1-2):90-106.

60. De Silva B, Dharmasiri K, MPAA B, KGNU R. Crime prevention through environmental

design (CPTED): A brief review. *Silva KBN Dharmasiri KS Buddhadasa MPAA Ranaweera KGNU 2021 Crime Prev Environ Des CPTED Brief Rev Acad Lett Artic*. 2021;2337.

61. Hasdell R. What we know about universal basic income. *Cross-Synth Standford Basic Income Lab*. Published online 2020.

62. Gennetian LA, Shafir E, Aber JL, De Hoop J. Behavioral insights into cash transfers to families with children. *Behav Sci Policy*. 2021;7(1):71-92.

63. Green BL, Ayoub C, Bartlett JD, et al. Pathways to prevention: Early Head Start outcomes in the first three years lead to long-term reductions in child maltreatment. *Child Youth Serv Rev*. 2020;118:105403.

64. Saitadze I, Dvalishvili DD. Protective role of informal social support and early childhood programs in reducing likelihood of subsequent reports of child maltreatment. *Child Abuse Negl*. 2024;149:106702.

65. Welti K, Manlove J. Unintended pregnancy in Delaware: estimating change after the first two years of an intervention to increase contraceptive access. *Bethesda MD Child Trends*. Published online 2018.

66. Health CD of P, Environment. Taking the unintended out of pregnancy: Colorado's success with long-acting reversible contraception. Published online 2017.

67. Malone NC, Williams MM, Smith Fawzi MC, et al. Mobile health clinics in the United States. *Int J Equity Health*. 2020;19:1-9.

68. Weiner D, Navalkha C, Abramsohn E, et al. Mobile resource referral technology for preventive child welfare services: Implementation and feasibility. *Child Youth Serv Rev*. 2019;107:104499.

69. Usher L, Watson AC, Bruno R, et al. Crisis intervention team (CIT) programs: A best practice guide for transforming community responses to mental health crises. *CIT Int*. Published online 2019.

70. Bolter K, McMullen I. The Kalamazoo Promise "Sweet 16," Summary Study Results: 16 Key Findings from 16 Years Studying The Kalamazoo Promise. Published online 2022.

71. Bartik TJ, Hershbein B, Lachowska M. The effects of the Kalamazoo Promise Scholarship on college enrollment and completion. *J Hum Resour*. 2021;56(1):269-310.

72. Rheingold AA, Zajac K, Chapman JE, et al. Child sexual abuse prevention training for childcare professionals: An independent multi-site randomized controlled trial of stewards of children. *Prev Sci*. 2015;16:374-385.

73. Sanders M, Higgins D, Prinz R. A population approach to the prevention of child maltreatment: Rationale and implications for research, policy and practice. *Fam Matters*. 2018;(100):62-70.

74. Higgins DJ, Lonne B, Herrenkohl TI, Klika JB, Scott D. Core components of public health approaches to preventing child abuse and neglect. In: *Handbook of Child Maltreatment*.

Springer; 2022:445-458.

75. Vaithianathan R, Benavides-Prado D, Dalton E, Chouldechova A, Putnam-Hornstein E. Using a machine learning tool to support high-stakes decisions in child protection. *AI Mag*. 2021;42(1):53-60.

76. Saxena D, Badillo-Urquiola K, Wisniewski PJ, Guha S. A human-centered review of algorithms used within the US child welfare system. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ; 2020:1-15.

77. Historical Information: SACWIS/TACWIS Information. *Child Bur*. Published online February 2024. https://www.acf.hhs.gov/cb/training-technical-assistance/state-tribal-info-systems/historical-info

78. Church CE, Fairchild AJ. In search of a silver bullet: Child welfare's embrace of predictive analytics. *Juv Fam Court J*. 2017;68(1):67-81.

79. Teixeira C, Boyas M. Predictive analytics in child welfare. Published online 2017.

80. Schwartz IM, York P, Nowakowski-Sims E, Ramos-Hernandez A. Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience. *Child Youth Serv Rev*. 2017;81:309-320. doi:https://doi.org/10.1016/j.childyouth.2017.08.020

81. Taylor J, Baldwin N, Spencer N. Predicting child abuse and neglect: ethical, theoretical and methodological challenges. *J Clin Nurs*. 2008;17(9):1193-1200.

82. Dare T, Gambrill E. Ethical analysis: Predictive risk models at call screening for Allegheny County. *Alleghany Cty Anal*. Published online 2017.

83. Richards T, Dang-Mertz T, Graham E. Language Bias in Child Welfare: Approaches to Identifying and Studying Biased Language to Advance Equitable Child Welfare Practice. Published online 2022. https://acf.gov/sites/default/files/documents/cb/language-bias.pdf

84. Zhu JJ, Yang M, Ren ZJ. Machine learning in environmental research: common pitfalls and best practices. *Environ Sci Technol*. 2023;57(46):17671-17689.

85. Grimmer J, Roberts ME, Stewart BM. Machine learning for social science: An agnostic approach. *Annu Rev Polit Sci*. 2021;24(1):395-419.

86. Hindman M. Building better models: Prediction, replication, and machine learning in the social sciences. *Ann Am Acad Pol Soc Sci*. 2015;659(1):48-62.

87. Guidotti R. Counterfactual explanations and how to find them: literature review and benchmarking. *Data Min Knowl Discov*. Published online 2022:1-55.

88. Lodi A, Ramírez-Ayerbe J. One-for-many Counterfactual Explanations by Column Generation. *ArXiv Prepr ArXiv240209473*. Published online 2024.

89. Schleich M, Geng Z, Zhang Y, Suciu D. GeCo: quality counterfactual explanations in real

time. *Proc VLDB Endow*. 2021;14(9).

90. Aguilera-Ventura C, Herzig A, Liu X, Lorini E. Counterfactual reasoning via grounded distance. In: *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol 19. ; 2023:2-11.

91. Pavoine S, Vallet J, Dufour AB, Gachet S, Daniel H. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos*. 2009;118(3):391-402.

92. Keane MT, Kenny EM, Delaney E, Smyth B. If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual xai techniques. *ArXiv Prepr ArXiv210301035*. Published online 2021.

93. Slack D, Hilgard A, Lakkaraju H, Singh S. Counterfactual explanations can be manipulated. *Adv Neural Inf Process Syst*. 2021;34:62-75.