CHILD OPPORTUNITY INDEX 3.0

TECHNICAL DOCUMENTATION

March 13, 2024

Clemens Noelke, Nancy McArdle, Brian DeVoe, Madeline Leonardos, Yang Lu, Robert W. Ressler and Dolores Acevedo-Garcia

BRANDEIS UNIVERSITY

PREFERRED CITATION

Noelke, C., McArdle, N., DeVoe, B., Leonardos, M., Lu, Y., Ressler, R.W., & Acevedo-Garcia, D. (2024). Child Opportunity Index 3.0 Technical Documentation. diversitydatakids.org, Brandeis University. Retrieved from diversitydatakids.org/research-library/coi-30-technical-documentation.

Brandeis | THE HELLER SCHOOL FOR SOCIAL POLICY AND MANAGEMENT Institute for Child, Youth and Family Policy

diversitydatakids.org
data for a diverse and equitable future

DOCUMENT VERSION

This document will be updated as needed. Versions are uniquely identified by the date on title page and in the footer. The current version for the Child Opportunity Index 3.0 is available at diversitydatakids.org/research-library/coi-30-technical-documentation. Technical documentation for COI 2.0 continues to be available at diversitydatakids.org/research-library/research-brief/how-we-built-it.

Version	Version notes
March 13, 2024 (current)	Initial publication of COI 3.0 technical documentation.

ACKNOWLEDGEMENTS

We gratefully acknowledge the generous funding provided by the W.K. Kellogg Foundation and the Robert Wood Johnson Foundation. The findings and opinions expressed in this report are our own and do not represent the views of our funders.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACS	American Community Survey
ADI	Area Deprivation Index
BMF	Business Master File
CDC	Centers for Disease Control and Prevention
CEP	Community Eligibility Provision
CCD	Common Core of Data
CRDC	Civil Rights Data Collection
COI	Child Opportunity Index
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
FRPL	Free or Reduced-Price Lunch
ICE	Index of Concentration at the Extremes
MICE	Multiple Imputation by Chained Equations
NCCS	National Center for Charitable Statistics
NCHS	National Center for Health Statistics
NCES	National Center for Education Statistics
PEP	Population Estimates Program
OLS	Ordinary Least Squares
RSEI	Risk-Screening Environmental Indicators Model
SaMD	Software as Medical Devices
SEDA	Stanford Education Data Archive
SNAP	Supplemental Nutrition Assessment Program
SVI	Social Vulnerability Index
USALEEP	U.S. Small-Area Life Expectancy Project

INTRODUCTION

WHAT IS THE CHILD OPPORTUNITY INDEX, AND WHY IS IT NEEDED?

Neighborhoods are important for families and children, shaping the economic, social and environmental contexts of their everyday lives and influencing their long-term socioeconomic and health outcomes.[1-16] Neighborhoods differ in the extent to which they provide access to important resources such as good schools, high quality jobs and safe and healthy outdoor spaces, producing a profoundly inequitable geography of neighborhood opportunity. Residential segregation racializes these inequities, so that Black, Hispanic and Indigenous children grow up in neighborhoods with much lower levels of opportunity than those of White and Asian children.[1; 2; 13; 17] Understanding how racial/ethnic segregation intersects with the geography of neighborhood opportunity is crucial for developing effective remedies that benefit all children.

The Child Opportunity Index (COI) is a composite index of neighborhood features that help children thrive, capturing variation in opportunity across U.S. neighborhoods and over time. The COI has been used to understand racial/ethnic inequities in access to neighborhood opportunity, the importance of neighborhoods in children's healthy development and to design and implement policies, programs and interventions to increase equity for families and children.[18] The COI allows users to answer questions such as: In a given metro area, state or across the country, which neighborhoods have the highest and lowest levels of child opportunity? How large is the gap between lower and higher opportunity neighborhoods within and between metro areas? Do all children enjoy access to higher opportunity neighborhoods, or are there racial/ethnic inequities? The COI has been used by over 100 peer-reviewed publications to assess the impact of children's neighborhood environments on health and other developmental outcomes.[19]

Many studies and applications use single indicators, such as the neighborhood poverty rate, to represent neighborhood resources or disadvantage. Single indicators fail to capture the multi-dimensional, intersecting and cumulative effects by which neighborhoods shape child development.[1; 2; 8; 14] Compared to a single metric of neighborhood quality, the Child Opportunity Index has greater content validity because it captures the many ways in which the neighborhood environment impacts children and families. Additionally, it recognizes the compounding benefits that multiple sources of advantage provide. For example, the COI identifies neighborhoods that have better educational opportunities as well as healthier physical environments and more economic and social resources. The COI captures a multiplicity of causal pathways at the neighborhood-level and should therefore be more predictive of children's outcomes than any single metric of neighborhood quality.

The composite nature of the Child Opportunity Index means it better reflects the legacy of structural racism on the geography of contemporary neighborhood opportunity. Structural racism has led to opportunity hoarding in affluent, predominantly White communities and the concentration of disadvantage in many Black, Hispanic and Native American communities.[13; 20-22] For example, compared to White children, Black children more often live in neighborhood environments with lower-performing schools, higher rates of pollution and fewer economic opportunities.[23; 24] The mechanisms generating these compounding inequities are multifold, including historical redlining, exclusionary zoning, racial/ethnic discrimination in the drawing of school attendance boundaries or placement of polluting industry sites.[22; 25-33] A multi-dimensional composite index such as the COI is best suited to measure how structural racism contributes to racial/ethnic inequity in access to neighborhood resources across multiple domains.

While the Child Opportunity Index is focused on neighborhood attributes important for child development, it is also an appropriate metric for assessing the quality of neighborhood contexts for adults and has shown to be a strong predictor of adult outcomes. For example, a recent study found sizeable associations between the COI and child mortality risk, but it found even stronger associations with the mortality risks for a child's adult caregivers.[8] The validation analysis performed below also suggests that the COI outperforms other composite indices constructed for the general population in predicting adult health and intergenerational socioeconomic mobility. Adult neighborhood contexts mirror the neighborhood contexts experienced in childhood and adolescence.[34-37] Residing in a low-opportunity neighborhood as a child can have cumulative effects on education, employment, health and social networks that influence the type of neighborhood where that child will eventually reside as an adult and raise their own children.[3; 4; 11; 13]

COI versions 1.0 and 2.0 have been used in over 100 peer-reviewed articles to study children's health, education and housing as well as the association between health and longevity in the general population.[8; 38] In this document, we describe the rationale and construction of the newest version of the COI—version 3.0—and present evidence on its validity relative to other widely-used metrics of neighborhood quality.

HOW DOES THE CHILD OPPORTUNITY INDEX 3.0 DIFFER FROM PREVIOUS VERSIONS?

The Child Opportunity Index (COI) 3.0 is the most recent version of the COI, succeeding COI 1.0, published in 2014, and COI 2.0, published in 2020, by diversitydatakids.org. Each successive version builds on the last, but adds more relevant indicators, improves the underlying methodology and increases the amount of available data. The following table summarizes key differences across the three versions and charts the evolution of the COI.

NUMBER OF INDICATORS AND DOMAINS. The indicators that comprise the COI are selected for their relevance to child wellbeing, their quality, reliability and availability across all census tracts. The number of indicators has grown from 19 in COI 1.0 to 29 in COI 2.0 to 44 in COI 3.0. Tables 3 through 5 list the 44 indicators included in COI 3.0. Indicators are grouped into three domains: education, health and environment and social and economic. In COI 1.0 and 2.0, the COI was calculated as an overall index and for each of the three domains. For COI 3.0, we additionally calculate and publish an index for each of 14 subdomains.

TABLE 1. CHILD OPPORTUNITY INDICES 1.0, 2.0 AND 3.0

	COI 1.0	COI 2.0	COI 3.0
Number of indicators	19	29	44
Domain and subdomain metrics	3 domains	3 domains	3 domains, 14 subdomains
Years	2012	2012, 2017	2012 to 2021
Source geography	Census tract	Census tract	Census block*

**Only census tract data is publicly available.*

COI AVAILABILITY BY YEAR. COI 1.0 was available for one year (2012), COI 2.0 for two years (2012 and 2017), and COI 3.0 is available for ten consecutive years (2012 through 2021). With COI 3.0, we have changed our convention for labeling years. COI component indicators combine data from multiple years. While the number of years used to construct to each estimate varied across indicators for COI 2.0, we harmonized all COI 3.0 indicators to cover a five-year period. For COI 2.0, we labeled five-year periods by their mid-point year, i.e., “2015” for data covering 2012-2017. For COI 3.0, we label five-year periods by their terminal year, i.e. “2017” for data covering 2012-2017.

SOURCE GEOGRAPHY. COI 1.0 and 2.0 were constructed for census tracts (2010 Decennial Census definition). COI 3.0 is constructed at the census block level, for both the 2010 and 2020 Decennial Census blocks. Constructing COI 3.0 at the block level supports flexible aggregation to multiple geographic summary levels (such as census tracts, cities, etc.). COI 3.0 is publicly available for 2010 census tracts.

KEY METHODOLOGICAL CHANGES BETWEEN COI 2.0 AND COI 3.0. Because of changes in its construction, COI 3.0 cannot be compared to COI 2.0 data. We changed data sourcing of some component indicators to improve measurement quality and facilitate more timely updates. We added new indicators and revised the methods used to process component indicators and construct the composite index. We relied on new tools, such as embeddings published by OpenAI to classify food retailers, and machine learning algorithms (Lasso Regression, Ridge Regression, and Random Forests) to improve data quality and predictive validity. We added improved measures for school quality and access to green space, new measures for social capital, and introduced new subdomains and measures for socioeconomic inequity, housing quality and wealth.

For COI 2.0, we grouped component indicators into three domains (education, health and environment and social and economic). We constructed the overall index by first averaging indicators within these domains into three domain scores, and then computed the overall index by averaging the three domain scores. For COI 3.0, we group similar indicators into 14 subdomains (e.g., elementary education, educational resources). We then compute the overall COI scores and the three domain scores from the subdomain scores. While COI 2.0 was constructed from the three domain scores, the COI 3.0 is constructed from 14 subdomain scores. By increasing the number of domains/subdomains from which the overall index is constructed from three domains (COI 2.0) to 14 subdomains (COI 3.0), we enable users to perform more fine-grained analyses with subdomain metrics (e.g., for early childhood education, elementary education, secondary and post-secondary education or educational resources) in addition to the broader domain metrics (e.g., education).

UPDATES AND FUTURE VERSIONS OF THE COI

We intend to update COI 3.0 biannually, using the same protocols for the foreseeable future. Small adjustments will likely be necessary and implemented during updates in order to improve the validity of the COI. For example, these adjustments may affect newly released data and/or previously published COI 3.0 data if some underlying data are updated, corrected or filled in retroactively. For example, certain measures of school quality in COI 3.0 are time-constant, e.g., test scores. In future COI updates, we expect to replace these time-constant with time-varying school quality indicators.

Such changes will have a very small effect on the COI because it is constructed from many different indicators, making the overall index robust to changes in a subset of its components. Nevertheless, even very small adjustments can result in changes in Child Opportunity Scores and Opportunity Levels for at least some neighborhoods. For example, neighborhoods that ranked just above or below the cutoff separating moderate- and high-opportunity neighborhoods could see their Opportunity Levels change.

We are committed to comprehensive improvement of the Child Opportunity Index. We continually review both existing and emerging data sources for potential component indicators, along with exploring innovative methods for constructing the index. We closely follow the rapidly evolving scientific research base and explore newly published databases and technologies. While we will continue to update COI 3.0, we aim to develop and launch a new version of the COI once we reach a critical mass of innovations that significantly justifies such a release.

HOW DOES THE CHILD OPPORTUNITY INDEX DIFFER FROM OTHER COMPOSITE INDICES?

Table 2 summarizes key differences between COI 3.0 and two widely used composite neighborhood indices, the Area Deprivation Index (ADI)[39] and Social Vulnerability Index (SVI)[40]. The ADI and SVI are exclusively sourced from the American Community Survey (ACS); the COI contains some of the same ACS-based measures included in the ADI and SVI but also indicators from many other sources. The COI is most comprehensive in terms of scope, measuring many aspects of neighborhoods that ADI and SVI omit, such as the school and physical environment. The COI captures more distinct content areas (domains) using more refined measurement approaches. A key limitation of the ADI relative to the other indices, discussed further below, is an error in its construction: the component indicators measured on different scales were not standardized prior to combining them into a composite index. Other differences are as follows:

UPDATE FREQUENCY. The SVI has been updated every other year. The ADI has recently been updated annually. COI 3.0 will be updated annually.

YEARS AVAILABLE. Annual COI 3.0 data is available from 2012 to 2021. ADI data is available for 2020 and 2021. The SVI is available for 2000, and for every other year from 2010 to 2020. All three indices are built from indicators that cumulate data over a five-year period.

SMALLEST GEOGRAPHY. COI 3.0 is published at the census tract-level, but constructed at the census block-level. Constructing the COI at the census block-level allows us to flexibly aggregate to other geographic summary levels, including census block groups and census tracts. The smallest geography for the ADI is census block groups, and for the SVI, census tracts.

STANDARDIZATION OF COMPONENT INDICATORS. To combine component indicators measured on different scales, such as percentages or dollars, standardization is necessary. This process ensures that the indicators are mapped onto a common scale before they are combined into an index. The ADI component indicators have not been standardized.[41-43] For the SVI, component indicators are converted into percentile ranks: tracts are sorted on the raw indicator value, grouped into 100 ordered groups of the same size and labelled from 1 to 100. For COI 3.0, component indicators are standardized using the z-score transformation.

COMPONENT INDICATOR WEIGHTS. Each COI component is weighted as a function of the correlation between the component indicator and four socioeconomic and health outcomes. Components that demonstrate a stronger association with these outcomes receive a larger weight. The SVI weighs each component equally when combining them into domain and overall scores. The ADI uses principal component loadings obtained by Singh from a county-level principal component analysis (PCA).[44; 45]

COMPARABLE OVER TIME. The ADI metrics are based on ranking census block groups for a given year, which limits their comparability over time. SVI metrics are similarly based on ranking census tracts for a given year. As a result, changes in neighborhood ranks over successive years do not capture changes in overall conditions (e.g., a recession or economic boom) that affect all block groups equally. COI component indicators are standardized for comparability over time, unlocking additional modes of analysis, such as descriptive analyses of trends in opportunity, and multivariate analyses that leverage over-time variation in census tract measures.

DOMAIN-SPECIFIC METRICS. The ADI is available only as a single metric. The SVI and COI allow for more fine-grained analyses of vulnerability/opportunity across different domains. The SVI is constructed by first grouping component indicators into four domains (themes) and then deriving domain-specific metrics, from which an overall SVI is constructed. In addition to the overall COI, we also publish three domain and 14 subdomain metrics.

MEASURES OF RACIAL/ETHNIC COMPOSITION INCLUDED. The COI and ADI do not include measures of racial/ethnic composition, such as the percentage Hispanic residents or the percentage of residents who do not speak English well. The SVI includes a measure of language proficiency and the percentage of residents who are not non-Hispanic white. The rationale for including measures of racial/ethnic composition is discussed further below.

REGIONALLY NORMED VERSIONS. U.S. states and metros differ in terms of their average opportunity. All three metrics therefore provide—in addition to a nationally-normed metric—state-normed versions for each U.S. state. The COI additionally provides metro-normed data for each of the 100 largest metro areas. Regionally-normed metrics can better reveal regional spatial inequalities than nationally-normed metrics. For example, to determine eligibility for a federal subsidy, we might rank all census tracts in the U.S. in terms of income and determine that the bottom 20% were eligible for the subsidy. However, this eligibility criterion may not work well for a state-level subsidy. Massachusetts residents, for example, have higher income overall than the U.S. as a whole, and we would find that far fewer than 20% of tracts meet the nationally-defined eligibility criterion. Massachusetts lawmakers may therefore prefer a state-normed eligibility criterion that defines the 20% of census tracts with the lowest incomes in Massachusetts as eligible.

TABLE 2. KEY DIFFERENCES BETWEEN THE CHILD OPPORTUNITY INDEX, AREA DEPRIVATION INDEX, AND SOCIAL VULNERABILITY INDEX

	Child Opportunity Index (COI) 3.0	Area Deprivation Index (ADI)	Social Vulnerability Index (SVI)
Publisher	diversitydatakids.org	Neighborhood Atlas, University of Wisconsin	Centers for Disease Control and Prevention
Update frequency	Annual	Annual	Every two years
Number of indicators	44	17	15
Content areas	Early childhood education, elementary education, secondary and post-secondary education, education resources, pollution, healthy environments, safety- and health-related resources, economic opportunities, economic resources, concentrated inequity, housing resources, social resources, wealth	Education, occupation, economic opportunities, economic resources, wealth, social resources, housing resources, transportation	Education, economic opportunities, economic resources, household composition, disability, racial/ethnic composition, housing resources, transportation
Smallest geography	Census tract	Census block group	Census tract
Years	2012-2021, annual data	2020, 2021	2000, 2010, 2014, 2016, 2018, 2020
Standardization of component indicator	z-score transformation	No	Percentile ranking
Track change over time	Yes	No	No
Component indicator weights	Based on correlation between component indicator and outcomes	Based on principal component analysis by Singh	None
Regionally normed versions?	National, state, metro areas	National and state	National and state
Domain-specific metrics?	Metrics for three domains and fourteen subdomains	No	Four domains
Measures of racial/ethnic composition included	No	No	Yes

COMPONENT INDICATORS

Neighborhood factors shape children's access to resources and experiences that promote healthy development. Neighborhoods are multi-dimensional, influencing child development through numerous features and causal pathways. We group neighborhood features into three domains, through which neighborhood contexts influence child development: education, health and environment and social and economic opportunity. Each domain in turn includes subdomains that capture distinct features, e.g., secondary education or exposure to environmental pollution.

EDUCATION DOMAIN

EARLY CHILDHOOD EDUCATION. There is strong evidence of sizable long-term effects of some high-quality preschool education programs targeted at children from disadvantaged backgrounds.[46-48] Similarly, evaluations of universal preschool programs at the city- or state-level find beneficial long-term effects, particularly for children from disadvantaged backgrounds.[46; 49-53] Unlike in K-12 education, private preschool providers account for a much larger share of enrollment and also utilize funding from both public and private sources.[54] We therefore included two measures of preschool enrollment, enrollment in public preschool and enrollment in private preschool.

ELEMENTARY EDUCATION. Standardized tests measure student proficiency in reading and math. Variation in test scores reflect variation in students' cognitive ability and learning-related socio-emotional skills[55] as well as variation in educational opportunities provided by families, schools and neighborhoods.[56] Schools are central to student learning, and a growing number of studies show that malleable school features impact student learning and long-term socioeconomic outcomes.[57-70] Therefore, variation in school quality likely contributes to the variation in student proficiency observed between schools and across school districts.[56]

School-level measures derived from standardized test scores are one metric to capture school quality.[56; 64; 66] For example, studies have found that children residing in neighborhoods that report higher test scores will tend to have higher income as adults,[37] and that attending schools with higher average test scores will boost student test scores as well.[64] However, it is difficult to measure the independent effect of school quality using test scores because student family background is strongly correlated with both school quality and achievement on test scores.[56; 71]

We measure the quality of elementary schools using metrics that are based on standardized reading/language arts and math tests administered in public schools from grades three to eight. In addition to average standardized test scores in math and reading, we also incorporate two metrics that are either weakly correlated or uncorrelated with student socioeconomic composition: growth in standardized test scores and poverty-adjusted standardized test scores. Reardon argues that growth in standardized test-scores is both a better measure of school quality than average school test scores and shows that is only very weakly with average school test scores.[56] Based on the work of Angrist et al., we compute a poverty-adjusted test score metric that is uncorrelated with the socioeconomic composition of the student body by construction.[71]

TABLE 3. EDUCATION DOMAIN INDICATORS

Indicator	Definition
Early childhood education subdomain	
Private pre-K enrollment	Percentage of 3- and 4-year-olds enrolled in private nursery school, preschool or kindergarten
Public pre-K enrollment	Percentage of 3- and 4-year-olds enrolled in public nursery school, preschool or kindergarten
Elementary education subdomain	
Reading and math test scores	Standardized test scores in math and reading/language arts
Reading and math test score growth	Growth in standardized test scores in math and reading/language arts
Poverty-adjusted reading and math test scores	Poverty-adjusted standardized test scores in math and reading/language arts
Secondary and post-secondary education subdomain	
Advanced Placement course enrollment	Percentage of 9th-12th graders enrolled in at least one AP course
College enrollment in nearby institutions	Percentage of 18-24-year-olds enrolled in college within a 20-mile radius
High school graduation rate	Percentage of ninth graders graduating from high school on time
Educational resources subdomain	
Adult educational attainment	Percentage of adults aged 25 and over with a Bachelor's degree or higher
Child enrichment-related non-profits	Density of non-profit organizations providing enrichment opportunities for children, such as after-school programs, recreational sports leagues and mentoring programs
Teacher experience	Percentage of teachers in their first and second year, reversed
School poverty	Percentage of students in elementary schools eligible for free or reduced-price lunches, reversed

SECONDARY AND POST-SECONDARY EDUCATION. Post-secondary education, particularly four-year college degrees, have a large impact on labor markets, socioeconomic status and health outcomes.[72; 73] However, there is considerable variation among neighborhoods in the educational opportunities they provide, especially regarding norms, expectations and access to post-secondary education.[13; 15; 74-77] We measure inequities in neighborhood-level post-secondary opportunities using three indicators: high school graduation rates, Advanced Placement (AP) course enrollment and college enrollment in nearby institutions. High school graduation rates and AP course enrollment reflect both neighborhood-level norms and expectations, but also the availability of educational resources within schools and communities that facilitate educational achievement in K-12 and beyond. High rates of college enrollment in nearby institutions reflects local access to post-secondary education, but can also impact student aspirations. AP course enrollment has been linked to increased college attendance and completion.[78-80] Studies find that incentives to complete high school and obtain post-secondary education, such as compulsory schooling laws and college proximity, not only increase educational attainment, but also have long-term effects on labor market outcomes, health and life expectancy.[81-88]

EDUCATIONAL RESOURCES. The educational resources subdomain includes indicators measuring both tangible and intangible neighborhood resources that benefit children's educational attainment. First, we have included measures of adult college attainment and school poverty, which capture variation in the socioeconomic composition of peers in schools and adults in the neighborhood. Both are associated with educational achievement.[23; 89-92] They capture aspects of the neighborhood social environment, including role models, norms, preferences and aspirations related to educational attainment. These factors, in turn, influence educational outcomes like high school graduation or college completion.[13; 15; 74; 76; 92] Both may also directly or indirectly impact learning-related resources in schools, including teacher quality and the quality of facilities, as well as peer effects such as educational aspirations and classroom disruptions. Teacher experience and child-enrichment-related non-profits capture more tangible resources that advance child development both inside and outside of schools. Teacher experience is a robust predictor of student learning and long-term outcomes.[65; 66; 93-97] Child enrichment non-profits (e.g., libraries, museums, recreation clubs, youth centers, after-school programs, youth sports leagues and Big Brothers & Big Sisters programs) provide important community resources outside school hours that support families and provide children with opportunities for physical, social and emotional development.[98-100]

HEALTH AND ENVIRONMENT DOMAIN

There are large spatial and racial/ethnic inequities in access to healthy neighborhood environments.[24; 101-109] Black, Hispanic and Indigenous children are more like to grow up in neighborhoods with higher rates of pollution and less access to green space. These neighborhood features impact adults and children alike, starting in utero, and exert long-term effects on health and other developmental outcomes. For COI 3.0, we grouped neighborhood health and environmental features into four subdomains: pollution, healthy environments, safety-related resources and health resources.

POLLUTION. We included two indicators of neighborhood air pollution: airborne microparticles (PM2.5) and ozone concentration. These indicators have been linked to adverse neurodevelopmental and birth outcomes, chronic illnesses and long-term adverse health and education outcomes.[110-123] We also included an index measuring exposure to over 600 toxic chemicals emitted into the air or water by US facilities, which require reporting to the EPAs Toxics Release Inventory. The index as well as the toxic chemicals it includes have been linked to multiple developmental and health issues in children.[124-129] The pollution subdomain also incorporates a measure of proximity to hazardous waste dump sites (uncleaned Superfund sites) that harm nearby residents through both air and water contamination, and increase the risk of adverse birth outcomes, reduce educational attainment and adversely affect health and life expectancy.[130-135]

TABLE 4. HEALTH AND ENVIRONMENT DOMAIN INDICATORS

Indicator	Definition
Pollution subdomain	
Airborne microparticles	Mean estimated microparticle concentration (PM2.5; micrograms per cubic meter), reversed
Ozone concentration	Mean estimated 8-hour average ozone concentration (parts per billion), reversed
Industrial pollutants in air, water or soil	Index of toxic chemicals released by industrial facilities, reversed
Hazardous waste dump sites	Average number of Superfund sites within a 2-mile radius, reversed
Healthy environments subdomain	
Fast food restaurant density*	Percentage of restaurants that serve fast food, reversed
Healthy food retailer density*	Percentage of food retailers selling healthy food
Extreme heat exposure	Number of summer days with maximum temperatures above 90F, reversed
NatureScore	NatureScore measures exposure to healthy natural environments using data on green space, tree canopies, parks, and air, noise and light pollution
Walkability	EPA Walkability Index
Safety-related resources	
Community safety-related non-profits	Density of non-profit organizations focused on increasing community safety
Vacant housing	Percentage of housing units that are vacant, reversed
Health resources	
Health-related non-profits	Density of non-profit organizations providing health-related services
Health insurance coverage	Percentage of individuals aged 0-64 with health insurance coverage

*Note: *The fast food and healthy food retailer density metrics are combined into the access to healthy food index. Only the latter is included in the calculation of the healthy environments subdomain score.*

HEALTHY ENVIRONMENTS. Due to global climate change, extreme heat exposure is worsening in the US, with low income and non-White communities at increased risk.[106; 136; 137] There is strong evidence that heat exposure is associated with higher risk of adverse birth outcomes, heat stress, heat-related illness, death in children and reduced academic achievement.[106; 138-145] A rapidly growing literature has examined the health benefits of exposure to natural environments, which has been linked to increased physical activity, reduced stress and improved mental wellbeing.[146-149] We measure exposure to natural environments using NatureScore. NatureScore improves upon single-indicator measures of green space to consider a range of features from numerous sources, including measures of green vegetation, bodies of water, and human-made structures, such buildings and roads.[150-152]

Neighborhood walkability has been linked to increased physical activity and reduced cardiovascular disease risk.[153-156] Walkable communities may also encourage social interaction.[157] Similarly, the neighborhood food environment is linked to improved nutritional quality, decreased diabetes risk and increased food security. We measure healthy food access using two indicators: access to healthy food retailers[158-163] and exposure to fast food restaurants.[160-162; 164; 165] While the benefits of healthy food retailers (e.g., supermarkets) is not definitive,[166; 167] we include it because proximity to healthy food stores is a prerequisite for a healthy diet.

SAFETY-RELATED RESOURCES. Children's exposure to violent crime has been linked to a range of adverse developmental outcomes, including reduced cognitive and socio-emotional skills as well as reduced intergenerational socioeconomic mobility.[3; 168-174] Due to the lack of a nationally comparable dataset on neighborhood exposure to violent crime, we use two neighborhood measures that are predictive of violent crime: the density of local non-profits focused on increasing community safety and housing vacancy rates. Sharkey et al. show that the density of non-profits dedicated to increasing community safety is associated with lower violent crime rates.[175] Several observational and experimental studies have linked the presence of vacant or abandoned housing with diminished perceptions of safety, reduced mental health and increased violent crime.[176-181]

HEALTH RESOURCES. The health resource subdomain includes both neighborhood health insurance coverage rates for individuals under age 65 and the density of health-related nonprofit organizations. Health insurance coverage rates are a marker of health care access, as coverage reduces costs and increases the demand for health care in a given area. However, expansions in health insurance coverage also impacts providers, who may increase the provision and quality of services to meet the increased demand.[182-186] Empirical research on the impact of health-related non-profits is limited, but there is suggestive evidence that their activities may improve community health.[99; 187]

SOCIAL AND ECONOMIC DOMAIN

The social and economic domain includes neighborhood indicators that measure different forms of economic and social capital, such as household income and friendship networks. These resources are unequally distributed across neighborhoods and highly predictive of children's short and long-term developmental outcomes. These indicators have also been interpreted as indicative of intangible features of the neighborhood social environment, such as norms and aspirations related to educational attainment, labor market participation and family formation. For example, neighborhood employment rates can influence the future employment status of children by connecting youth to employment opportunities through social networks and by shaping work-related norms and aspirations.

EMPLOYMENT AND ECONOMIC RESOURCES. Access to and the quality of employment is central to family economic wellbeing. In the employment subdomain, we include three measures capturing distinct facets of local economic opportunities: (1) the employment rate, which captures variation in access to jobs but might also measure less concrete resources, such as labor market networks that facilitate job finding and job mobility, (2) high-skill employment, which captures access to high-income jobs that also provide other amenities such as more stable employment and access to flexible work schedules, and (3) labor market earnings. In the economic resources subdomain, we include three measures capturing different facets of household income: (1) the poverty rate, (2) the public assistance rate, and (3) median household income.

TABLE 5. SOCIAL AND ECONOMIC DOMAIN INDICATORS

Indicator	Definition
Employment subdomain	
Employment rate	Percentage of adults aged 25-54 who are employed
High-skill employment rate	Percentage of individuals aged 16 and over employed in management, business, financial, computer, engineering, science, education, legal, community service, health, arts and media occupations
Full-time year-round earnings	Median earnings in the past 12 months for civilian employees working full-time, year-round
Economic resources subdomain	
Median household income	Median household income of all households
Poverty rate	Percentage of individuals living in households with incomes below 100% of the federal poverty limit, reversed
Public assistance rate	Percentage of households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assistance Program, reversed
Concentrated socioeconomic inequity subdomain	
Adults with advanced degrees	Percentage of individuals with master's, professional or doctoral degrees
Very high-income households	Percentage of households with incomes at or above $125,000
Adults without high school degrees	Percentage of individuals without a high school degree, reversed
Very low-income households	Percentage of households with incomes below $20,000, reversed
Housing resources subdomain	
Broadband access	Percentage of households with connections to high speed broadband internet
Crowded housing	Percentage of housing units with >1 occupant per room, reversed
Social resources subdomain	
Mobility-enhancing friendship networks	The prevalence of high-socioeconomic status (SES) friends among low-SES individuals (economic connectedness)
Single-parent families	Percentage of family households that are headed by a single-parent, reversed
Non-profit organizations	Density of non-profit organizations
Wealth subdomain	
Homeownership rate	Percentage of housing units that are owner-occupied
Aggregate home values	Aggregate home value (dollars) divided by the number of children aged 0-17
Aggregate capital income	Aggregate interest, dividends, or net rental income (dollars) divided by the number of children aged 0-17
Aggregate real estate taxes	Aggregate real estate taxes paid (dollars) divided by the number of children aged 0-17

These variables encompass some of the most widely-utilized single-indicator measures of neighborhood resources, and have been consistently identified as among the strongest neighborhood-level predictors of children's outcomes.[3; 9; 14; 173; 174; 188-190] Neighborhoods high in economic resources have more capital to invest in amenities that rely on local funding, such as schools, parks and after-school programs. Higher income neighborhoods also have greater purchasing power, attracting private businesses and service providers. High-quality jobs and the economic resources they provide also have less tangible benefits: The sociologist William Julius Wilson proposed a causal link between the disappearance of jobs in poor, urban neighborhoods and the resulting breakdown of social structures.[15] Recent studies have underscored the importance of jobs and economic resources in shaping inequality of opportunity across U.S. communities. These studies link economic shocks from globalization not only to reduced employment and earnings,[191] but also to reduced family formation, as well as increased mortality rates, notably due to drug and alcohol poisoning, HIV infection, and homicides.[191; 192]

CONCENTRATED SOCIOECONOMIC INEQUITY. While the employment and economic resource subdomains include measures of median earnings and median household income, measures of central tendency do not capture all important facets of the distribution of resources within neighborhoods. Research employing the Index of Concentration at the Extremes (ICE), an index measuring the prevalence of either very rich or very poor households, has been shown to demonstrate superior predictive validity compared to single-indicator measures of neighborhood resources, such as the poverty rate.[193; 194] Other studies have documented non-linear associations between area-level social determinants of health and health outcomes or life expectancy, where outcomes decline much more rapidly at the very low end of the income distribution.[195; 196] We therefore included two indicators capturing the extremes of the income distribution: the percentage of households with incomes above $125,000 per year and the percentage of households with incomes less than $20,000 per year. We also included two indicators capturing the extremes of the education distribution: the percentage of individuals with an advanced degree (master's, professional or doctoral) and the percentage of individuals without a high school degree.

HOUSING RESOURCES. Housing quality is one channel through which unequal opportunities translate into unequal outcomes.[197-199] The COI includes two indicators of housing quality that are predictive of children's outcomes. The first of these indicators is crowded housing conditions, which have been linked to children's health and educational attainment by disrupting sleep and study schedules and increasing the risk of contracting infectious diseases.[197; 198; 200-206] Second, broadband internet connection availability increases access to information which has been linked to improved education, labor market and health outcomes. Having internet access, particularly in-home broadband, improves employment outcomes by facilitating job search[207-209], may improve students educational outcomes[210-212] and may have mitigated learning losses during the COVID-19 pandemic.[213; 214] Additionally, it improves access to health information and telehealth medicine.[215-217]

SOCIAL RESOURCES. Sociologists have long highlighted the importance of social capital for individual and community well-being.[14; 218-220] Drawing on the work of Chetty et al., we have included a measure of economic connectedness, or the prevalence of high-socioeconomic-status friends for low-socioeconomic-status individuals.[221] High-status friends can be a source of information, shape aspirations, provide mentoring or job contacts.[221] Drawing on the work of Putnam, we also include a measure of non-profit organization density.

Non-profits provide an institutional infrastructure that facilitate volunteering and the provision of community-focused resources, and therefore contribute to the formation of community social capital.[99; 218; 222-224] Furthermore, high neighborhood rates of single-headed households have been shown to have direct independent effects on children's long-term outcomes, even after controlling for economic factors that are strongly correlated with high rates of single-headed households (e.g., lower family incomes).[10; 174; 225] Potential explanations for this effect are reduced availability of parental supervision and weakened informal social control, as well as fewer (male) role models.[226-229]

WEALTH. Present day inequities in community wealth, such as those observed in home ownership and home values, reflect a long history of discriminatory policies and practices in the housing and real estate sector. These include direct discrimination, government redlining and exclusionary zoning, which have perpetuated substantial wealth inequality across neighborhoods, as well as significant disparities in access to wealthy neighborhoods among children with different racial/ethnic backgrounds.[16; 28; 230; 231] Wealth provides families with economic safety and a means to invest in their children's future. Wealth represents an important channel for the intergenerational transmission of opportunity,[232-235] but it also benefits communities through property taxes. Property taxes account for a significant share of local government revenue, therefore communities with more valuable residential properties generate greater returns that are used to finance local amenities, including schools, parks or public libraries. Increased local tax revenue also generates demand for housing, resulting in higher home values.[236-243] Wealth also benefits community amenities through local philanthropy and charitable giving.[244] COI 3.0 includes four indicators capturing different facets of neighborhood wealth: home ownership, home values, real estate taxes paid and capital income.

SHOULD RACE/ETHNICITY BE INCLUDED AS A COMPONENT INDICATOR?

We do not include measures of neighborhood racial/ethnic composition, such as the percentage of Black residents, in the Child Opportunity Index. Including these measures would, as Dyer et al. note, confuse the causes and effects of structural racism.[245] While studies have found causal effects of neighborhood racial/ethnic composition, for example, on intergenerational mobility,[246; 247] it is not the presence of specific racial/ethnic groups per se that generates inequities in opportunity across neighborhoods. Rather, it is racial/ethnic discrimination and structural racism that have generated and sustain structural inequities across neighborhoods where children of different races/ethnicities reside. For example, government redlining in the 1930s has resulted in lower homeownership and lower accumulation of wealth in predominantly Black communities.[29] Our approach is to directly measure these structural features—lower homeownership and lower wealth—rather than use race as a proxy. Furthermore, by omitting racial/ethnic composition from the COI itself, we separate structural features of neighborhoods from the racial/ethnic composition of neighborhoods, allowing us to quantify racial/ethnic inequities in access to neighborhood opportunity. Those inequities will then not be confounded by the inclusion of race/ethnicity into our measure of neighborhood opportunity.

DATA AND METHODS

OVERVIEW

This section covers general concepts underlying the construction of COI 3.0 and describes the methodology used to construct the index from its component indicators. Appendices 1 through 4 contain more detailed information on the construction of component indicators, including definitions and sources (Appendix 1), data on schools (Appendix 2), data on non-profit organizations (Appendix 3) and point to block aggregation (Appendix 4). Appendix 5 describes our protocol for crosswalking data from 2010 census tracts geographies to 2020 census tract geographies and vice versa. Appendix 6 details our protocol for reconciling changes in 2010 census tract definitions.

All COI 3.0 component indicators were sourced and constructed to be comparable over time and across the U.S. COI 3.0 is constructed from 44 neighborhood indicators, listed in Tables 3 through 5 above and in Appendix 1. The indicators are drawn from different sources and harmonized into a single, common format: annual five-year moving-average census block data for census blocks covering the period from 2012 to 2021. We have published census tract-level COI 3.0 subdomain, domain and composite index metrics for 2010 census tracts, i.e., census tracts as defined for the 2010 Decennial Census.

COI component indicators were drawn from public sources, including, for example, the Census Bureau, National Center for Health Statistics (NCHS), U.S. Department of Education and the Environmental Protection Agency (EPA). Two proprietary data sets were used: NatureScore, a metric of healthy natural environments produced using satellite imagery, and data from DataAxle (formerly Infogroup) on the location of restaurants and food retailers.

The source data that we construct COI indicators from is published at different geographic levels, ranging from point data (e.g., school-level data) to ZIP codes. When multiple geographic levels of data were available, we used the most granular level. The only exceptions are indicators sourced from the American Community Survey where we used census tract-level data for every indicator because more granular block group data is not consistently available over time. The only metric sourced at a more aggregate level than census tracts is data on friendship networks, which was available for ZIP codes only.

COI source indicators are measured on different time scales, ranging from hourly, e.g., ambient temperature, to five-year averages, e.g., census tract data from the American Community Survey (ACS) such as the poverty rate. For consistency, all indicators that were not available as five-year averages in their raw form were converted to five-year averages. Pooling of observations across years improves the reliability of spatially granular estimates, smooths out short-term fluctuations and helps to identify persistent, structural inequities across neighborhoods. For example, we average temperature data over a five-year period to smooth over short-term weather variation and identify more persistent local and regional temperature differentials.

Most COI component indicators are available for all years covered by COI 3.0 and are constructed using consistent methodology over that time period. However, for some indicators, such as those derived from school-level data, we lack information from after 2018. For other indicators, time-varying data could not be obtained at all. Appendix 1 contains detailed information on the sources and temporal availability of each component indicator.

To combine component indicators measured on different scales, e.g., percent or dollars, into an index, the raw values of each indicator were standardized using the common z-score transformation. We then combined standardized indicator z-scores into 14 subdomain scores using weights that reflect the strength of the association between each indicator and four health and socioeconomic outcomes. Using the same weighting approach, the subdomain z-scores were then aggregated into an overall index z-score and into three domain scores: education, health and environment, and social and economic. Lastly, we constructed easily interpretable metrics, Child Opportunity Scores and Child Opportunity Levels, for every subdomain, domain and the overall index. Child Opportunity Levels group neighborhoods into five ordered levels, labeled very low-, low-, moderate-, high- and very high-opportunity. Child Opportunity Scores group neighborhoods into one hundred ordered groups, and assign numeric labels from 1 (lowest opportunity) to 100 (highest opportunity).

We generate three different versions of the Child Opportunity Scores/Levels, i.e., nationally-normed, state-normed and metro-normed. For nationally-normed Scores/Levels, census tracts are ranked and compared to all other tracts in the U.S.; for state-normed Scores/Levels, census tracts are ranked and compared to all other tracts in that state; for metro-normed Scores/Levels, census tracts are ranked and compared to all other tracts in that metropolitan area. Metro- and state-normed metrics are not comparable across metro areas and states, respectively. See the section “Choosing between metro-, state- and nationally-normed Child Opportunity Levels and Scores” below for further guidance.

GEOGRAPHIC DEFINITION AND SOURCES

For every Decennial Census, the Census Bureau reviews and revises the delineation of census blocks, block groups and tracts. Census blocks are the smallest geographic summary level defined by the Census Bureau, and all COI component indicators are harmonized at the census block level. Census blocks defined for the 2010 and 2020 Decennial Censuses differ in many instances. We therefore harmonized every component indicator for both 2010 and 2020 block geographies. Census tract-level COI data is derived as a final step of index construction, by aggregating census block-level subdomain, domain and overall index composite z-scores from block- to tract-level.

In social science and health research, census tracts are often used as proxies for neighborhoods. Tracts are drawn to include an area of about 8,000 residents, and their boundaries generally follow visible or identifiable local boundaries, such as intersections, roadways, streams or other bodies of water and boundaries of administrative entities (e.g., cities, towns and counties). Census blocks nest perfectly within census tracts.

Lists of 2010 and 2020 census blocks and census tracts, as well as their population weighted centroids, were obtained from Census Bureau TIGER/Line shapefiles. To crosswalk data collected for 2010 geographies to 2020 geographies and vice versa, we used the Census Bureau’s census block relationship files, which map 2000 to 2010 blocks and 2010 to 2020 blocks (see Appendix 5). Metropolitan areas were defined using 2020 Office of Management and Budget (OMB) definitions. The 100 largest metro areas are defined as the metropolitan statistical areas with the largest population of children aged 0-17 in 2020. COI 3.0 metro-normed data is only available for the 100 largest metro areas.

DATA USED FOR CALCULATING WEIGHTS AND VALIDATION ANALYSES

To calculate indicator weights and for validation analyses, we relied on several data sources providing census tract-level data on health outcomes, socioeconomic outcomes, racial/ethnic composition and other metrics of neighborhood opportunity, deprivation or vulnerability.

Data used to calculate indicator weights

We used four variables to calculate indicator weights. Two of these variables were drawn from the Opportunity Atlas measures of intergenerational mobility. Based on Census Bureau data linked to IRS tax records, Chetty et al. estimated socioeconomic outcomes in 2015 for the birth cohort 1978 to 1983, aggregated to the census tract that individuals resided in during childhood and adolescence.[248] We used the following Opportunity Atlas variables for calculating COI indicator weights:

- Mean household income rank at age 35 for children with parents at the 50th percentile (median) of the parent income distribution
- Probability of living in a low poverty census tract at age 35 for children with parents at the 50th percentile (median) of the parent income distribution

We also used two health outcome variables from the CDC PLACES dataset.[249] These data are estimated using multilevel modeling combined with poststratification using survey data from the 2019 Behavioral Risk Factor Surveillance System (BRFSS).[250; 251] We used the following indicators for calculating weights:

- Mental health not good for 14 or more days among adults aged 18 and older
- Physical health not good for 14 or more days among adults aged 18 and older

Data used for validation analysis

We used the following variables from the Opportunity Atlas[248] for the validation analysis, all measured for children at the median of the parent income distribution: the probability of being incarcerated, having a teenage birth, having any earnings, having an income in the top one percent of the household income distribution, having an income in the top 20 percent of the household income distribution, having an income in the top one percent of the personal income distribution, having an income in the top 20 percent of the personal income distribution and personal income rank.

The following PLACES[249] variables, measured as percentages of the population aged 18 or older at the census tract-level, were used for validation analyses: all teeth lost, arthritis, asthma, cancer, chronic kidney disease, COPD, coronary heart disease, depression, diabetes, high blood pressure, high cholesterol, obesity, stroke, general health, annual checkup, cervical cancer screening, cholesterol screening, colorectal cancer screening, core preventive services (men), core preventive services (women), dental visit, health insurance, mammography, taking blood pressure medication, binge drinking, current smoking, physical inactivity, sleeping less than 7 hours.

We also used census tract-level life expectancy data from the CDC USALEEP project for validation analyses.[249; 252]

It is important to note that Opportunity Atlas, PLACES and USALEEP data are modeled estimates. In particular, PLACES and USALEEP use data on racial/ethnic composition and income/poverty in the estimation of census tract level health and life expectancy estimates.[250-252]

Other neighborhood metrics used in validation analyses include:

- Child Opportunity Index 2.0 overall index z-scores, 2017, from diversitydatakids.org.[248; 249]
- Social Vulnerability Index (SVI), 2016, from the CDC. The CDC constructed the SVI to enable communities and policymakers to better prepare for natural disasters, such as hurricanes, or anthropogenic hazardous events, such as harmful chemical spills.[40]
- Area Deprivation index (ADI), 2015, is published by the Neighborhood Atlas.[39; 253] The ADI is a composite index at the census block group level that captures neighborhood deprivation, such as lack of income, employment, education or access to health care. To obtain census-tract level estimates, we averaged ADI national percentiles (1-100) across block groups within census tracts using the number of children aged 0-17 in each block group as weights. The published ADI is constructed from component indicators that were not standardized before combining them into an index.[41] The ADI values are therefore nearly fully determined by the component indicators for home values and median household income, which are measured on a dollar scale while most of the remaining variables are proportions and vary between 0 and 1.[41; 43]
- Area Deprivation Index (ADI), corrected, 2017: We computed a corrected version of the Area Deprivation Index using 2017 5-year ACS data. Our corrected version uses the same component indicators (sourced at the census tract level), but standardizes them using the z-score transformation and then combining the standardized indicators into a composite z-scores using weights published by Kind et al.[253]
- Median household income, 2017, from 5-year ACS data (5-year ACS Table B19013, api.census.gov). Our analyses show that median household income was among the strongest single predictors of health and socioeconomic outcomes.
- Index of Concentration at the Extremes (ICE), 2017: The ICE is computed here from 5-year ACS data as the number of individuals residing in households with incomes above $125,000 minus the number of individuals residing in households with incomes less than $20,000.[193] This difference is divided by the total number of individuals in the census tract. The resulting metric varies between 1 (all individuals reside in households with incomes above $125,000) and -1 (all individuals reside in households with incomes below $20,000). While median household income measures variation across census tracts in terms of the income of the median household, the ICE captures variation in the prevalence of either very rich or very poor households.

To examine the association between different metrics of neighborhood opportunity and neighborhood racial/ethnic composition, we obtained data on the number of children of a given race/ethnicity from the 2017 American Community Survey (5-year ACS Table B01001*, api.census.gov) at the census tract-level.

ALGORITHM TO CONSTRUCT COMPOSITE INDEX

To construct the composite index, we take component indicators measured in their original scale, e.g., dollars or percent, and convert them to 5-year moving averages at the census block-level for the period from 2012 to 2021. The main steps of the algorithm that combines component indicators into subdomain, domain and overall index z-scores are as follows:

- Combine component indicators within a subdomain into a single dataset and remove census blocks that are missing more than 50% of the component indicators
- Top- and bottom-code raw component indicators
- Compute means and standard deviations for component indicators
- Standardize indicators by subtracting their mean and dividing by their standard deviation
- Top- and bottom-code every standardized indicator
- Compute indicator weights
- Combine block-level standardized and top/bottom-coded indicator data into a composite subdomain z-score using component indicator weights

We outline the algorithm for the construction of subdomain composite z-scores below. The algorithm to construct domain and overall index composite z-scores from subdomain composite z-scores is identical except that the inputs are subdomain composite z-scores rather than unstandardized component indicators.

Methodological differences for national-, state- and metro-normed COI

The algorithm produces three different sets of composite z-scores for subdomains, domains and the overall COI: metro-normed, state-normed and nationally-normed z-scores. For nationally normed z-scores, indicator means and standard deviations are computed across all census blocks in the U.S. For metro- and state-normed z-scores, indicator means and standard deviations are computed for each of the 100 largest metros (metro-normed) and each of the 50 U.S. states plus D.C. (state-normed).

Within-region (metro or state) standardization is necessary for the estimation of region-specific indicator weights, a new feature of COI 3.0. For COI 2.0, we estimated component indicator weights using data for all U.S. census tracts. For COI 3.0, we still follow this strategy for the nationally-normed version. For region-specific versions, however, we estimate region-specific component indicator weights that vary across metros (metro-normed) and states (state-normed) using ridge regression (REF).

Census block exclusions

We excluded all blocks for which one or more of the following criteria were met:

- The block was fully covered by water.
- More than 50% of component indicators combined within a subdomain had missing data.
- The block did not have at least one child aged 0-17 in either the 2010 or 2020 Decennial Census.

The total number of 2010 census tracts with COI 3.0 index data varies between 72,587 and 72,650 across years covered by COI 3.0 (2012-2021).

Top and bottom coding of outliers

Some component indicators had skewed distributions, for example, the metric measuring exposure to hazardous waste sites. Extreme skewness can produce very large z-score values after standardization, resulting in outliers that exert an outsize influence on the composite z-score. To reduce the impact of extreme outliers, we top- and bottom-coded every indicator in two steps. First, every raw indicator was bottom-coded at the 0.1th percentile and top-coded at the 99.9th percentile. This step was performed year by year for every indicator. Second, after standardization, we bottom-coded indicators with z-scores below –4 standard deviations and top-coded indicators with z-scores above +4 standard deviations, which impacted only a small subset of the component indicators.

Standardization

Component indicators are measured on different scales, for example, percent (e.g., poverty rate), U.S. dollars (e.g., median household income) or parts per billion (ozone concentration). In order to combine indicators measured on different scales into a composite index, some form of standardization is necessary. Standardization rescales the indicators so that they are all measured on a common scale. Failure to standardize would result in some variables exercising outsize influence on the composite index just because of the scale they are measured on.[41; 43]

For the nationally-normed version, we performed the common z-score standardization for each raw or untransformed indicator x_j using the following formula:

$$z_{ijt} = \left(x_{ijt} - \mu_j\right) / \sigma_j$$

where i denotes census blocks, j denotes indicators, t represents time in years, and z_{ijt} is the standardized version of the untransformed indicator x_{ijt}. μ_j and σ_j are the unweighted arithmetic mean and unweighted standard deviation of indicator j, computed over all years t and all census blocks i for the nationally-normed version of the COI. The resulting z-scores for indicator j are comparable both across census blocks and over time. Standardization rescales each component indicator to have a mean of zero and standard deviation of one. The standardized (z-score transformed) component indicators are now on a common scale and can be combined into composite z-scores.

For the metro- and state-normed version versions, standardization was performed within metro areas or states. We computed unweighted means and standard deviations for every component indicator and every metro area and state separately, and then used the formula above to standardize component indicators for each metro area and state. Metro- and state-standardized indicator z-scores are only comparable within a given metro or state.

To ensure that higher values always indicate more opportunity, we standardized the directionality of each indicator by multiplying the standardized z-scores of some indicators by -1. Those indicators are labeled as “reversed” in Tables 3, 4 and 5.

General considerations for the derivation of indicator weights

When combining component indicators, an important consideration is the weight each indicator should receive. For COI 3.0, we use empirically-derived weights that reflect how important a given indicator is as a predictor of children's outcomes. A strong empirical determinant of children's long-term outcomes should have greater weight in the construction of subdomain, domain and overall scores. We obtain these weights by (1) computing bivariate correlations between an index of health and socioeconomic outcomes and component indicators within a subdomain and (2) rescaling the resulting correlation coefficients to reduce their variability.

Rescaling is necessary to ensure that subdomain scores are not completely determined by a single indicator that is much more strongly associated with outcomes than other indicators in the subdomain. The COI is designed to be a composite index, and our approach seeks to ensure that each component contributes meaningfully to the overall index. We also reduce variability because the associations between component indicators and outcomes are biased measures of their true causal effects. Potential biases can either artificially inflate or reduce the component indicator-outcome correlation and the indicator weights derived from them. By rescaling the correlation coefficients, we guard against potential biases resulting in an indicator having an outsize weight (or no weight at all) in the construction of a subdomain composite score. Rescaling preserves the ordering of component indicators in terms of the weight assigned to them within a subdomain but shrinks the variation across weights.

We report descriptive statistics on component indicator weights as part of our descriptive analysis below. Overall, our approach is conceptually very similar to COI 2.0. However, for COI 3.0, we compute separate weights for the national-, state-, and metro-normed COI versions, with the nationally-normed weights constructed using OLS regression and the state- and metro-normed weights constructed using ridge regression, as explained below.

Derivation of indicator weights for nationally-normed Child Opportunity Index

For the nationally-normed version of the COI, weights are obtained from bivariate regression of an outcome index on a given component indicator. The outcome index is based on the four outcome measures described above: mean household income rank and the probability to reside in a low-poverty census tract from the Opportunity Atlas, and metrics for mental and physical health from the PLACES database. All four outcomes are measured for 2010 census tracts. While health outcomes are measured in 2021, socioeconomic outcomes are measured in 2015. We use 2018 component indicator z-scores in the weight estimation.

Because the outcome index is measured for 2010 census tracts, we first aggregate the component indicator z-scores from 2010 census blocks to 2010 census tracts using block-level child population counts as weights. To compute the outcome index, we standardize the four outcome variables and norm the two health variables directionally by multiplying their standardized values by -1 so that higher values indicate better outcomes. We then average the four outcomes to create the final outcome index. We construct datasets that include each of the component indicator z-scores along with the outcome index, drop observations with missing data for either the outcomes or the component indicator z-score, and then standardize each to ensure that both outcome and predictor have a mean of 0 and standard deviation of 1. We then run a bivariate OLS regression of the standardized outcome index on the standardized component indicator and retrieve the beta regression coefficient for the component indicator z-score. In this way, we obtain standardized beta regression coefficients for all component indicators in a given subdomain.

We then rescale the beta coefficients to derive the final component indicator weights. We first bottom code the coefficients at 0.05. Some component indicators are only very weakly, or in some cases, negatively associated with the outcome index.[248] Bottom-coding ensures that every indicator at least has a minimum (positive) weight and contributes some variation to a subdomain composite score.

Furthermore, if, for a given subdomain, the ratio between the largest and smallest coefficient exceeds a constant c, we shrink the weights towards constant m by averaging each beta coefficient within the given subdomain with constant m. For a given subdomain, m is defined as

$$m = \beta_{max} - \frac{c \times \beta_{min}}{c - 1}$$

To compute m, we set c to be equal to 5, i.e., we shrink the beta coefficients towards m for a given subdomain whenever the largest beta coefficient, β_{max}, exceeds the smallest beta coefficient, β_{min}, by a factor of 5 or more. m is defined such that the maximum shrunken coefficient is no larger than c times (5 times) the minimum inflated coefficient. We then average each coefficient with the constant m. Coefficients with values greater than m are shrunk and coefficients with values less than m are inflated towards m. We apply shrinkage only if the ratio between the largest and smallest beta coefficient exceeds 5.

Lastly, regardless of whether or not bottom-coding or shrinkage was applied, we construct the final weights by rescaling so that they sum to one within each subdomain.

Derivation of indicator weights for state- and metro-normed Child Opportunity Index

For the state- and metro-normed versions, we use ridge regression to compute the underlying regression coefficients. While the following description focuses on the metro-normed version, we use the same approach for the state-normed version. The basic rationale for computing metro-area specific weights is that the extent to which certain indicators matter for overall opportunity can vary across metro areas. For example, the effect of summer time heat may vary across metro areas as a function of metro-area variation in average summer time temperatures and green space. Moreover, indicators that show less spatial variation within and more spatial variation across metro areas should be less important for explaining neighborhood differences in outcomes within metro areas, but more important for explaining variation across metro areas. For example, temperatures vary less across neighborhoods within a metro area than they do nationally. Therefore, metro-level weights for heat exposure should be smaller than the national weights for heat exposure. We therefore compute component indicator weights for each of the 100 largest metro areas (metro-normed version) and all 50 states plus D.C. (state-normed version).

Replicating the OLS regression approach used to derive the beta coefficients for the national version turns out to be problematic in a metro-based context. For at least some indicators and metro areas there is not enough information to robustly identify the coefficients. Some of the 100 largest metro areas have less than 100 census tracts, making it difficult to robustly identify the indicator-outcome association because of a relatively small number of cases. Some of the component indicators remain skewed despite the normalizations we apply and/or show relatively little variation within a metro area. As a result, and in particular when these issues arise in combination, we found that bivariate metro-specific OLS regression results could be sensitive to outliers and overfitting.

Instead of conducting metro-specific OLS regressions, we utilized outcome and component indicator data for all census tracts in the 100 largest metro areas combined. We then estimated a ridge regression model with metro area-specific coefficients that shrank poorly identified metro-specific coefficients towards a national average beta coefficient. The metro-specific ridge coefficients were close to their OLS analogs—the beta coefficient from a metro-specific OLS regression—whenever there is a clear, well-identified association between the outcome index and component indicator in a given metro area. When this association is less well-identified, for example, because the number of tracts in a metro area is small or the outcome or component indicator (or both) are strongly skewed or lack variation, the metro area-specific beta coefficient is shrunk towards a global national average coefficient estimated using all census tracts from the 100 largest metro areas. In other words, metro area-specific beta coefficients are indeed only metro area-specific if sufficient information is available to robustly identify them. Otherwise, they approximate the average metro area-coefficient common to all metro areas.

To implement this approach, we use a dataset that includes the outcome index and a given component indicator for all census tracts in the 100 largest metro areas, both standardized within metro areas. We then create an identical copy of this dataset and stack original (dataset 1) and copy (dataset 2). For a given indicator, the linear model is specified as follows:

$$y_{cm} = \alpha + \beta x_{cm} + \gamma_m d + \vartheta_m d x_{cm} + \varepsilon_{cm}$$

where y_{cm} is the outcome index y in census tract c and metro area m, and ε_{cm} is an idiosyncratic error term. α is a common intercept and β is the common standardized regression coefficient for indicator x. β quantifies the strength of the correlation between outcome index and component indicator across all census tracts in the 100 largest metro areas. $\gamma_m d$ and $\vartheta_m d x_{cm}$ are interactions between metro-specific intercepts (γ_m) and metro-specific slope coefficients ($\vartheta_m d$) specified as the interaction with an indicator variable d=1 for dataset 2 and zero otherwise, i.e., metro-specific intercepts and slopes are only estimated for observations in dataset 2, while α and β are estimated using all observations from both datasets. This model allows us to estimate the association between outcome index and component indicators across all census tracts in the 100 largest metro areas (β), and estimate 100 metro-specific associations (ϑ_m) as deviations from β, where we let the ridge estimator shrink the ϑ_m towards β. We need two copies of the data because otherwise it would not be possible to jointly estimate both β and one ϑ_m for each of the 100 largest metro areas in the same model.

We estimate this model using ridge regression with five-fold cross-validation using the glmnet R library (REF to Tibshirani/Hastie), specifying penalty factors equal to one for the ϑ_m coefficients and zero for all other coefficients. A penalty factor of zero exempts an estimated coefficient from shrinkage, i.e., we only shrink the ϑ_m coefficients. The resulting model allows us to estimate the average correlation between outcome and predictors across all census tracts in the 100 largest metros (β), and any metro-specific deviations from that national average (ϑ_m) that may be shrunk towards the national average β whenever metro-specific associations are poorly identified.

We extract the beta and ϑ_m coefficients, and obtain the metro-specific standardized correlation coefficient by summing β and ϑ_m for each metro area m. Lastly, we process—for a given subdomain and metro—the beta coefficients as described above, by optionally bottom coding and shrinking, and then rescaling the final weights to sum up to one across indicators for a given subdomain and metro area.

Construction of composite z-scores

After component indicator weights have been derived and rescaled so that they sum up to one within a subdomain, we compute subdomain composite z-scores by multiplying each indicator with its respective weight and sum over the products. If data on an indicator is missing for a given block, the weights for that block no longer sum up to one. In these cases, the weights for the indicators with available data for a given block are rescaled to sum up to one.

The subdomain composite z-score just constructed in turn becomes an input in the index construction algorithm just described. We use exactly the same algorithm that we used to combine component indicators into subdomain composite z-scores to combine subdomain z-scores into domain scores, and to combine subdomain z-scores into an overall index score. The algorithm thus produces, for both 2010 and 2020 census blocks, standardized z-scores for all component indicators and composite z-scores for all subdomains, domains and the overall index for the period from 2012 to 2021.

CHILD OPPORTUNITY METRICS

We use the composite subdomain, domain and overall z-scores to compute two easy-to-work-with metrics: Child Opportunity Scores and Child Opportunity Levels. Child Opportunity Scores divide neighborhoods into one hundred ordered groups and take values from 1 to 100. Child Opportunity Levels divide neighborhoods into five ordered groups that we label "very low," "low," "moderate," "high" and "very high." We provide Child Opportunity Scores and Levels for convenience and to facilitate use of the COI data. We also encourage users to use the composite z-scores as appropriate, for example to derive their own version of Child Opportunity Scores or Levels within their analysis sample. Composite z-scores, Child Opportunity Levels and Child Opportunity Scores are published as metro-normed, state-normed and nationally-normed versions.

COI 3.0 Child Opportunity Levels

Child Opportunity Levels are constructed so that they divide neighborhoods into five groups containing 20% of the child population each in 2021. We ordered neighborhoods in 2021 from lowest to highest in terms of their composite z-score. We then computed cut points (percentiles) that divide the neighborhoods in 2021 into five ordered groups containing 20% of the child population each (children aged 0-17, 2021 5-year ACS). We then applied the same cut points to data from other years as well. The resulting metric defines Child Opportunity Levels in terms of the 2021 distribution of children across neighborhoods with varying levels of opportunity. Child Opportunity Levels were defined in relation to the 2021 distribution of children across neighborhoods, therefore we observe exactly 20% of children at each level in 2021 data only.

Specifically, census tracts with composite z-scores at or below the 20th population-weighted 2021 percentile were sorted into the "very low" group. Tracts above the 20th and at or below the 40th population-weighted 2021 percentile were classified as "low opportunity." Tracts above the 40th and at or below the 60th

population-weighted 2021 percentile were classified as "moderate opportunity," tracts above the 60^{th} and at or below the 80^{th} population-weighted 2021 percentile were classified as "high opportunity" and tracts above the 80^{th} population-weighted 2021 percentile were classified as "very high opportunity."

In empirical analyses and for visualization, Child Opportunity Levels are often treated as a categorical variable. In our COI mapping application (diversitydatakids.org/maps), we visualize five distinct, uniquely color-coded opportunity levels. In empirical analyses, many studies have computed separate descriptive statistics for each of the five levels, or estimated separate regression coefficients for each of the five levels (omitting one of them as the reference group).

COI 3.0 Child Opportunity Scores

While Child Opportunity Levels divide neighborhoods into five ordered groups, Child Opportunity Scores divide them into 100 ordered groups. We ordered neighborhoods in 2021 from lowest to highest in terms of their composite z-score. We then computed cut points (percentiles) that divide the neighborhoods in 2021 into 100 ordered groups containing 1% of the child population each, and applied the same cut points to data from other years as well. The lowest ranked group of tracts was assigned a score of 1, the next lowest was assigned a score of 2, and so forth, until the top-ranked group, which was assigned a score of 100. Because Child Opportunity Scores are defined in relation to the 2021 distribution of children across neighborhoods, we observe exactly 1% of children at each level in 2021 data only.

In empirical studies, researchers have utilized Child Opportunity Scores as a continuous dependent variable that can be modeled using OLS regression.[27] As appropriate, users can also employ Child Opportunity Scores as a continuous predictor variable to summarize the association between an outcome and neighborhood opportunity using a single variable/coefficient, rather than the five Child Opportunity Level categories. If sample sizes permit, researchers can also treat Child Opportunity Scores as a categorical variable, which enables the identification of non-linear associations between an outcome and neighborhood opportunity at the very low and very high end of the opportunity distribution.

COI 3.0 Composite Z-Scores

Child Opportunity Scores and Levels collapse the information encoded in the composite z-scores into 100 distinct values (Scores) or 5 distinct values (Levels). Composite z-scores are the most detailed composite metric of neighborhood opportunity we publish. We encourage users to utilize the composite z-scores as appropriate. For example, users may use composite z-scores to derive their own version of Child Opportunity Levels whenever they find that the published Opportunity Levels do not efficiently divide their analysis dataset into five groups of roughly equal size. This can occur for several reasons:

- If data are drawn from an area with relatively high opportunity, there will be a disproportionate number of high- and very high-opportunity neighborhoods, and relative few low- and very low-opportunity neighborhoods.
- Opportunity Levels may also not be balanced in a sample from the early 2010s: Because opportunity grew over time (see following section), there are fewer high- and very high-opportunity neighborhoods in the earlier years covered by the COI.

- Users are working with data for a geographic area that does not correspond to an area for which we have published state- or metro-normed metrics, for example, if the area crosses state and/or metro-area boundaries, such as a hospital service area.
- For administrative uses or program implementation, users may require specific, exact divisions of their geographic areas into different opportunity levels.

By defining new opportunity levels within their analysis data, users will obtain the most balanced distribution of their analysis units (e.g., children, patients, health care encounters) across opportunity levels. Users can then compare their own opportunity levels to published Child Opportunity Levels to determine where in the national opportunity distribution their sample falls.

CHOOSING BETWEEN METRO-, STATE- AND NATIONALLY-NORMED CHILD OPPORTUNITY LEVELS AND SCORES

For users interested in data for a given metro area (state) only, we recommend the metro-normed (state-normed) opportunity metrics as a default. For users interested in areas located outside the 100 largest metro areas within the same state or when comparing some metro and some non-metro areas within the same state, we recommend the state-normed metrics. In all other cases, we recommend the use of nationally-normed metrics. Metro-normed (state-normed) opportunity metrics are only comparable within a given metro area (state) and are not comparable across metro areas (states).

States and metro areas across the U.S. differ in their levels of neighborhood opportunity. For metro areas that have high-opportunity levels compared to other metro areas nationwide, using the nationally-normed index conceals within-metro area inequalities because a disproportionate number of neighborhoods are assigned to the high- and very high-opportunity levels. However, there are situations in which even users interested in exploring a specific metro area may benefit from examining nationally-normed Opportunity Scores or Levels, as those metrics may provide important contextual information because they identify where a neighborhood ranks compared to all other neighborhoods in the U.S. The same reasoning and caveats apply to using state- versus nationally-normed data.

DESCRIPTIVE STATISTICS

The following section reports descriptive statistics on COI component indicators and composite z-scores. Data in Table 6 is based on census blocks, while data in Tables 7 and 8 is based on census tracts. Table 6 contains selected percentiles of the raw component indicator distributions in their native scale (e.g., percentages or dollars). All component indicators were harmonized so that they are measured at the same geographic scale, i.e., census blocks, and temporal scale, i.e., five-year moving averages. Here and below, we only report descriptive statistics for 2010 census blocks, i.e., census blocks as defined for the 2010 Decennial census. Table 6 reports percentiles for 2012, covering the five-year period from 2008 to 2012, and 2021, covering the period from 2017 to 2021.

Table 6 is based on component indicator data in their native scales after applying the inclusion criteria described above. The number of blocks with valid data varies between 7.2 and 8.1 million across indicators and years, with an average of 7.6 million blocks. Table 7 reports the same percentiles after component indicators were standardized, top and bottom coded and aggregated from 2010 census blocks to 2010 census tracts. The number of tracts with valid data varies between 70,425 and 72,650 across indicators and years, with an average of 72,380. Table 8 reports the same percentiles for composite subdomain, domain, and overall index z-scores. The number of tracts with valid data varies between 72,302 and 72,648 across indicators and years, with an average of 72,569.

We only drop a block if it is missing more than 50% of component indicators within a subdomain. If 50% or more of the indicators have valid data, we derive the composite z-scores using only the indicators with non-missing data. The same applies to blocks with missing data on subdomain z-scores when combining them into domain and overall index z-scores. Our approach results in a low overall missing rate that is mainly determined by our block inclusion criteria defined above. We observe 72,585 tracts in 2021 with non-missing overall composite index z-scores, or 99.4% of all 2010 census tracts.

Tables 6 through 8 provide evidence of an overall increase in neighborhood opportunity between 2012 and 2021. The 50th percentile of the overall index z-score grew by 26% of a standard deviation (Table 8). The increase in composite z-scores is largest for the first percentiles (+0.37) and then declines monotonically with the smallest increase at the 99th percentile (+0.19). We observe higher composite z-scores in 2021 than 2012 for virtually all subdomains, with the strongest gains driven by reductions in pollution, increased access to health resources, reductions in extreme socioeconomic inequities and improvements in secondary and access to post-secondary education, housing resources and employment. Growth at the 50th percentile is fastest in the health and environment domain, followed by the social and economic, and education domains.

At the component indicator-level (Table 7), we observe reductions in air pollution[254], and increases in health insurance coverage following the passage of the Affordable Care Act in 2010[255], as well as increases in high school graduation rates[256] as drivers of the overall increase in opportunity. We also observe gains in educational attainment—both a decrease of adults without high school degrees and an increase of adults with Bachelor's and advanced degrees—and improved economic indicators reflecting a period of sustained labor market expansion following the Great Recession[257], including increased employment rates, increased median incomes (though not at the first percentile) and reduced poverty rates. The decline in extreme heat exposure is due to the choice of reference periods: The 2012 estimate combines data from 2008 through 2012, which saw stronger heat waves than the 2017 to 2021 period.[258]

While growth has been faster in the lower tail of the neighborhood opportunity distribution (Table 8), the overall gaps in opportunity are still extremely large across neighborhoods. For example, the 25th percentile increased by 30% of a standard deviation, from -0.8 to -0.5, while the 75th percentile increased by 25% of a standard deviation, from 0.55 to 0.80. If these rates persist over the coming decades, it would take about 320 years for the 25th percentile to catch up with the 75th percentile.

Table 9 reports the component indicator weights used to construct the nationally-normed COI and reports the means and standard deviations for the metro- and state-specific weights derived. Our approach constrains the largest weight within a subdomain to be no larger than five times the smallest weight. For example, within the elementary education domain, reading and math test scores are assigned a weight of 0.65, while test score growth has a weight of 0.21 and poverty-adjusted test scores has the smallest weight of 0.13 (one fifth of the largest weight). In the social and economic domain, weights differ little from one another within subdomains. There is more variability in the subdomains in the education and health and environment domain than in the social and economic domain. The non-profit density indicators have a relatively small weights within their respective subdomains, accounting for much of the within subdomain variation in weights.

The metro- and state-normed weights display relatively little variation. The average standard deviation across component indicators 0.03 for state- and 0.04 for metro-specific weights. Two notable exceptions are the weights for NatureScore and heat exposure. The association between NatureScore and the outcomes used to calculate the weights is notably stronger within metro areas than within states or across the country. Within metro areas, access to green nature is typically higher in suburban areas, where we also tend to observe better health and socioeconomic outcomes compared to urban areas. For heat exposure, we observe stronger associations nationally compared to states and metro areas, reflecting the greater variation in climates across the U.S. compared to within regions.

TABLE 6. SELECTED PERCENTILES OF THE RAW COMPONENT INDICATOR DISTRIBUTIONS, 2010 CENSUS BLOCK DATA

	p1		p25		p50		p75		p99	
	2012	2021	2012	2021	2012	2021	2012	2021	2012	2021
Early childhood education										
Public pre-K enrollment	0.0	0.0	9.9	0.0	24.2	21.9	40.9	43.2	100.0	100.0
Private pre-K enrollment	0.0	0.0	0.0	0.0	11.8	5.7	29.3	26.1	100.0	100.0
Elementary education										
Reading and math test scores	-1.8	-1.8	-0.4	-0.4	0.1	0.1	0.6	0.6	1.9	2.0
Reading and math test score growth	-2.0	-2.1	-0.6	-0.6	-0.1	-0.1	0.4	0.5	1.9	1.9
Poverty-adjusted reading and math test scores	-1.8	-1.8	-0.4	-0.4	0.1	0.1	0.6	0.6	1.8	1.8
Secondary and postsecondary education										
Advanced Placement course enrollment	0.0	0.0	1.0	0.4	8.2	8.4	17.3	18.2	46.1	50.9
High school graduation rate	51.0	60.1	75.6	83.1	84.0	89.0	90.2	93.1	97.5	97.9
College enrollment in nearby institutions	10.5	9.6	25.2	24.6	32.7	32.2	43.6	43.2	82.5	81.1
Educational resources										
School poverty	2.7	6.4	35.5	39.3	52.4	55.0	67.8	71.3	95.6	97.2
Teacher experience	1.1	1.2	6.4	6.5	9.5	9.7	13.3	13.6	29.4	30.1
Adult educational attainment	3.4	4.2	12.6	15.4	18.6	22.8	29.6	35.9	74.3	80.3
Child enrichment-related non-profits	0.6	1.0	2.8	3.8	4.1	5.3	5.9	7.3	18.0	22.6
Healthy environments										
Walkability	1.8	1.8	4.5	4.5	6.7	6.8	10.2	10.5	17.8	18.0
NatureScore	0.0	0.0	0.3	0.4	0.5	0.5	0.7	0.7	1.0	0.9
Extreme heat exposure	0.0	0.0	10.6	6.0	33.0	23.4	78.4	60.0	170.6	173.0
Fast food restaurant density	23.1	31.3	56.4	58.8	62.8	65.3	68.8	71.2	85.2	87.0
Healthy food retailer density	13.3	10.7	41.4	35.1	50.6	43.9	60.2	53.4	89.5	83.3
Pollution										
Airborne microparticles	5.4	4.5	8.4	7.1	9.7	8.1	10.8	8.8	13.1	13.1
Ozone concentration	29.4	30.0	37.6	36.0	39.6	37.0	41.8	38.2	49.6	50.1
Industrial pollutants in air, water or soil	0.0	0.0	81.6	57.8	562.6	412.9	2919.4	1941.7	106539.3	56568.9
Hazardous waste dump sites	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.0
Safety-related resources										
Vacant housing	0.0	0.0	6.7	5.5	11.4	10.5	18.4	18.2	58.5	58.0
Community safety-related non-profits	0.1	0.1	0.6	0.7	0.9	1.1	1.5	1.7	5.8	7.2

	p1		p25		p50		p75		p99	
	2012	2021	2012	2021	2012	2021	2012	2021	2012	2021
Health resources										
Health insurance coverage	54.6	62.0	77.5	85.2	84.2	91.1	89.8	95.1	98.3	99.6
Health-related non-profits	0.1	0.2	0.8	1.1	1.3	1.7	2.1	2.5	8.0	9.7
Employment										
Employment rate	34.6	36.7	69.0	71.5	76.2	78.9	81.8	84.6	91.1	94.2
High-skill employment rate	9.1	10.0	24.1	26.6	30.4	34.0	38.6	43.6	68.9	74.7
Full-time year-round earnings	28439.0	28965.3	42223.2	43908.5	49330.7	51205.2	59313.9	62164.2	128158.5	139096.7
Economic resources										
Poverty rate	1.1	0.8	7.6	6.3	12.8	10.9	20.1	17.8	49.9	46.9
Public assistance rate	0.0	0.0	5.6	5.4	10.4	10.2	17.2	17.3	45.9	48.1
Median household income	23803.1	23315.6	48596.5	52387.8	61284.5	67051.3	79268.8	87718.7	175967.2	202827.0
Concentrated socioeconomic inequity										
Adults with masters, professional or doctoral degrees	0.4	0.5	3.6	4.5	5.9	7.5	10.4	13.4	37.8	43.3
Very high-income households	0.0	0.6	4.2	9.4	7.6	15.9	14.3	26.8	55.7	69.2
Adults without a high school degree	0.9	0.4	7.7	5.3	12.7	9.3	19.9	15.3	48.8	42.5
Very low-income households	1.9	0.8	11.2	7.5	17.5	12.4	25.3	19.1	52.7	46.0
Housing resources										
Crowded housing	0.0	0.0	0.5	0.3	1.5	1.5	3.3	3.5	20.9	19.7
Broadband access	15.7	18.1	46.0	53.0	59.9	66.4	73.2	78.4	92.2	94.3
Social resources										
Single-parent families	4.0	0.0	20.9	20.3	30.4	31.2	42.3	45.0	83.5	88.2
Non-profit organizations	1.5	2.5	6.4	8.5	9.6	12.1	14.2	17.4	59.8	69.9
Mobility-enhancing friendship networks	0.4	0.4	0.7	0.7	0.8	0.8	1.0	1.0	1.4	1.4
Wealth										
Homeownership rate	13.2	13.0	63.8	62.0	76.2	76.1	83.9	84.8	96.6	97.5
Aggregate home value per capita	34670.2	0.0	160205.2	173156.0	244327.2	280354.1	398890.0	475450.8	2153041.2	2806518.7
Aggregate capital income per capita	34.4	0.0	1841.0	1794.9	4063.3	4717.4	8611.8	10955.6	82789.9	107579.2
Aggregate real estate taxes per capita	181.6	0.0	1267.7	1411.4	2402.4	2756.2	4390.6	5246.3	19034.3	25052.4

TABLE 7. SELECTED PERCENTILES OF THE STANDARDIZED COMPONENT INDICATOR DISTRIBUTIONS, 2010 CENSUS BLOCK DATA

	p1		p25		p50		p75		p99	
	2012	2021	2012	2021	2012	2021	2012	2021	2012	2021
Early childhood education										
Public pre-K enrollment	-1.18	-1.18	-0.87	-1.18	-0.18	-0.25	0.61	0.68	3.20	3.20
Private pre-K enrollment	-0.90	-0.90	-0.90	-0.90	-0.27	-0.47	0.66	0.47	3.33	3.33
Elementary education										
Reading and math test scores	-2.21	-2.15	-0.71	-0.67	-0.04	-0.02	0.60	0.62	2.18	2.19
Reading and math test score growth	-2.38	-2.38	-0.66	-0.65	-0.03	-0.01	0.59	0.60	2.20	2.23
Poverty-adjusted reading and math test scores	-2.38	-2.32	-0.64	-0.62	0.01	0.03	0.66	0.67	2.48	2.53
Secondary and postsecondary education										
Advanced Placement course enrollment	-1.10	-1.10	-0.77	-0.79	-0.24	-0.23	0.42	0.49	2.55	2.82
High school graduation rate	-3.49	-2.56	-1.08	-0.29	-0.19	0.35	0.50	0.81	1.32	1.36
College enrollment in nearby institutions	-1.72	-1.81	-0.70	-0.77	-0.09	-0.14	0.65	0.61	3.27	3.18
Educational resources										
School poverty	-1.63	-1.68	-0.71	-0.85	0.01	-0.14	0.79	0.61	2.00	1.89
Teacher experience	-3.08	-3.48	-0.44	-0.54	0.18	0.14	0.69	0.66	1.50	1.50
Adult educational attainment	-1.42	-1.36	-0.84	-0.66	-0.36	-0.10	0.47	0.81	2.74	2.98
Child enrichment-related non-profits	-1.16	-1.03	-0.59	-0.30	-0.23	0.11	0.28	0.71	3.99	4.00
Healthy environments										
Walkability	-1.63	-1.63	-0.76	-0.76	-0.07	-0.07	0.93	0.93	2.00	2.00
NatureScore	-1.81	-1.75	-0.66	-0.54	0.04	0.18	0.70	0.84	1.61	1.76
Extreme heat exposure	-2.63	-2.70	-0.53	-0.19	0.45	0.64	0.81	0.89	1.01	1.01
Fast food restaurant density*	-1.95	-2.09	-0.42	-0.61	0.21	0.02	0.89	0.67	2.43	2.23
Healthy food retailer density*	-1.87	-2.12	-0.44	-0.83	0.17	-0.26	0.82	0.38	2.25	1.91
Access to healthy food*	-1.90	-2.15	-0.38	-0.68	0.24	-0.05	0.92	0.62	2.56	2.24
Pollution										
Airborne microparticles	-2.76	-2.58	-1.18	-0.02	-0.57	0.48	0.16	0.98	1.94	2.57
Ozone concentration	-2.82	-3.11	-0.64	0.09	-0.10	0.43	0.46	0.70	2.55	2.23
Industrial pollutants in air, water or soil	-3.91	-2.25	0.03	0.13	0.20	0.23	0.25	0.25	0.26	0.26
Hazardous waste dump sites	-3.68	-3.55	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

	p1		p25		p50		p75		p99	
	2012	2021	2012	2021	2012	2021	2012	2021	2012	2021
Safety-related resources										
Vacant housing	-4.00	-4.00	-0.66	-0.53	0.07	0.21	0.55	0.67	1.18	1.18
Community safety-related non-profits	-0.93	-0.86	-0.51	-0.34	-0.17	0.04	0.36	0.65	4.00	4.00
Health resources										
Health insurance coverage	-3.57	-2.50	-0.96	-0.04	-0.18	0.56	0.45	0.95	1.28	1.37
Health-related non-profits	-0.96	-0.87	-0.51	-0.30	-0.17	0.10	0.37	0.71	4.00	4.00
Employment										
Employment rate	-4.00	-3.89	-0.67	-0.36	0.05	0.32	0.58	0.85	1.52	1.74
High-skill employment rate	-1.89	-1.74	-0.78	-0.56	-0.22	0.05	0.52	0.86	2.51	2.82
Full-time year-round earnings	-1.33	-1.30	-0.67	-0.60	-0.26	-0.19	0.33	0.45	3.32	3.85
Economic resources										
Poverty rate	-3.65	-3.17	-0.60	-0.33	0.20	0.34	0.71	0.77	1.22	1.20
Public assistance rate	-3.21	-3.30	-0.36	-0.38	0.34	0.31	0.78	0.75	1.18	1.18
Median household income	-1.51	-1.54	-0.78	-0.66	-0.34	-0.18	0.27	0.50	3.02	3.60
Concentrated socioeconomic inequity										
Adults with masters, professional or doctoral degrees	-1.15	-1.11	-0.75	-0.62	-0.38	-0.17	0.32	0.65	3.50	3.92
Very high-income households	-1.13	-1.11	-0.87	-0.52	-0.58	0.01	0.06	0.89	2.66	3.41
Adults without a high school degree	-3.46	-2.74	-0.54	-0.16	0.19	0.44	0.65	0.80	1.16	1.17
Very low-income households	-3.86	-3.03	-0.86	-0.27	-0.04	0.38	0.57	0.82	1.33	1.35
Housing resources										
Crowded housing	-4.00	-3.69	0.00	-0.05	0.43	0.40	0.64	0.63	0.71	0.71
Broadband access	-2.75	-2.49	-0.82	-0.42	-0.01	0.31	0.67	0.88	1.53	1.60
Social resources										
Single-parent families	-3.01	-3.09	-0.67	-0.74	0.11	0.09	0.73	0.74	1.78	1.77
Non-profit organizations	-0.87	-0.78	-0.48	-0.29	-0.21	0.02	0.25	0.54	4.00	4.00
Mobility-enhancing friendship networks	-1.93	-1.93	-0.71	-0.71	-0.07	-0.07	0.68	0.68	2.41	2.41
Wealth										
Homeownership rate	-2.86	-2.89	-0.66	-0.74	0.22	0.18	0.79	0.78	1.47	1.47
Aggregate home value per capita	-0.76	-0.82	-0.47	-0.45	-0.22	-0.13	0.26	0.44	4.00	4.00
Aggregate capital income per capita	-0.45	-0.45	-0.37	-0.36	-0.24	-0.19	0.06	0.21	4.00	4.00
Aggregate real estate taxes per capita	-0.88	-0.93	-0.61	-0.60	-0.28	-0.20	0.35	0.53	4.00	4.00

*Note: *The fast food and healthy food retailer metrics are combined into the access to healthy food index. Only the latter is included in the calculation of the healthy environments subdomain score.*

TABLE 8. SELECTED PERCENTILES OF THE COMPOSITE SUBDOMAIN, DOMAIN AND OVERALL INDEX Z-SCORES, CENSUS BLOCK DATA FOR 2012 AND 2021

	p1		p25		p50		p75		p99	
	2012	2021	2012	2021	2012	2021	2012	2021	2012	2021
Subdomains										
Early childhood education	-1.19	-1.19	-0.75	-0.85	-0.25	-0.37	0.66	0.48	3.26	3.26
Elementary education	-2.40	-2.32	-0.69	-0.65	-0.06	-0.03	0.58	0.60	2.29	2.32
Educational resources subdomain	-1.71	-1.67	-0.72	-0.68	-0.11	-0.06	0.62	0.72	2.32	2.43
Secondary and post-secondary education	-2.79	-2.13	-0.96	-0.47	-0.25	0.10	0.49	0.72	2.16	2.22
Pollution	-3.42	-2.84	-0.89	0.07	-0.36	0.56	0.26	0.98	1.70	2.08
Health resources	-3.48	-2.42	-0.92	-0.01	-0.16	0.60	0.46	1.00	1.49	1.81
Safety-related resources	-4.00	-4.00	-0.70	-0.49	0.07	0.28	0.58	0.75	1.34	1.42
Healthy environments	-2.35	-2.48	-0.35	-0.36	0.30	0.29	0.88	0.82	1.91	1.78
Concentrated socioeconomic inequity	-2.38	-1.88	-0.86	-0.39	-0.25	0.19	0.41	0.87	2.20	2.56
Employment	-2.22	-2.05	-0.79	-0.56	-0.17	0.07	0.51	0.78	2.43	2.82
Economic resources	-2.84	-2.67	-0.62	-0.47	0.07	0.17	0.62	0.70	1.89	2.07
Housing resources	-2.64	-2.25	-0.75	-0.40	0.05	0.31	0.70	0.86	1.51	1.57
Social resources	-2.37	-2.40	-0.70	-0.71	0.00	0.02	0.70	0.74	2.15	2.20
Wealth	-2.31	-1.68	-0.58	-0.55	-0.13	-0.04	0.47	0.68	4.00	4.00
Domains										
Education	-1.99	-1.82	-0.78	-0.64	-0.17	-0.07	0.57	0.64	2.45	2.44
Health and environment	-3.09	-2.34	-1.04	-0.24	-0.21	0.48	0.48	1.03	1.67	1.97
Social and economic	-2.27	-1.99	-0.75	-0.52	-0.10	0.12	0.58	0.81	2.22	2.45
Overall Index										
Overall Child Opportunity Index	-2.15	-1.78	-0.82	-0.52	-0.16	0.10	0.55	0.80	2.22	2.41

TABLE 9. COMPONENT INDICATOR WEIGHTS WITHIN SUBDOMAINS FOR NATIONAL VERSION, MEANS/STANDARD DEVIATIONS OF WEIGHTS FOR STATE- AND METRO-NORMED VERSIONS

	National version weights	State version average weights	Standard deviation of state version weights	Metro version average weights	Standard deviation of metro version weights
Early childhood education					
Public pre-K enrollment	0.17	0.17	0.00	0.17	0.00
Private pre-K enrollment	0.83	0.83	0.00	0.83	0.00
Elementary education					
Reading and math test scores	0.65	0.63	0.04	0.60	0.06
Reading and math test score growth	0.21	0.18	0.04	0.18	0.05
Poverty-adjusted reading and math test scores	0.13	0.19	0.06	0.23	0.07
Secondary and postsecondary education					
Advanced Placement course enrollment	0.25	0.26	0.04	0.22	0.06
High school graduation rate	0.45	0.49	0.06	0.52	0.07
College enrollment in nearby institutions	0.30	0.25	0.05	0.26	0.09
Educational resources					
School poverty	0.39	0.39	0.02	0.36	0.02
Teacher experience	0.15	0.16	0.02	0.21	0.04
Adult educational attainment	0.38	0.37	0.02	0.36	0.02
Child enrichment-related non-profits	0.08	0.08	0.00	0.07	0.00
Healthy environments					
Fast food restaurant density*	0.79	0.60	0.09	0.60	0.12
Healthy food retailer density*	0.21	0.40	0.09	0.40	0.12
Healthy food access*	0.34	0.18	0.07	0.12	0.04
Walkability	0.13	0.16	0.06	0.11	0.02
NatureScore	0.15	0.44	0.15	0.51	0.09
Extreme heat exposure	0.38	0.23	0.09	0.26	0.11
Pollution					
Airborne microparticles	0.25	0.23	0.07	0.33	0.15
Ozone concentration	0.22	0.22	0.06	0.14	0.07
Industrial pollutants in air, water or soil	0.33	0.28	0.06	0.30	0.12
Hazardous waste dump sites	0.20	0.28	0.05	0.23	0.09
Safety-related resources					
Vacant housing	0.83	0.80	0.05	0.81	0.06
Community safety-related non-profits	0.17	0.20	0.05	0.19	0.06

	National version weights	State version average weights	Standard deviation of state version weights	Metro version average weights	Standard deviation of metro version weights
Health resources					
Health insurance coverage	0.83	0.83	0.00	0.83	0.00
Health-related non-profits	0.17	0.17	0.00	0.17	0.00
Employment					
Employment rate	0.31	0.31	0.02	0.28	0.05
High-skill employment rate	0.34	0.34	0.02	0.36	0.03
Full-time year-round earnings	0.35	0.36	0.01	0.36	0.02
Economic resources					
Poverty rate	0.33	0.33	0.01	0.33	0.01
Public assistance rate	0.34	0.34	0.01	0.34	0.01
Median household income	0.33	0.32	0.01	0.32	0.01
Concentrated socioeconomic inequity					
Adults with masters, professional or doctoral degrees	0.22	0.21	0.01	0.22	0.01
Very high-income households	0.26	0.26	0.01	0.26	0.01
Adults without a high school degree	0.23	0.26	0.01	0.26	0.01
Very low-income households	0.29	0.27	0.02	0.26	0.02
Housing resources					
Crowded housing	0.28	0.40	0.04	0.42	0.04
Broadband access	0.72	0.60	0.04	0.58	0.04
Social resources					
Single-parent families	0.43	0.44	0.03	0.45	0.03
Non-profit organizations	0.09	0.10	0.00	0.10	0.01
Mobility-enhancing friendship networks	0.47	0.47	0.02	0.46	0.02
Wealth					
Homeownership rate	0.21	0.23	0.02	0.26	0.03
Aggregate home value per capita	0.28	0.28	0.01	0.27	0.01
Aggregate capital income per capita	0.19	0.19	0.02	0.18	0.02
Aggregate real estate taxes per capita	0.32	0.30	0.02	0.29	0.02

*Note: *The fast food and healthy food retailer density metrics are combined into the access to healthy food index. Only the latter is included in the calculation of the healthy environments subdomain score.*

TABLE 10. SUBDOMAIN WEIGHTS WITHIN OVERALL AND DOMAINS INDICES

	National version weights	State version average weights	Standard deviation of state version weights	Metro version average weights	Standard deviation of metro version weights
Overall index					
Early childhood education	0.06	0.06	0.01	0.05	0.01
Elementary education	0.08	0.07	0.00	0.07	0.00
Educational resources	0.09	0.09	0.00	0.09	0.00
Secondary and post-secondary education	0.06	0.07	0.00	0.07	0.01
Pollution	0.02	0.02	0.00	0.03	0.01
Health resources	0.07	0.07	0.01	0.07	0.01
Safety-related resources	0.05	0.04	0.01	0.04	0.02
Healthy environments	0.03	0.04	0.01	0.04	0.01
Concentrated socioeconomic inequity	0.10	0.10	0.00	0.09	0.00
Employment	0.10	0.09	0.00	0.09	0.00
Economic resources	0.10	0.10	0.00	0.09	0.00
Housing resources	0.09	0.09	0.00	0.09	0.00
Social resources	0.09	0.09	0.00	0.09	0.00
Wealth	0.08	0.08	0.00	0.08	0.01
Education domain					
Early childhood education	0.19	0.19	0.02	0.18	0.02
Elementary education	0.27	0.26	0.01	0.26	0.01
Educational resources	0.33	0.32	0.01	0.31	0.01
Secondary and post-secondary education	0.22	0.23	0.01	0.24	0.02
Health and environment domain					
Pollution	0.09	0.10	0.02	0.15	0.06
Health resources	0.45	0.44	0.06	0.38	0.06
Safety-related resources	0.32	0.25	0.07	0.23	0.07
Healthy environments	0.14	0.20	0.07	0.24	0.05
Social and economic domain					
Concentrated socioeconomic inequity	0.18	0.18	0.00	0.18	0.00
Employment	0.17	0.17	0.00	0.17	0.00
Economic resources	0.18	0.18	0.00	0.18	0.01
Housing resources	0.15	0.16	0.01	0.17	0.01
Social resources	0.17	0.17	0.00	0.16	0.01
Wealth	0.13	0.14	0.01	0.14	0.01

VALIDATION ANALYSES

The following sections present evidence supporting the predictive and equity validity of the Child Opportunity Index. The analyses focus on the nationally-normed version of the COI, but the pattern of results is similar for the state- and metro-normed versions.

PREDICTIVE VALIDITY. Compared to single-indicator metrics of neighborhood context, such as neighborhood median household income, multi-dimensional neighborhood metrics capture numerous distinct neighborhood features that are causally linked to positive outcomes. By virtue of measuring numerous neighborhood features and causal pathways linked to socioeconomic and health outcomes, multi-dimensional neighborhood metrics have superior predictive validity compared to single indicator metrics of neighborhood context. We operationalize predictive validity of neighborhood metrics by measuring the strength of the association between neighborhood metrics and health/socioeconomic outcomes. COI 3.0 includes more distinct domains, or measures of distinct causal pathways; therefore it should have better predictive validity compared to other composite indices.

EQUITY VALIDITY. Structural racism has generated large racial/ethnic inequities in access to neighborhood opportunity. White and Asian children are growing up in neighborhoods that have more resources across multiple dimensions (economic, social, environmental, etc.), while Black, Hispanic, and Indigenous children grow up in neighborhoods that lack resources across multiple dimensions. Because composite metrics capture cumulative, multi-dimensional neighborhood advantages or disadvantages, they should be more strongly associated with neighborhood racial/ethnic composition than single-indicator metrics of neighborhood opportunity. We term this property of neighborhood metrics equity validity, and we quantify equity validity as the strength of the association between neighborhood metric and neighborhood racial/ethnic composition: A metric with high (low) equity validity is strongly (weakly) associated with neighborhood racial/ethnic composition. Specifically, we will quantify the extent to which children of different races/ethnicities are concentrated at either the top or bottom of the neighborhood distribution using different neighborhood metrics. Because COI 3.0 includes more distinct dimensions/domains, it should have greater equity validity compared to other composite indices.

DATA AND METHODS

The validation analyses use 29 health and 7 socioeconomic outcome variables (see above) to quantify predictive validity. Health outcomes include life expectancy and measures of health status, conditions and health care access taken from the CDC PLACES database. Socioeconomic outcomes include measures of intergenerational social and economic mobility from the Opportunity Atlas. All variables are measured at the census tract-level and are available for virtually all U.S. census tracts. COI data is measured for 2018. All other variables are measured between 2015 and 2021.

To measure predictive validity, we run bivariate OLS regressions of a given health or socioeconomic outcome on a given predictor. Predictors are the COI and its components as well as other neighborhood metrics, such as neighborhood median household income or the Area Deprivation Index. Outcomes and predictors are standardized over the set of observations with non-missing data for both variables. We run OLS regressions of each outcome on each predictor and record the R^2 values. For a given predictor, we then quantify predictive validity as the average R^2 value over all socioeconomic outcomes and as the average R^2 value over all health outcomes.

To measure equity validity, we do not use OLS regression to quantify the association between neighborhood metrics and racial/ethnic composition because it tends to produce results that were sensitive to minor specification changes. This sensitivity itself reflects the strong relationship between race/ethnicity and neighborhood opportunity. Different racial/ethnic groups are concentrated in different parts of the neighborhood opportunity distribution, which makes it difficult to quantify the association between neighborhood racial/ethnic composition and neighborhood metrics using linear models.[188; 259]

Instead, we construct an index to quantify the extent to which a given neighborhood metric reproduces racial/ethnic inequities in access to neighborhood opportunity. This equity validity index I_j follows the definition of the Index of Concentration at the Extremes[193] and is defined as

$$I_j = \frac{(h_j - k_j)}{n_j}$$

where h_j is the number of children aged 0-17 belonging to racial/ethnic group j residing in the top 20% of neighborhoods as defined by a given neighborhood metric. k_j is the number of children belonging to racial/ethnic group j residing in the bottom 20% of neighborhoods as defined by a given neighborhood metric. n_j is the total number of children belonging to racial/ethnic group j. The resulting metric will vary between -1 (all children in group j reside in bottom 20% of neighborhoods) and +1 (all children in group j reside in the top 20% of neighborhoods). We compute I_j separately for Asian, Black, Hispanic, Indigenous (American Indian or Alaska Native) and non-Hispanic White children. We find that for all metrics, Asian and White children have I_j values greater than zero for all neighborhood metrics, while Black, Hispanic and Indigenous children have values less than zero for all neighborhood metrics.

To simplify presentation of results, we take the absolute value of I_j and average the I_j across all racial/ethnic groups for a given neighborhood metric. The resulting metric measures the degree to which children of different races/ethnicities are concentrated at either the top or bottom of the neighborhood distribution. Larger values indicate greater concentration at either top or bottom of the neighborhood distribution.

RESULTS

Predictive validity of COI component indicators

The following analyses show that the COI components are predictive of important health and socioeconomic outcomes. We will also discuss some limitations to the predictive validity analyses.

Using census tract level data, we estimated OLS regressions of socio-economic outcomes on COI 3.0 component indicators. For a given component indicator, we then averaged the R^2 values from these regressions separately across health and socioeconomic outcomes. We report the resulting two averaged R^2 values for every component indicator in Figures 1a and 1b. The sign of the regression coefficients was positive in all but a few cases. If the association between a component indicator and outcomes was negative, we multiplied the R^2 value from these regressions with -1 so that we can still display results for those indicators alongside the others in Figure 1b.

Overall, COI component indicators are more predictive of health than socioeconomic outcomes. This in part reflects the fact that the health outcomes are model-based estimates, and the models used to estimate them include indicators taken from the American Community Survey. Some of these indicators (or closely related indicators calculated from the same underlying sample) are included in the Child Opportunity Index, but also other composite neighborhood indices such as the Area Deprivation Index (ADI) or Social Vulnerability Index (SVI). The socioeconomic outcomes, while also model-based, are not estimated as a function of ACS variables, and therefore there is no "built-in" association between them and COI components or neighborhood metrics.

In part reflecting this deterministic relationship, most of the variables that are strongly associated with health and socioeconomic outcomes are sourced from the American Community Survey (ACS). School poverty, reading and math test scores and mobility-enhancing friendship networks are among the most predictive variables that are not—or in case of friendship networks, only partly—sourced from the ACS. These three variables are also among the top-five most powerful predictors of socioeconomic outcomes. Around half of the component indicators have average R^2 values of 10% or less, and about one third have average R^2 values of five percent or less.

Figure 1 also shows that median household income has the highest predictive validity across both sets of outcomes. This is why we have selected it as the single-indicator metric of neighborhood context in the analysis comparing multi- and single-indicator metrics below.

We caution against placing too much emphasis on the predictive validity analysis in the overall assessment of neighborhood metrics. The bivariate associations reported here are biased estimates of the causal effect of neighborhood variables on outcomes. The estimated associations are likely upwardly biased by failing to control for important individual-level confounders. For health outcomes, the estimated associations are also upwardly biased because these outcomes are estimated using variables that are included in the COI, as well as the ADI and SVI. Similarly, estimated association might sometimes be downward biased. For example, we observe small and even negative associations for public pre-K enrolment, likely reflecting a selective rollout of public pre-K in lower opportunity areas, while private pre-K has a moderately strong association with different outcomes. The negative association for NatureScore likely reflects the prevalence of high NatureScore values in rural areas with older, less healthy populations. These limitations are further discussed below.

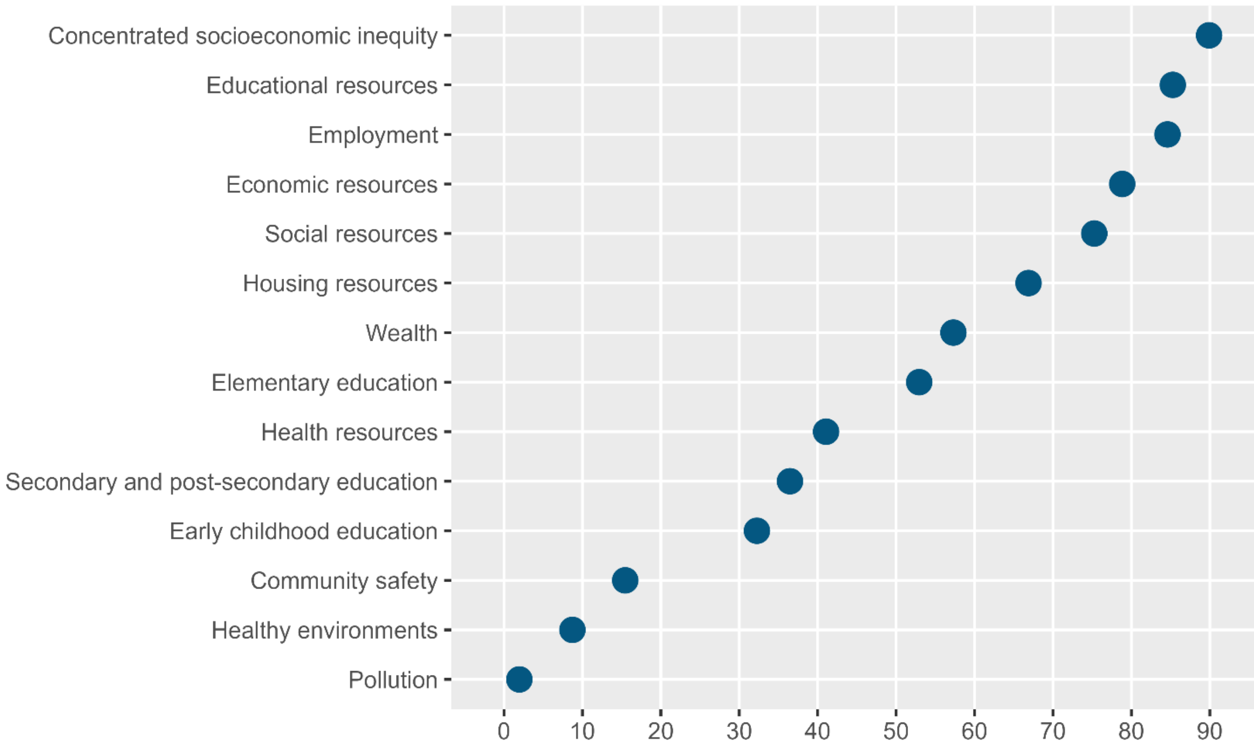
Association between the COI 3.0 overall index and its components

Figure 2 reports R^2 values from bivariate OLS regressions of the COI 3.0 overall composite index z-score on the subdomain z-scores. It is constructed using 2018 COI data. The figure shows that the overall COI is strongly associated with many of its subdomain components. The shared variation between the overall COI and its subdomains exceeds 50% for eight of the 14 subdomains, and exceeds 25% for 11 of the 14 subdomains. Three of the four subdomains associated with the health and environment domain—community safety, healthy environments and pollution—share the least variation with the overall COI.

FIGURE 1. AVERAGE PERCENT VARIANCE EXPLAINED IN HEALTH AND SOCIOECONOMIC OUTCOMES BY COI 3.0 COMPONENT INDICATORS

Note: Using census tract level data, we estimated OLS regressions of 29 health and 7 socioeconomic outcomes on COI component indicators listed here. For a given predictor, we averaged the R^2 values from these regressions separately across health and socioeconomic outcomes. In a few cases, the average association between component indicators and outcomes was negative. To preserve this information, we multiplied the R^2 value from these regressions with -1 for display in this figure.

FIGURE 2. PERCENT VARIANCE EXPLAINED IN OVERALL COI 3.0 COMPOSITE Z-SCORE BY COI 3.0 SUBDOMAIN Z-SCORES



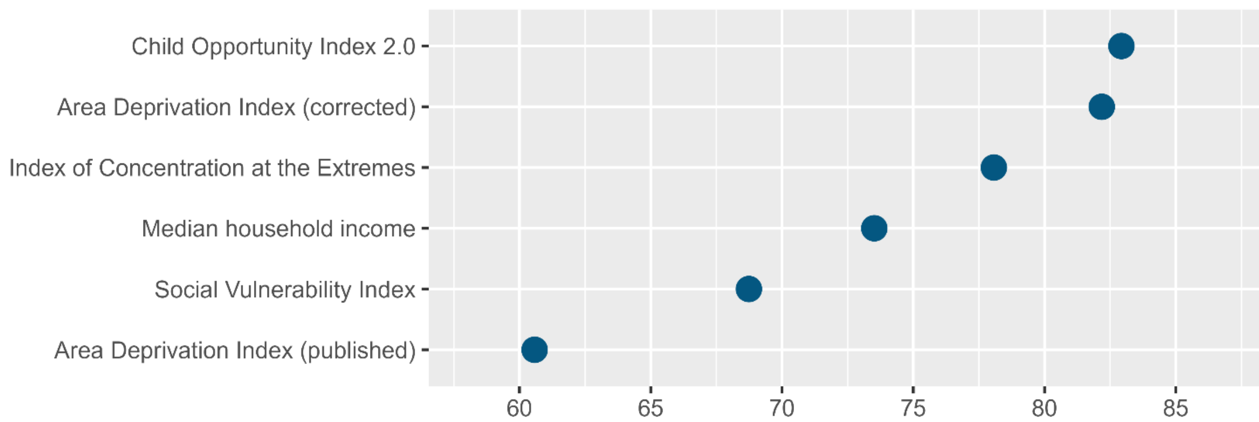
Note: The COI 3.0 composite z-score is a weighted average of 14 subdomain composite z-scores. The subdomain-specific weights for the nationally-normed version of COI 3.0 are reported in Table 10. For this figure, we report R^2 values from bivariate OLS regressions of the standardized overall COI 3.0 composite z-score on the standardized composite subdomain z-scores it is constructed from.

These patterns show that the overall COI strongly reflects socioeconomic inequities across U.S. neighborhoods, i.e., inequities in schooling, educational attainment, employment, income and wealth, which are central to understanding why neighborhoods differ in terms of the opportunities they provide. At the same time, the COI also includes indicators and subdomains that show no or only a weak association with the overall COI, i.e., variables that are largely uncorrelated with school-related and socioeconomic variables reflected in the overall index. While uncorrelated with neighborhood school and socioeconomic variables, they nevertheless represent important neighborhood features that have well-documented effects on health and socioeconomic outcomes.

Association between COI 3.0 and other neighborhood metrics

Figure 3 reports R^2 values from bivariate OLS regressions of the COI 3.0 overall index z-score on other neighborhood metrics. The other metrics are: Child Opportunity Index 2.0 overall index z-score, a corrected version of the Area Deprivation Index (ADI) with component indicators properly standardized, the ADI as published by the Neighborhood Atlas, the Social Vulnerability Index published by the CDC, the Index of Concentration at the Extremes (income-version) and census tract median household income from the American Community Survey.

FIGURE 3. PERCENT VARIANCE EXPLAINED IN OVERALL COI 3.0 COMPOSITE Z-SCORE BY OTHER NEIGHBORHOOD METRICS



Note: This figure reports R^2 values from bivariate OLS regressions of the standardized overall COI 3.0 composite z-score on other neighborhood metrics.

Figure 3 indicates that COI 3.0 has a moderate to strong association with all neighborhood metrics examined here. COI 3.0 is most strongly associated with the COI 2.0 overall index z-score (R^2 = 82.9%) and the corrected Area Deprivation Index (R^2 = 82.2%). It is weakly correlated with the published Area Deprivation Index. While the associations between COI and the other metrics are strong, the metrics are not identical. The differences between metrics will become consequential once they are used, for example, to define eligibility for subsidies or services.[43; 260]

Predictive and equity validity of COI 3.0 and other neighborhood metrics

For a comparative analysis of predictive validity, we ran bivariate OLS regression for every outcome-metric pair and recorded the R^2 values from each regression. For a given metric, we measure predictive validity as the average R^2 value over all socioeconomic outcomes and as the average R^2 value over all health outcomes. Figure 4 summarizes the results.

As we found for the component indicators of COI 3.0, all neighborhood metrics explain more variation in health than in socioeconomic outcomes. COI 3.0 explains more variation in health (average R^2 = 36.2%) and socioeconomic outcomes (average R^2 = 26.8%) than any other metric. It is followed by COI 2.0 and the corrected Area Deprivation Index (ADI). The Social Vulnerability Index (SVI) explains the least variation in both health and socioeconomic outcomes. The predictive validity advantage of COI 3.0 relative to its rivals is larger for socioeconomic than health outcomes. For example, the average R^2 for health outcomes is 36.2% for COI 3.0 and 33.6% for the corrected Area Deprivation Index, an eight percent advantage. The corresponding average R^2 values for socioeconomic outcomes are 26.8% for COI 3.0 and 22.8% for the corrected ADI, an 18% advantage.

To quantify equity validity, we computed an index that measures the extent to which children of a given race/ethnicity are concentrated at either the top or bottom of the neighborhood distribution. We computed this index for five racial/ethnic groups using each of the neighborhood metrics. For Figure 3, we averaged the absolute value of the index across racial/ethnic groups for a given metric, and multiplied it by 100. Larger values indicate greater concentration of racial/ethnic groups at either top or bottom of the neighborhood distribution measured using one of the seven metrics. COI 3.0 again outperforms the other metrics in terms of equity validity, with the exception of the SVI. The SVI has high equity validity, but it is also the only metric that includes one component indicator measuring racial/ethnic composition (percentage not non-Hispanic White) and a measure that is highly correlated with migrant status (English language skills), which contribute to its superior performance in this test.

For Figure 5, we averaged the two R^2 values shown for each metric in Figure 4 (health and socioeconomic outcomes) and plotted this average predictive validity estimate against the equity validity index results also shown in Figure 4. We hypothesized that multi-indicator, composite indices outperform single-indicator neighborhood metrics both in terms of predictive and equity validity, which is largely supported by the data shown in Figure 5. COI 3.0, COI 2.0 and the corrected ADI outperform the ICE (two indicators), published ADI (de-facto two indicators), and median household income (single indicator).

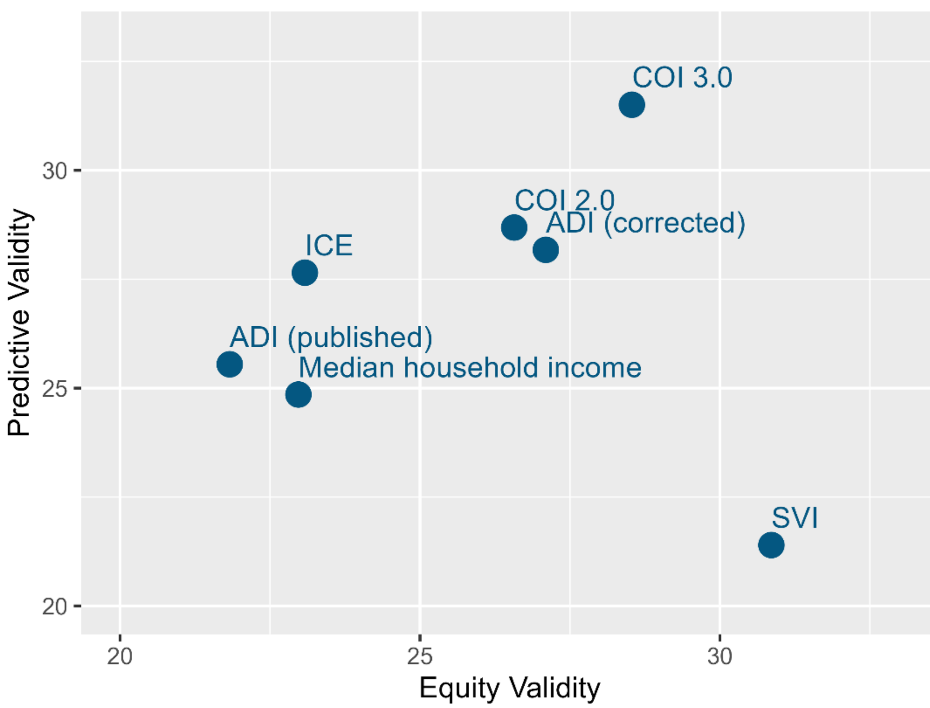
FIGURE 4. ASSOCIATION OF SEVEN NEIGHBORHOOD METRICS WITH NEIGHBORHOOD HEALTH OUTCOMES, SOCIOECONOMIC MOBILITY AND RACIAL/ETHNIC COMPOSITION

Note: To quantify the strength of the association between seven neighborhood metrics and census tract health and socioeconomic outcomes, we estimated bivariate OLS regressions for every pair of outcomes (29 health and 7 socioeconomic mobility outcomes) and neighborhood metric. For a given neighborhood metric, we averaged the R^2 values from these regressions separately across health and socioeconomic outcomes. To quantify the association between neighborhood metrics and neighborhood racial/ethnic composition, we computed an index that measures the extent to which children of a given race/ethnicity are concentrated at either the top or bottom of the neighborhood distribution. Larger values indicate greater concentration of racial/ethnic groups at either top or bottom of the neighborhood distribution measured using one of the seven metrics.

The results for the two versions of the ADI, which use the same components but different methods for index construction, support both hypotheses concerning the superior predictive and equity validity of composite indices. The published ADI nearly exclusively reflects variation in two variables: neighborhood home values and neighborhood income. In contrast, the corrected ADI is a true composite index that reflects variation in all of its component indicators. The results show that the corrected ADI – because it reflects variation in 17 rather than two variables – outperforms the published ADI in terms of both predictive and equity validity.

Figure 5 also shows that the ICE outperforms median household income predictively. Both metrics measure neighborhood income. The ICE measures variation in the tails of the neighborhood income distribution, i.e., variation in the prevalence of either very high- or very low-income households, while median household income focuses on the center of the neighborhood income distribution. In other words, for prediction, it is more important to know whether there are many very high- or very low-income households in the neighborhood than knowing the median household income.

FIGURE 5. PREDICTIVE AND EQUITY VALIDITY OF SEVEN NEIGHBORHOOD METRICS



Note: For Figure 5, we averaged the average R^2 values for health and socioeconomic outcomes for each neighborhood metric displayed in Figure 4. This average is plotted on the y-axis (predictive validity). The x-axis shows the index measuring the concentration of racial/ethnic groups at either top or bottom of the neighborhood distribution measured using one of the seven metrics as shown in Figure 4.

DISCUSSION

The preceding analyses show that COI 3.0 is moderately to highly correlated with other leading composite indices, but outperforms these indices both in terms of predictive validity and equity validity. It is more strongly associated with adult health and socioeconomic outcomes (predictive validity) and neighborhood racial/ethnic composition (equity validity) than either COI 2.0, published ADI, corrected ADI or SVI. The SVI does well in terms of equity validity, in part because it includes measures of racial/ethnic composition, but it is the worst performing metric among the ones tested here in terms of predictive validity.

The results presented here suggest that the COI 3.0 has desirable properties for health equity-focused research and applications. Scholars seeking a summary measure of neighborhood context will tend to find the neighborhood effects or inequities using COI 3.0 that are as large as or larger than those found using alternate composite indices. Similarly, policymakers and practitioners will find that the COI powerfully measures neighborhood structural inequities that are highly predictive of long-term health and socioeconomic outcomes. They will also find that the COI exhibits a stronger association with racial/ethnic neighborhood composition than most rival metrics, even though the COI does not include measures of racial/ethnic composition among its components. Therefore, utilizing the COI for the spatial allocation of resources will tend to benefit groups historically most affected by structural racism, and help to alleviate the profound racial/ethnic inequities in access to neighborhood opportunity generated by structural racism.

The ADI, SVI and ICE have been designed to quantify neighborhood inequities across age groups, whereas the COI has been designed to focus specifically on children. There are three potential explanations for the COI's relatively strong performance compared to these general population metrics in predicting adult socioeconomic and health outcomes: First, adult neighborhood contexts mirror the neighborhood contexts experienced in childhood and adolescence. Residing in a low-opportunity neighborhood as an adult therefore often reflects exposure to low-opportunity neighborhood environments from an early age, which has cumulative effects on educational attainment, labor market outcomes, health and social networks. Second, high-opportunity neighborhoods might reduce parental stress by providing families with more resources that help their children thrive. Third, regardless of whether adults have children or not, the majority of the COI component indicators measure features that affect adults too, including access to social and economic resources and healthy neighborhood environments.

An important limitation of the predictive validity analysis using health outcomes is that these outcomes are constructed in part as a function of the component indicators included in the different metrics compared here. As a result, all composite indices using ACS variables have a built-in association with health outcomes, and that association will be stronger the more closely the index components resemble the variables used to estimate the health outcomes. In a similar way, the high equity validity of the SVI reflects a built-in association with neighborhood racial/ethnic composition. Further research is required that compares the predictive performance of different composite indices using individual-level data. For example, three recent studies have used large individual datasets to compare the predictive performance of COI, SVI and ADI.[261-263] Aris et al. examine incident asthma and childhood BMI and find larger and more robust associations using the COI vs. SVI.[261; 262] Beyer et al. examine externalizing behaviors and find larger and more robust associations using the

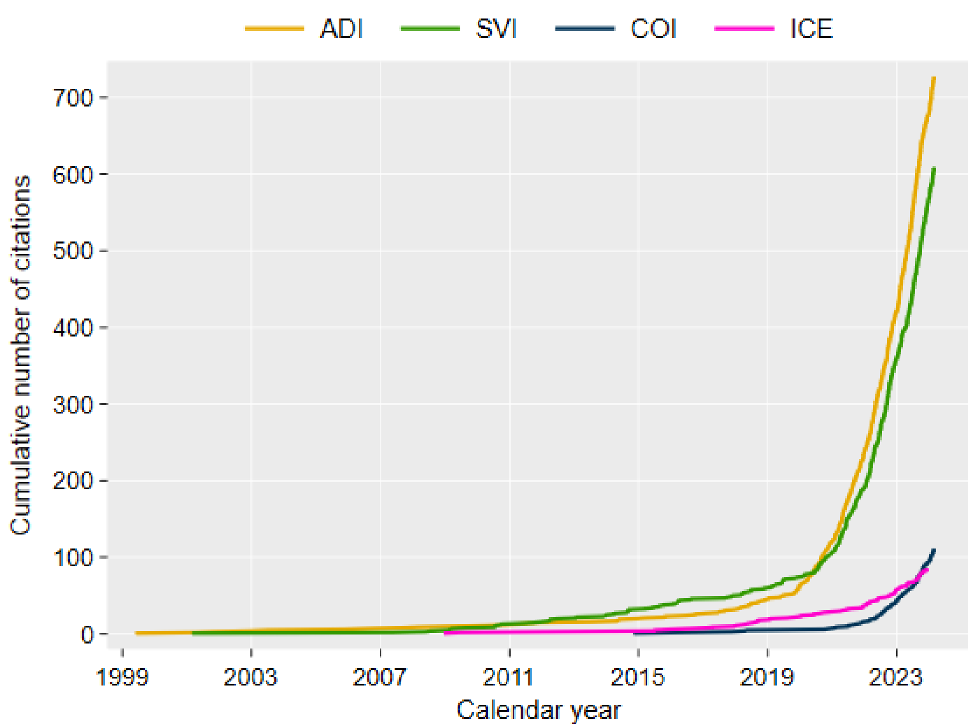
ADI vs. COI.[263] Similar studies, using a broader set of neighborhood metrics and outcomes, are needed to further validate composite neighborhood indices.

LOOKING AHEAD

Composite indices are appealing for many reasons. They turn large amounts of publicly available neighborhood-level data into comprehensive measures of neighborhood quality for researchers. They provide efficient and actionable metrics for policymakers and program administrators. Figure 6 shows the number of papers listed on PubMed citing at least one of the following indices: Area Deprivation Index, Social Vulnerability Index, Index of Concentration at the Extremes and/or Child Opportunity Index. The exponential growth in the adoption of composite indices among scholars reflects a growing recognition of the importance of residential environments for health inequity and a growing recognition that residential environments are multidimensional. While there are no statistics on the applied use of composite indices by companies, organizations and government agencies, we believe that the uptake has been no less rapid, and may be further fueled by the rapidly growing evidence base.

More research on the construction, validation and application of composite indices is required to put their use on a more solid scientific foundation. First, it is of critical importance that data providers are funded not just for the development and dissemination of a new index, but for instituting rigorous quality control mechanisms and providing detailed documentation on its construction. Second, funders, data providers and users need to develop a shared understanding of the protocols and procedures required to establish that an index is properly constructed.

FIGURE 6. CUMULATIVE NUMBER OF PAPERS ON PUBMED CITING ONE OF FOUR NEIGHBORHOOD INDICES, 1999-2024



Source: PubMed.

For example, in its guidelines on the development and use of software as medical devices (SaMD), the U.S. Food and Drug Administration presupposes that development and product deployment follow accepted quality management principles in the software development/product life cycle. The guidelines also highlight the necessity to demonstrate analytical validation, which “measures the ability of a SaMD to accurately, reliably and precisely generate the intended technical output from the input data.”[264] Similarly, Acevedo-Garcia et al. and Rehkopf and Phillips put forward conceptual frameworks and criteria that should govern the development, dissemination and application of composite neighborhood indices.[42; 265]

Third, little is known about how content and construction of composite neighborhood indices impact their validity and application. For example, Petterson investigates the distributional effects of using two versions of the Area Deprivation Index, i.e., which neighborhoods stand to benefit or lose from the application of a given metric, for example, in the distribution of government funds.[43]

Lastly, there is little research examining how the methods of index construction impact their content, predictive and equity validity, and consequently it is unclear how validity can be optimized. Existing indices could certainly be improved by refining the use of existing data or tapping into new data sources to measure thus far unmeasured constructs. Existing indices could also benefit from refining the methods used to combine component indicators into a composite index. Improvements along these dimensions should increase the effectiveness of indices in applied settings, for example, by providing better measures of eligibility for a program or policy and therefore potentially increase equity in program or policy effects. Lastly, “ground-truthing” index components and construction is an important facet of validation. Previous research has shown that community feedback can deepen our understanding of composite indices and their limitations, and provide important evidence supporting further refinements.[266; 267]

With the Child Opportunity Index 3.0, we have put forward the most comprehensive composite neighborhood index available to the public to date. In this document, we provide detailed documentation on its construction, its content and its predictive validity. COI 3.0 is available to the public for download on our website and for exploration on our interactive mapping platform, where it can be analyzed in relation to racial/ethnic residential segregation and historical government redlining. We hope that the index and the research supporting it will be useful to a wide community of scholars and practitioners across different fields and sectors, and we encourage scholars and practitioners to reach out to us with their questions and comments. We look forward to continue our interactions with an ever-growing community of COI users, and we look forward to supporting and learning from them.

APPENDIX 1. INDICATOR DESCRIPTIONS

EDUCATION DOMAIN

Early childhood education subdomain

Public pre-K enrollment

Description:	Percentage of 3- and 4-year-olds enrolled in public nursery school, preschool or kindergarten
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B14003
Source geography:	Census tract
Definition:	The number of 3- and 4-year-olds enrolled in public nursery school, preschool or kindergarten divided by the number of 3- and 4-year-olds for whom enrollment status is known, times 100.

Private pre-K enrollment

Description:	Percentage of 3- and 4-year-olds enrolled in private nursery school, preschool or kindergarten
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B14003
Source geography:	Census tract
Definition:	The number of 3- and 4-year-olds enrolled in private nursery school, preschool or kindergarten divided by the number of 3- and 4-year-olds for whom enrollment status is known, times 100.

Elementary education subdomain

Reading and math test scores

Description:	Standardized test scores in math and reading/language arts
Years:	2008/09 – 2017/18 school years
Scale:	Cohort standardized (standard deviation)
Source:	Stanford Education Data Archive, Version 4.1[268; 269]
Source geography:	School latitude/longitude
Definition:	Math and reading/language arts standardized test scores, averaged over students, grades, subjects and school years. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.
Notes:	School-level estimates are cohort-standardized and comparable across the U.S.

Reading and math test score growth

Description:	Growth in standardized test scores in math and reading/language arts
Years:	2008/09 – 2017/18 school years
Scale:	Cohort standardized (standard deviation)
Source:	Stanford Education Data Archive, Version 4.1[268; 269]
Source geography:	School latitude/longitude
Definition:	Math and reading/language arts standardized test score growth, averaged over students, grades, subjects and school years. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.
Notes:	School-level estimates are cohort-standardized and comparable across the U.S.

Poverty-adjusted reading and math test scores

Description:	Poverty-adjusted standardized test scores in math and reading/language arts
Years:	2008/09 – 2017/18 school years
Scale:	Standardized (standard deviation)
Source:	Stanford Education Data Archive, Version 4.1[268; 269]
Source geography:	School latitude/longitude
Definition:	Poverty-adjusted math and reading/language arts standardized test scores, averaged over students, grades, subjects and school years. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.
Notes:	School-level estimates are standardized and comparable across the U.S.

Secondary and post-secondary education subdomain

Advanced Placement course enrollment

Description:	Percentage of 9th to 12th graders enrolled in at least one AP course
Years:	2011/12, 2013/14, 2015/16, 2017/18 school years
Scale:	Percent
Source:	U.S. Department of Education Office for Civil Rights Data Collection (CRDC)[270]
Source geography:	School latitude/longitude
Definition:	Percentage of 9th-12th graders enrolled in at least one AP course, averaged across school years. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.

College enrollment in nearby institutions

Description:	Percentage of 18- to 24-year-olds enrolled in nearby colleges or graduate schools
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B14004
Source geography:	Census tract
Definition:	The number of individuals aged 18-24 years enrolled in nearby public or private colleges, universities or graduate schools divided by the number of adults aged 18-24 years, times 100.
Notes:	We allocated the census tract counts of individuals aged 18-24 (denominator) and individuals aged 18-24 who are enrolled in college or graduate school (numerator) to the census block level using block-level child population estimates as weights. We then aggregated the numerator and denominator within a convex hull defined around each census block centroid as described in Appendix 4.

High school graduation rate

Description:	Percentage of ninth graders graduating from high school on time
Years:	2009/10 – 2018/19 school years
Scale:	Percent
Source:	U.S. Department of Education ED*Facts* Four-Year Adjusted-Cohort Graduation Rates Data Files[271]
Source geography:	School latitude/longitude
Definition:	Adjusted four-year high school graduation rate. All students who enter ninth grade for the first time form a cohort that is subsequently adjusted for transfers and deaths. The four-year adjusted cohort graduation rate is then defined as the percentage of students of that adjusted cohort that graduate from high school with a regular diploma in four years or less. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.

Educational resources subdomain

Adult educational attainment

Description:	Percentage of adults aged 25 and over with a college degree or higher
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B15002
Source geography:	Census tract
Definition:	The number of adults aged 25 years and older who have completed a bachelor's degree or higher divided by the number of adults aged 25 years and older, times 100.

Child enrichment non-profit organizations

Description:	Density of non-profit organizations providing enrichment opportunities for children, such as after-school programs, recreational sports leagues and mentoring programs
Years:	2012 – 2021
Scale:	Number of organizations per 1000 children
Source:	National Center for Charitable Statistics IRS Business Master File[272]
Source geography:	Non-profit organization latitude/longitude
Definition:	See Appendix 3 for further details on the data source and Appendix 4 for the aggregation of non-profit location data to census blocks.

Teacher experience

Description:	Percentage of teachers in their first and second year
Years:	2011/12, 2013/14, 2015/16, 2017/18 school years
Scale:	Percent
Source:	U.S. Department of Education Office for Civil Rights Data Collection (CRDC)[270]
Source geography:	School latitude/longitude
Definition:	Percentage of teachers in their first and second year, averaged across school years. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.
Notes:	According to the CRDC documentation, "[t]he number of year(s) of teaching experience including the current year but not including any student teaching or other similar preparation experiences. Experience includes teaching in any school, subject or grade; it does not have to be in the school, subject, or grade that the teacher is presently teaching."[273]

School poverty

Description:	Percentage of students in elementary schools eligible for free or reduced-price lunches
Years:	2007/08 to 2020/21 school years
Scale:	Percent
Source:	National Center for Education Statistics Common Core of Data (CCD)[274]
Source geography:	School latitude/longitude
Definition:	The number of students in grades 1 through 5 who are eligible for free or reduced-price lunches divided by the total number of students enrolled in grades 1 through 5, times 100. See Appendix 2 for details on the processing of school-level data and Appendix 4 for aggregation of school-level data to census blocks.

HEALTH AND ENVIRONMENT DOMAIN

Pollution subdomain

Airborne microparticles

Description:	Mean estimated microparticle concentration (PM2.5; micrograms per cubic meter)
Years:	2008 – 2020
Scale:	Micrograms per cubic meter ($\mu g/m^3$)
Source:	EPA Fused Air Quality Surface Using Downscaling (FAQSD) output files[275; 276]
Source geography:	Census tracts
Definition:	Microparticle exposure is defined as the mean estimated daily 24-hour average microparticle (PM 2.5) concentration. For every census tract, we computed the annual average across all daily observations.
Notes:	The EPA used output from a Bayesian space-time downscaling fusion model called “downscaler model” (DS). The DS combines air quality data from State and Local Air Monitoring Stations (SLAMS) and the Community Multiscale Air Quality (CMAQ) model to predict daily concentrations for all U.S. census tracts (2010 definition) in the contiguous U.S. Census tract data for Alaska and Hawaii is not available.

Ozone concentration

Description:	Mean estimated ozone concentration
Years:	2008 – 2020
Scale:	Parts per billion (ppb)
Source:	EPA Fused Air Quality Surface Using Downscaling (FAQSD) output files[275; 276]
Source geography:	Census tracts
Definition:	Ozone concentration is defined as the daily maximum 8-hour average ozone concentration within 3 meters of the surface of the earth. For every census tract, we computed the annual average across all daily observations.
Notes:	The EPA used output from a Bayesian space-time downscaling fusion model termed “downscaler model” (DS). The DS combines air quality data from State and Local Air Monitoring Stations (SLAMS) and the Community Multiscale Air Quality (CMAQ) model to predict daily concentrations for all U.S. census tracts (2010 definition) in the contiguous U.S. Census tract data for Alaska and Hawaii is not available.

Industrial pollutants in air, water or soil

Description:	Index of toxic chemicals released by industrial facilities (RSEI Score)
Years:	2008 – 2020
Scale:	See definition
Source:	EPA 2020 Aggregated Grid Cell Microdata Core files provided by Abt Associates[277]
Source geography:	Raster grid cell centroid latitude/longitude
Definition:	The RSEI Score is calculated by combining information on the toxicity and estimated concentration based on emissions data of over 600 toxic chemicals.
Notes:	The Risk-Screening Environmental Indicators (RSEI) index measures the release, the fate and transport through the environment, size and location of the exposed population and toxicity level of over 600 toxic chemicals. The RSEI model uses the reported quantities of EPA Toxics Release Inventory (TRI) to estimate the risk-related impacts associated with each type of toxic air and water release or transfer by every TRI facility. The model relies on identifying where facilities are located, where people live in relation to facilities and attributes of the physical environment, such as meteorology, in the areas surrounding each facility. To locate the facilities and attribute corresponding data, the model describes the U.S. and territories on an 810m by 810m grid. We used the Aggregated Grid Cell Microdata, which contain a grid-level index

(ToxConc) defined as the estimated concentration of chemicals for a grid cell multiplied by an inhalation toxicity weight, summed over all chemicals impacting the grid cell. To map grid cells to latitude/longitude values, we used a shapefile provided by the EPA containing latitude and longitude of every grid cell centroid. We then assigned each census block the ToxConc value of the nearest grid cell.

Hazardous waste dump sites

Description:	Average number of Superfund sites within a 2-mile radius
Years:	2008 – 2021
Scale:	Count
Source:	EPA Superfund National Priorities List (NPL)[278]
Source geography:	Point data (latitude/longitude)
Definition:	We linked each census block to all Superfund sites within a 2-mile radius from the block centroid that were uncleaned in a given year, and counted the number of uncleaned sites meeting these criteria for every block and year.

Healthy environments subdomain

Fast food restaurant density

Description:	Percentage of restaurants that serve fast food
Years:	2008 – 2021
Scale:	Percent
Source:	DataAxle company database
Source geography:	Point data (latitude/longitude)
Definition:	Percentage of nearby restaurants that serve fast food.
Notes	We first limit our universe of entities to those within the restaurant industry using Standard Industrial Classification (SIC) codes self-reported by businesses. We then conduct both manual coding and entity name keyword searching to assemble a training dataset comprised of restaurants that serve fast food and those that do not. Next, we utilize natural language processing techniques, specifically employing word embeddings derived from the OpenAI text-embedding-ada-002 model. These embeddings allow us to encode semantic information from company names into numerical vectors. By combining these embeddings with SIC codes and other entity-specific information such as business size and number of employees, we apply LASSO for feature selection. We employ a random forest classifier based on the selected features to classify entities as fast food vs. non-fast food restaurants. We remove entities identified as fast food restaurants from the universe of entities and repeat the process to classify the remaining entities either as full-service restaurants or non-restaurants. To quantify the accuracy of our approach, we sampled observations at random and manually classified them. The accuracy of our model for fast food restaurants and all types of restaurants were 89% and 87%, respectively. We then construct the fast food density indicator as the percentage of restaurants (fast food and full service) that are fast food restaurants. Appendix 4 describes how we aggregate the point-level information on the location of restaurants to the block-level.

Healthy food retailer density

Description:	Percentage of retailers selling healthy food
Years:	2008 – 2021
Scale:	Percent
Source:	DataAxle company database
Source geography:	Point data (latitude/longitude)
Definition:	Percentage of nearby food retailers selling healthy food. See Appendix 4 for aggregation of point-level data to census blocks.

Notes	We first limit our universe of entities to those within the food retail industry using Standard Industrial Classification (SIC) codes self-reported by businesses. We then conduct both manual coding and entity name keyword searching to assemble a training dataset comprising businesses that are food retailers and those that are not (for example, liquor or conveniences stores). Next, we utilize natural language processing techniques, specifically employing word embeddings derived from the OpenAI text-embedding-ada-002 model. These embeddings allow us to encode semantic information from company names into numerical vectors. By combining these embeddings with SIC codes and other entity-specific information such as business size and number of employees, we apply LASSO for feature selection. We employ a random forest classifier based on the selected features to classify entities as healthy food retailers vs. unhealthy food retailers. We remove entities identified as healthy food retailers from the universe of entities and repeat the process to classify the remaining entities either as food retailers or non-food retailers. To quantify the accuracy of our approach, we randomly sampled observations and manually classified them. The accuracy of our model for healthy food retailers and all types of food retailers were 89% and 91%, respectively. Appendix 4 describes how we aggregate the point-level information on the location of restaurants to the block level.

Healthy food access

Description:	Index of fast food restaurant and healthy food retail density
Years:	2008 – 2021
Scale:	Index value (standard deviations)
Source:	DataAxle company database
Source geography:	Census block
Definition:	Index of fast food restaurant and healthy food retail density
Notes:	The index is computed as the weighted average of the fast food restaurant density and healthy food retailer density component indicators. To combine the two component indicators into the healthy food index, we used the same algorithm that was used to combine component indicators into subdomain scores. Only the healthy food access index is included in calculation of the healthy environments subdomain score.

Extreme heat exposure

Description:	Number of summer days with maximum temperatures above 90 degrees Fahrenheit
Years:	2008 – 2021
Scale:	Count
Source:	North American Land Data Assimilation System Phase 2 (NLDAS-2) Primary Forcing Data (NLDAS_FORA0125_H), NASA Goddard Earth Sciences Data and Information Services Center[279]
Source geography:	Raster grid cell centroid latitude/longitude
Definition:	Number of summer days with maximum temperatures above 90 degrees Fahrenheit.
Notes:	The number of extreme heat days, or days over 90F, was calculated using data from NASA's North American Land Data Assimilation System Phase 2 (NLDAS). These data consist of hourly temperature measurements covering the contiguous U.S. states. Data for Alaska and Hawaii is not available. Temperature measurements are interpolated to fill a 1/8 by 1/8 degree raster grid. Each census block was assigned the 5-year average number of extreme heat summer days of the grid cell nearest to the block's centroid.

NatureScore

Description:	NatureScore measures exposure to healthy natural environments using data on green space, tree canopies, parks, and air, noise and light pollution
Years:	2018, 2020
Scale:	Index units ranging from 0 to 1
Source:	NatureQuant

Source geography: Census block
Definition NatureScore is a proprietary index measuring the quantity and quality of healthy, green and natural environments. It is constructed from datasets of different environmental features, including satellite images of vegetation and land use, data on parks, tree canopy, noise levels, artificial light, air pollution, buildings, roads, and aerial and street view images. The index uses machine learning to construct a weighted average, where weights reflect the components association with health outcomes.[151; 248]
Notes NatureScore data is not publicly available. NatureScore was not available before 2018. We linearly interpolated the 2019 value from 2018 and 2020 data. We used the 2018 values for the period from 2012 to 2017 and the 2020 value for 2021.

Walkability

Description: EPA Walkability Index
Years: 2020
Scale: Index units ranging from 1 (least walkable) to 20 (most walkable)
Source: U.S. Environmental Protection Agency and U.S. General Services Administration Smart Location Database, version 3, January 2021[280; 281]
Source geography: Census block groups
Definition: The walkability index was developed by the EPA and uses 2018 Census TIGER/Line geographic definitions. It is a weighted average of four block group features that predict the likelihood of residents making walk trips: (1) street intersection density, weighted to reflect connectivity for pedestrian and bicycle travel; (2) distance from population centers to nearest transit stop in meters; (3) the mix of employment types in a block group (such as retail, office or industrial) and (4) the mix of employment types and occupied housing. A block group with a diverse set of employment types (such as office, retail and service) plus many occupied housing units will have a relatively high value. Blocks were ranked on each score and assigned a rank score from 1 to 20 based on their quantile position, where a higher score indicates a greater probability of walking. To calculate the index, the four rank scores are averaged, where intersection density and proximity to transit stops receive a weight of 1/3 and employment mix and household mix receive a weight of 1/6, respectively. Source variables were gathered for somewhat different time points that represent conditions over the period from 2018 to 2020.[280]

Safety-related resources subdomain

Community safety-related non-profits

Description: Density of non-profit organizations focused on increasing community safety (number of organizations per 1,000 children)
Years: 2012 – 2021
Scale: Number of organizations per 1,000 children
Source: National Center for Charitable Statistics IRS Business Master File[272]
Source geography: Non-profit organization latitude/longitude
Definition: See Appendix 3 for further details on the data source and Appendix 4 for the aggregation of non-profit location data to census blocks.

Vacant housing

Description: Percentage of housing units that are vacant
Years: 2012 – 2021
Scale: Percent
Source: 5-year ACS (api.census.gov), Table B25002
Source geography: Census tract

Definition:	The number of vacant housing units, excluding housing units for seasonal, recreational and occasional use, divided by the number of total housing units, times 100.
Notes:	Information on vacancy status in the ACS was obtained both through internet self-responses and personal interviews. Before 2013, it was obtained only via personal interviews for a sample of cases.[282]

Health resources subdomain

Health-related non-profits

Description:	Density of non-profit organizations providing health-related services (number of organizations per 1,000 children)
Years:	2012 – 2021
Scale:	Number of organizations per 1,000 children
Source:	National Center for Charitable Statistics IRS Business Master File[272]
Source geography:	Non-profit organization latitude/longitude
Definition:	See Appendix 3 for further details on the data source and Appendix 4 for the aggregation of non-profit location data to census blocks.

Health insurance coverage

Description:	Percentage of individuals aged 0-64 with health insurance coverage
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B27001
Source geography:	Census tract
Definition:	The number of individuals aged 0-64 with health insurance coverage divided by the number of individuals aged 0-64, times 100.

SOCIAL AND ECONOMIC DOMAIN

Employment subdomain

Employment rate

Description:	Percentage of adults aged 25-54 years who are employed
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B23001
Source geography:	Census tract
Definition:	The number of adults aged 25-54 years who are employed in the civilian labor force divided by the number of adults aged 25-54 years, times 100.

High-skill employment rate

Description:	Percentage of individuals aged 16 years or older who are employed in high-skill occupations
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table C24010
Source geography:	Census tract
Definition:	The number of individuals aged 16 years and over who are employed in management, business, financial, computer, engineering, science, education, legal, community service, health care practitioner, health technology or arts and media occupations divided by the number of individuals aged 16 years and over, times 100.

Full-time, year-round earnings

Description: Median earnings in the past 12 months for civilian employees working full-time, year-round
Years: 2012 – 2021
Scale: 2021 U.S. Dollars
Source: 5-year ACS (api.census.gov), Table B24022
Source geography: Census tract
Definition: Median earnings in the past 12 months for civilian employees working full-time, year-round. Full-time, year-round work is defined as working 35 hours or more per week for 50 to 52 weeks in the past 12 months.

Economic resources subdomain

Median household income

Description: Median household income
Years: 2012 – 2021
Scale: 2021 U.S. Dollars
Source: 5-year ACS (api.census.gov), Table B19013
Source geography: Census tract
Definition: Median income across all households.

Poverty rate

Description: Percentage of individuals living in households with income below 100% of the federal poverty threshold
Years: 2012 – 2021
Scale: Percent
Source: 5-year ACS (api.census.gov), Table B17001
Source geography: Census tract
Definition: The number of individuals of all ages living in households with incomes below 100% of the federal poverty threshold divided by the number of individuals of all ages living in households for whom poverty status could be determined, times 100.

Public assistance rate

Description: Percentage of households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assessment Program (SNAP)
Years: 2012 – 2021
Scale: Percent
Source: 5-year ACS (api.census.gov), Table B19058
Source geography: Census tract
Definition: The number of households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assessment Program (SNAP) divided by the number of households, times 100.

Concentrated socioeconomic inequity subdomain

Adults with advanced education degrees

Description: Percentage of adults aged 25 and over with master's, professional or doctoral degrees
Years: 2012 – 2021
Scale: Percent
Source: 5-year ACS (api.census.gov), Table B15002
Source geography: Census tract
Definition: The number of individuals aged 25 or older with a master's degree, professional school degree, or doctorate degree divided by the number of all individuals aged 25 or older, times 100.

Very high-income households

Description:	Percentage of households with income greater than $125,000 in the past 12 months
Years:	2010 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B15002
Source geography:	Census tract
Definition:	The number of households with income greater than $125,000 in the past 12 months divided by the number of households, times 100.

Adults without a high school degree

Description:	Percentage of individuals aged 25 and older without a high school degree
Years:	2010 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B15002
Source geography:	Census tract
Definition:	The number of individuals aged 25 or older without a high school degree, divided by the number of all individuals aged 25 or older, times 100.

Very low-income households

Description:	Percentage of households with income less than $20,000 in the past 12 months
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B15002
Source geography:	Census tract
Definition:	The number of households with income less than $20,000 in the past 12 months divided by the number of households, times 100.

Housing resources subdomain

Broadband access

Description:	Percentage of households with connections to high speed broadband internet
Years:	2017 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table S2801
Source geography:	Census tract
Definition:	The number of households with connections to high speed broadband internet (including cable, fiber optic and DSL connections) divided by the number of households, times 100.

Crowded housing

Description:	Percentage of housing units with more than one occupant per room
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), B25014
Source geography:	Census tract
Definition:	The number of households with more than one occupant per room divided by the number of households, times 100.
Notes:	The following types of rooms are included in the count: “living rooms, dining rooms, kitchens, bedrooms, finished recreation rooms, enclosed porches suitable for year-round use, and lodger's rooms. […] [P]ullman kitchens, bathrooms, open porches, balconies, halls or foyers, half-rooms,

utility rooms, unfinished attics or basements, or other unfinished space used for storage" are not counted.[282]

Homeownership rate

Description:	Percentage of occupied housing units that are owner-occupied
Years:	2012 – 2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), B25003
Source geography:	Census tract
Definition:	Number of occupied housing units that are owner-occupied divided by the number of occupied housing units, times 100.

Social resources subdomain

Mobility-enhancing friendship networks

Description:	Prevalence of high-socioeconomic status (SES) friends among low-SES individuals (economic connectedness)
Years:	2022
Scale:	Index units
Source:	Opportunity Insights[221; 283]
Source geography:	2022 ZIP codes
Definition:	Economic connectedness is defined as two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the ZIP code.
Notes	ZIP code estimates of economic connectedness were allocated to 2010 census tracts using USPS ZIP Code Crosswalk Files published by the U.S. Department of Housing and Urban Development Office of Policy Development and Research (PD&R).[284; 285] For a census tract that intersects with more than one ZIP code, the tract estimate is computed as the weighted average of the intersecting ZIP codes, where weights are the proportion of the census tract's addresses with a given ZIP code.

Single-headed households

Description:	Percentage of single-parent family households
Years:	2012-2021
Scale:	Percent
Source:	5-year ACS (api.census.gov), Table B17010
Source geography:	Census tract
Definition:	The number of single-parent (male householder with no wife present or female householder with no husband present) family households with children aged 0-17 years related to the householder divided by the number of family households with children aged 0-17 related to the householder, times 100.
Notes:	For ACS tabulations, a family "consists of a householder and one or more other people living in the same household who are related to the householder by birth, marriage, or adoption. All people in a household who are related to the householder are regarded as members of his or her family. A family household may contain people not related to the householder, but those people are not included as part of the householder's family in tabulations."[282]

Non-profit organizations

Description:	Density of non-profit organizations (number of organizations per 1,000 children)
Years:	2012 – 2021
Scale:	Number of organizations per 1,000 children

Source:	National Center for Charitable Statistics IRS Business Master File[272]
Source geography:	Non-profit organization latitude/longitude
Definition:	See Appendix 3 for further details on the data source and Appendix 4 for the aggregation of non-profit location data to census blocks.

APPENDIX 2. SCHOOL DATA

OVERVIEW

COI 3.0 includes seven indicators calculated from school-level data listed in Table A2.1. In this section, we describe the protocol for processing school-level data up until the point of aggregation to the block level, which is described in Appendix 4.

We lack school-level data for several indicators after 2017/18. Standardized testing and reporting of school-level data were disrupted by the COVID pandemic starting with the 2019/20 school year, and post-pandemic data from several of the sources utilized here is not yet available nationwide. We expect to retroactively update these indicators with future updates of the Child Opportunity Index.

TABLE A2.1. COI 3.0 COMPONENT INDICATORS SOURCED FROM SCHOOL DATA

Indicator	Description (Source)	Universe	School Years
Reading and math test scores	Average school-wide standardized test scores in math and reading/language arts (SEDA)	Public elementary schools	Single estimate covering 2007/08 through 2017/18
Reading and math test score growth	Average school-wide standardized test score growth rates in math and reading/language arts (SEDA)	Public elementary schools	Single estimate covering 2007/08 through 2017/18
Poverty-adjusted reading and math test scores	Poverty-adjusted average school-wide standardized test scores in math and reading/language arts (SEDA)	Public elementary schools	Single estimate covering 2007/08 through 2017/18
Advanced Placement (AP) course enrollment	Percentage of 9^{th}-12^{th} graders enrolled in at least one AP course (CRDC)	Public high schools	2009/10 to 2018/19 (biannual)
High school graduation rate	Percentage ninth graders graduating from high school on time (EdFacts)	Public high schools	2010/11 to 2018/19 (annual)
Teacher experience	Percentage teachers in their first and second year, reversed (CRDC)	Public elementary schools	2009/10 to 2018/19 (biannual)
School poverty	Percentage students in elementary schools eligible for free or reduced-price lunches, reversed (CCD)	Public high schools	2007/08 to 2020/21 (annual)

We utilized school data from the following sources:

- National Center for Education Statistics Common Core of Data (CCD): The CCD is an annual database of all public elementary and secondary schools published by the U.S. Department of Education's National Center for Education Statistics. From the CCD, we drew a comprehensive list of public schools as well as information on total and grade-specific enrollment counts, total enrollment counts by race/ethnicity, number of students eligible for free and reduced-price lunches (FRPL), school location (longitude and latitude) and other school features, including whether it is subject to the Community Eligibility Provision (see below).[274; 286]
- Stanford Education Data Archive (SEDA) Version 4.1: The SEDA data files contain school-level data on math and reading proficiency that is comparable across states. It is based on standardized tests administered in grades three to eight across all U.S. public schools during the school-years 2007/08 through 2017/18.[268; 269]
- U.S. Department of Education EDFacts State Assessments in Mathematics and Reading/Language Arts Data Files: We used EDFacts assessment files to retrieve information on the number of students that are economically disadvantaged for all public schools enrolling students in grades three to eight. Annual data was available from the 2009/10 to 2018/19 school years.[287]
- U.S. Department of Education EDFacts Four-Year Adjusted-Cohort Graduation Rates Data Files: We used publicly available EDFacts assessment files to retrieve information on the number of students that are economically disadvantaged for all public high schools, as well as high school graduation rates for all public high schools. Annual data was available from the 2009/10 to 2018/19 school years.[271]
- U.S. Department of Education Office for Civil Rights Data Collection (CRDC): The CRDC is a biennial survey required by the U.S. Department of Education's Office for Civil Rights (OCR). It collects data from all public local educational agencies (LEA) and schools.[270] We retrieved data on student's enrollment in at least one AP course and teachers in their first or second year of teaching. We used data from the following school-years: 2011/12, 2013/14, 2015/16, 2017/18 school years. Data from 2009/10 was omitted because of quality concerns.

UNIVERSE OF SCHOOLS

To define our universe of schools, we begin with all schools listed in the NCES Common Core of Data (CCD). We then remove schools if they meet any of the following criteria:

- Schools located outside the 50 states and Washington, D.C.
- Schools with missing data on latitude or longitude, or schools matching any of the following coordinates: zero degrees latitude and longitude, degrees of latitude greater than zero or degrees of longitude greater than -50
- Schools for which kindergarten is the highest grade
- Schools for which adult education is the highest grade
- Schools with zero or missing total enrollment

- Schools that are virtual only. We include all schools that are non-virtual or have missing virtual status, a category inflated by the COVID pandemic for the 2019/20 and 2020/21 school years
- We excluded schools designated as "special education" schools. Furthermore, we excluded schools with names designating that they are serving either children with special needs or are affiliated with hospitals, children's homes, prisons or jails or schools for rehabilitating juvenile offenders. Specifically, we excluded schools if their name matched a list of 37 terms such as, "for the blind," "jail," "prison," "psychiatric," "children's home" or "transition services."

We define two universes of schools: public elementary schools that have at least one student enrolled in grades one to five and public high schools that have at least one student enrolled in grades nine to 12. We omit schools solely enrolling students in grades six to eight from our universe. Each of our school indicators (see Table 1) is computed for one of those two universes. We use all schools enrolling at least one student in grades one to five to compute school poverty, academic proficiency and teacher experience, and all schools enrolling at least one student in grades nine to 12 to compute high school graduation and AP enrollment. Across the 2007/8 to 2020/21 school years, the resulting dataset includes between 92,000 and 93,000 schools per school year.

The following sections describes the processing of SEDA, CCD, CRDC, and EDFacts school-level datasets used to generate the school level indicators listed in Table 1. Appendix 4 describes how census blocks were linked to nearby schools and how school data described in this section was aggregated to the block level to construct annual, block-level indicators based on school data.

SCHOOL POVERTY

School poverty is defined as the percentage of students in grades one to five eligible for free and reduced-price lunches (FRPL) as reported in the NCES CCD. With the introduction of the Community Eligibility Provision (CEP) in a growing number of states beginning with the 2011/12 school year, all students within a district gain access to free meals, regardless of the student composition of a specific school.[286] Thereby, the introduction of CEP systematically biases the spatial measurement of school poverty using FRPL-eligible students counts within and across school districts. Under the CEP, a school-level estimate of 100% FRPL eligibility might reflect district-wide policies and likely overstates the amount of economic need in at least some schools in the district.[269] Because we equate FRPL status with poverty, the CEP likely inflates FRPL eligibility for some schools in poorer districts.

As proposed by Fahle et al., we correct for this bias by setting FRPL status to missing for schools subject to the CEP and then impute it.[269] Before CEP's introduction, there already were a limited number of schools with missing FRPL status—7% in the 2008/09 school year. However, after setting FRPL to missing for schools reporting to be eligible for CEP, the number of schools with missing FRPL status increases steadily over time and reaches 23% of schools in our universe for the school poverty indicator in the 2020/21 school year. After finding evidence of FRPL misreporting during the introduction of CEP in California, we set FRPL counts to missing for all Californian schools in the 2011/12 school year.

The increasing rate of missingness biases estimates of school poverty, because missingness becomes a function of economic need. For this reason, we impute missing FRPL status using, first, Multiple Imputation by Chained

Equations (MICE) to impute missing feature data, and then Random Forests to predict FRPL status.[288; 289] We use school-level racial/ethnic composition— the proportion of students who are economically disadvantaged (EDFacts)—and school latitude and longitude as main features predicting FRPL status. We take advantage of the panel structure of the data by using lead and lagged feature data and lead and lagged FRPL status. We first create MICE imputed datasets for a given school year with contemporaneous, lead and lag features, and then run a Random Forest year by year to predict the proportion of students eligible for FRPL. The average cross-validated R^2 in the training data is 0.96 and shows no trend over time (min = 0.88, max = 0.98). We observe the lowest R^2 (0.88) in 2012, the year in which the CEP was introduced. We then use the Random Forest model results to predict the proportion of students eligible for FRPL whenever missing. The resulting dataset contains fully imputed racial/ethnic composition, percent students economically disadvantaged and counts of students eligible for FRPL status for all schools in our universe.

We use this data to compute the school poverty COI component indicator at the census block-level using the methods described in Appendix. We also use this dataset to process and impute school-level data on teacher experience and academic proficiency. Lastly, we created a similar dataset for our universe of public high schools that we use for processing and imputing school data on AP enrollment and high school graduation.

READING AND MATH TEST SCORES

We obtained school-level data on reading and math proficiency from the Stanford Education Data Archive (SEDA), Version 4.1. The SEDA files include school-level estimates of students' reading and math proficiency based on internal Department of Education data. SEDA reading and math proficiency estimates are based on school-level data collected through nationally mandated standardized tests administered across grades 3-8 over the school years 2008/9 through 2017/18. Test score data is processed so that the resulting proficiency metrics are comparable across U.S. states that administer different tests and apply different reporting standards. We obtain two metrics of learning/proficiency from the SEDA data, average school-level reading/math proficiency and average school-level reading/math test score growth. We also construct a third metric, poverty adjusted reading/math proficiency, described further below. School-level proficiency and test score growth estimates combine test scores from all school years, grades and groups of students and combine math and reading test results. Therefore, for each school, we have a single estimate of reading and math proficiency (and reading/math proficiency growth) that combines data collected across the 2008/09 through 2017/18 school years.

We obtain variables from the SEDA data files that measure reading and math proficiency (pooled, cohort-standardized Empirical Bayes estimates) at the school- and district-level, as well as reading and math proficiency growth rates at the school- and district-level. We merge this school- and district-level proficiency data onto the fully imputed data file described in the preceding section containing annual school-level data, which captures all schools in our universe of elementary schools observed from the 2008/09 school year onwards.

In the resulting merged data file, 4% of observations had missing data on average reading/math proficiency, and 21% of schools lacked data on reading/math test score growth. We examined whether the missing pattern for each variable shows a gradient with respect to the proportion of students eligible for FRPL. The presence of a gradient would likely result in biased block-level proficiency estimates, for example, if schools in poorer communities were more likely to have missing proficiency data. We detected no gradient for the missing rate on average reading/math proficiency, and therefore dropped all observations with missing data on average reading/math proficiency. We found an inverse FRPL gradient for missingness on reading/math test score growth. We observed the highest percentage of missing data (24%) among the 20% of schools with the lowest proportion of students eligible for FRPL, and we observe the lowest percentage of missing data (17%) among the 20% of schools with the highest proportion of students eligible for FRPL. We therefore imputed reading/math test score growth first using MICE to impute features with missing data (four percent missing on reading/math test scores) and then Random Forest to impute missing test-score growth data. The cross-validated R^2 in the training data was 0.92. After imputing reading/math test score growth, we took total enrollment weighted averages of the imputed observations across school years, and also computed total enrollment weighted proportions of students eligible for FRPL. We then combined time-invariant estimates of reading/math proficiency, partially imputed estimates of reading/math test score growth and the proportion of students eligible for FRPL in a dataset for further processing.

The use of unadjusted measures of average reading/math proficiency as indicators of school quality is often criticized because of the very strong relationship between student composition and proficiency scores. Because test scores are strongly associated with features of students' home environments, they might primarily reflect educational resources available within families and say relatively little about the quality of instruction and overall effectiveness of schools.[56; 71] For example, the correlation between proportion students eligible for FRPL and unadjusted average student proficiency is very strong (Pearson's rho = -0.79).

To address the concern that (unadjusted) reading/math proficiency primarily reflects home and not school environments, the elementary education subdomain in COI 3.0 combines reading/math proficiency with test score metrics that are weakly correlated or uncorrelated with school poverty: test score growth, which is only weakly correlated with the proportion of students eligible for FRPL (Pearson's rho = -0.15), and poverty-adjusted test scores, which are uncorrelated with school poverty by construction.

To construct poverty-adjusted proficiency estimates, we follow the approach outlined by Angrist et al. and regression-adjust average school reading/math proficiency using the proportion of students eligible for FRPL.[71] We first standardize school average proficiency and percentile-transform the proportion FRPL, by first ranking schools on FRPL and then grouping them into 100 groups containing one percent of schools each. We then regress standardized proficiency non-parametrically on percentile-transformed FRPL status. The residual from this regression is our poverty-adjusted measure of average school reading/math proficiency. It is uncorrelated with the proportion of students eligible for FRPL by construction, but still strongly correlated with unadjusted average school reading/math proficiency (Pearson's rho = 0.61).

The resulting data-file contains three time-invariant proficiency estimates per school. We standardize each estimate across schools using the z-score transformation.

TEACHER EXPERIENCE

We obtain school-level data on the proportion of teachers in their first or second year of teaching from the NCES CRDC data. CRDC data is available biennially for the 2011/12, 2013/14, 2015/16, and 2017/18 school years. We merge this data with the fully imputed file, capturing our universe of elementary schools described in the "School Poverty" section, and also merge in time-invariant reading and math proficiency estimates described in the preceding section. We subset the resulting data file to those school years with available CRDC data. The annual missing rate for teacher experience varies between three and five percent. We impute the missing observations first using MICE and then Random Forest using an approach similar to the one described in the "School Poverty" section. The cross-validated R^2 in the training data was 0.65. Lastly, we take total enrollment weighted averages of the number of teachers in their first/second year and the total number of teachers across the four school-years included in the analysis. The resulting data file contained time-invariant measures of the number of teachers in their first/second year and the total number of teachers.

HIGH SCHOOL GRADUATION RATE

The high school graduation rate indicator is defined as the four-year adjusted cohort graduation rate: All students who enter ninth grade for the first time form a cohort that is subsequently adjusted for transfers and deaths. The four-year adjusted cohort graduation rate is then defined as the percentage of students of that adjusted cohort who graduate from high school with a regular diploma in four years or less. Annual, school-level data on high school graduation was available for the 2010-11 through 2018-19 school years.

For privacy protection in the publicly released EDFacts data files, graduation rates are bounded into intervals in many cases. Roughly 50% of values are reported as intervals that are five- or 10-percentage-point-wide intervals, e.g. 90-95% or 70-80%. Twenty percent of values are bounded into wider intervals, and about 30% are reported as integer percentages, e.g., 91%. To obtain comparable data across all schools, we impute integer percentages for all graduation rates reported as intervals.

We first merged the EDFacts data files with our data file containing the universe of high schools and fully imputed school features described in the Section "School Poverty (CCD)" above and subset to the school years with available high school graduation rate data. The resulting dataset has between 20,000 and 21,000 high schools each year for the 2010/11 to 2018/19 school years. Next, we generated a high school graduation rate variable for training a Random Forest. The variable contains either percentage graduation rates reported with integer precision or the midpoint of interval-reported graduation rates that are reported as a five-percentage-point-wide interval. We then predict this variable using a Random Forest, taking advantage of the panel structure of the data by using lead and lagged feature data and lead and lagged graduation rates, as well as the lower and upper bound of the interval-reported graduation rates. We first create MICE imputed datasets for a given school year with fully imputed contemporaneous, lead and lag features, and then run the Random Forest. The average cross-validated R^2 in the training data is >0.99 across all school years. Lastly, we use the Random Forest model results to predict the integer percentage graduation rate for all interval-reported graduation rates, including those reported with a five-percentage-point-wide interval.

AP ENROLLMENT

AP courses are courses sponsored by the College Board, through which students may earn college credit and advanced college placement by demonstrating mastery on accompanying standardized AP exams. We obtain school-level data on the number of students enrolled in AP courses from the NCES CRDC data, and also retain data on the number of students taking SAT or ACT tests that will be used to impute AP enrollment in some cases. CRDC data is available biennially for the 2011/12, 2013/14, 2015/16, and 2017/18 school years. We merged this data with the fully imputed file capturing our universe of high schools described in the "School Poverty (CCD)" section, and also merge in data on high school graduation rates described in the preceding section. We subset the resulting data file to those school years with available CRDC data. We define AP enrollment as the school-reported percentage of students enrolled in grades nine through 12 (NCES CCD) who are taking at least one AP course (CRDC). The AP enrollment rate was missing for 19% of observations across all years. We impute the missing observations first using MICE and then Random Forest using an approach similar to the one described in the "School Poverty" section. The cross-validated R^2 in the training data was 0.92. Lastly, we took 9th-12th grade enrollment weighted averages of the AP enrollment rate across all school years with available CRDC data. The resulting data file contained time-invariant measures of students enrolled in AP courses and the number of students enrolled in grades nine through 12.

APPENDIX 3. NON-PROFIT DATA

Non-profit organizations generate social capital and provide community-focused services.[218; 222; 223] They provide community members with opportunities for volunteering and socializing, sometimes across social groups and classes. They also provide educational, health, and safety-related services to communities and take on important social safety-net functions. Scholars have highlighted the role non-profit organizations play to address racial, economic and other inequalities.[290; 291] Non-profit organizations play a crucial role in neighborhoods by fostering institutional environments where knowledge sharing, social connections, and positive community experiences thrive. The COI 3.0 captures three pathways through which non-profit community organizations are associated with improved child neighborhood opportunity: provision of educational and health-related services and opportunities, reduction in violence and crime, and providing civic infrastructure that supports the development of positive social capital.[98; 175; 222; 292; 293]

We measure the local density of non-profit organizations by counting–at the census block level–the number of "nearby" (see below) 501c3s. 501c3 organizations represent a special sector of public-oriented organizations that are tax exempt because they exist for charitable purposes. The IRS defines charitable purposes to broadly include religious, educational, scientific, athletic, and social support organizations that engage in pro-social activities, including, for example: poverty relief, community service, educational activities, and amateur sports leagues.[294] 501c3 organizations are distinct from advocacy organizations in that they are prohibited from direct or indirect participation in any political campaigns, and are distinct from other social enterprises in that they are subject to a non-distribution constraint, meaning they must reinvest any profits into the organization rather than distributing them to any private shareholder or individual.[294]

The original data for the 501c3 non-profit component indicators for COI 3.0 comes from yearly tax-exempt filings provided by every non-profit organization to the IRS. Due to the nature of the agreement between non-profits and the IRS regarding their tax-exempt status, the IRS maintains a public database of 501c3s that is continually updated. The National Center for Charitable Statistics (NCCS) has worked with the IRS to provide at least yearly snapshots of this data in downloadable tables that date back to the 1980s. The Business Master File (BMF) contains key variables on every registered non-profit entity in existence at that time. The BMF data includes every organization registered as a non-profit for tax exemption regardless of size, unlike other sources of non-profit data that are derived from tax forms only required from organizations with an annual revenue of $50,000 or greater. Because the NCCS did not release the Business Master File in 2021, we linearly interpolated census block indicators based on the IRS data for 2021 using 2020 and 2022 data.

For every organization that receives tax exempt status at the time of observation, the BMF contains their employer identification number, national taxonomy of exempt entities (NTEE) classification label, and address. Because many national organizations with multiple chapters use the same tax filing address, we identified multiple filings at the same address through matching organization root names and removed duplicate organizations that shared a name and major NTEE code. We submitted all address information for geocoding through a commercial geocoding service (geocod.io) that takes all available address information and codes it to the nearest geographic point (in degrees latitude and longitude), which can be a rooftop, a point in front of a parcel on the street, or a ZIP code centroid.

We used organizations' NTEE codes to define three groups of non-profits:

- Child enrichment non-profits that provide service or facilitate activities that are physically, socially or educationally enriching for children, for example: museums, recreation clubs, youth centers, after-school programs, youth sports leagues and Big Brothers & Big Sisters programs.
- Safety-related non-profits that provide services or facilitate activities increasing community safety. We used the classification system developed by Sharkey et al. (2017) to identify non-profits that have been linked to reductions in violent crime, including organizations focused on crime prevention, neighborhood development and job training.[175]
- Health-related non-profits that focus on the provision of health-related services that benefit children and families, including hospitals, community clinics and community mental health centers.

There is overlap across groups, because some NTEE codes were used for more than one group. For example, Big Brothers & Big Sisters non-profits were included in both the child enrichment and safety-related non-profit groups. Using methods described in the following section (Appendix 4), we then counted up the number of non-profits belonging to each group for a polygon (see section on convex hulls in Appendix 5) defined around each census block to compute three indicators measuring the density of child enrichment, safety-related and health-related non-profits. We also computed a fourth indicator that simply measured the density of non-profits of any type.

APPENDIX 4. POINT TO BLOCK AGGREGATION

Several COI 3.0 component indicators were computed from data on schools, food retailers or non-profit organizations that had exact data on their exact geographic location measured in degrees latitude and longitude. For school data, we relied on the latitude and longitude information in the Common Core of Data (CCD). For college enrollment in nearby institutions, we allocated census tract data to blocks and used block-level population-weighted centroids. For food retailers, latitude and longitude was provided by the vendor, DataAxle (formerly Infogroup). For non-profit organizations (501c3s), we used a commercial geo-coding service to convert address data to latitude and longitude.

To derive census block-level estimates from point-level (latitude/longitude) data, we used one of two approaches: To aggregate point data on the location of non-profit organizations (501c3), restaurants and food retailers, we constructed a convex hull around each block centroid and obtained the inverse-distance weighted count of entities/children within each hull, further described below. To aggregate school-level point data, we took spatially weighted averages across “nearby” (defined below) schools. Data on industrial pollutants and temperature was available in raster format. We allocated to each block the value of the nearest raster grid cell using the distance between the grid cell centroid and the population-weighted block centroid.

CONVEX HULLS TO FOR BLOCK-LEVEL ESTIMATES DERIVED FROM POINT DATA

To measure the density of non-profit organizations, for example, around a census block, we counted the number of non-profits within a convex hull and divide it by the number of children residing in blocks located within the convex hull. A convex hull is defined as the smallest convex polygon that contains all points in a set. (The convex polygon cannot have indentations or concave portions.) We algorithmically defined a convex hull around each census as follows: We matched each block to all other blocks within a 20-mile radius. For a given block, we ranked nearby blocks by distance (block centroid to block centroid) from nearest to farthest. We then defined an initially empty set of blocks, and added blocks to the set from nearest to second-nearest and so forth until two criteria were met: The last block added to the set was at least two miles away from the focal block and the sum of children residing in the blocks in the set (including the focal block) was equal to or exceeded 8,000 children. The convex hull around this set of blocks is the smallest polygon that contains all points in the set. Lastly, we summed up the number of, e.g., non-profits and children within each hull using inverse-distance weighting.

While census tracts are drawn to contain approximately 1,000 children, we drew convex hulls to contain approximately 8,000 children. Unlike census tracts, we did not take into consideration topographical or other geographic features. Most convex hulls were approximately circular around the focal block, but the focal block can also be on the boundary of the convex hull in cases where it borders a sparsely populated area. In remote rural areas, the population criterion threshold was sometimes not met if there were not enough children residing within the 20-mile radius of a given block. In densely populated areas, the population threshold was sometimes exceeded because every hull has to contain at least one block that is two miles away from the focal block.

Underlying this approach is the assumption that the proximity of geographically accessible resources varies with population density. Residents in sparsely populated, rural areas can travel further in the same amount of time, because rural areas suffer less from traffic congestion. We assumed that families can travel at least two miles to, e.g., a non-profit offering an after-school program, and that entities further than 20 miles away do not contribute to opportunities available in a given block. The algorithm then drew smaller polygons in densely populated urban areas, often not exceeding the minimum distance of two miles between the focal block and the furthest block on the convex hull, because the population criterion was already met at this distance. It drew larger polygons in rural areas, where the population threshold was only reached at larger distances. This logic mirrors a feature of census tracts which are often smaller in urban areas and can be very large in rural areas. The population threshold of 8,000 children corresponds to the number of children residing in about eight 2010 census tracts. Our approach results in polygons that have a roughly equal number of children, but differ in size. The same is true for census tracts, which can be very small in densely populated areas and very large in rural areas.

To derive the numerator for block-level estimates, we computed inverse-distance weighted sums of the number of entities, e.g., non-profits, within the convex hull for the numerator. The inverse-distance weights applied to each entity were constructed as follows: They are constant, equal to one, from zero to two miles and then decline linearly from one to 0.2 at 20 miles distance. The further an entity is outside the two-mile radius, the less it contributes to the count of entities accessible from a given block. Lastly, the weights were rescaled so that they sum to the number of entities within the set.

To derive the denominator, we computed either the sum of children within the hull (non-profit indicators) or the number of entities in the underlying universe (food indicators). For the latter, the universes are the number of all restaurants (including, e.g., full-service restaurants) for the fast food indicator, and the number of all food retailers (including, e.g., convenience stores) for the food retail indicator. In either case, we applied the same inverse-distance weights used for the numerator.

We computed the block-level indicator as the ratio of the number of non-profits per 1,000 children (non-profit indicators), the percentage of restaurants (food retailers) that serve fast food (sell healthy food). For college enrollment in nearby institutions, we computed the inverse-distance weighted total number of individuals aged 18-24 enrolled in college or graduate school and divided it by the inverse-distance weighted number of individuals aged 18-24.

BLOCK-LEVEL ESTIMATES FROM DATA ON NEARBY SCHOOLS

Our goal was to create block-level estimates of school features of those schools that children residing in a given block are likely to attend based on proximity. We began by uniquely assigning each block and each school to either a primary and a secondary geographic school district, or to a unified geographic school district. For each block, we defined a subset of in-district schools which are nearest to the block centroid. We tried to include a sufficient number of schools to obtain robust indicator estimates without adding schools that are so distant that children are unlikely to attend them. We then applied inverse distance weighting to place a greater emphasis on the nearest schools and their characteristics. With the exception of high school graduation and AP enrollment, which are measured for high schools, school indicators are measured for elementary schools. For precise definitions of the school universes, see Appendix 2.

Data on schools, school districts and census blocks

We relied on the National Center for Education Statistics (NCES) Composite School District Boundary shapefiles, which combine primary, secondary, and unified school district boundaries from the Census Bureau's TIGER/Line shapefiles. These shapefiles are published annually, reflecting incremental changes in districts and district boundaries over time, and encode the spatial areas covered by U.S. primary, secondary and unified school districts. Data on school location (latitude and longitude) was sourced from the NCES Common Core of Data (CCD). We used annual data on school location because some schools move. Census blocks are spatially defined using their population-weighted centroid, obtained from the Census Bureau's TIGER/Line shapefiles for 2010 and 2020 census blocks, i.e., census blocks as defined for the 2010 and 2020 Decennial Censuses.

Every block was linked – year by year – to either a primary or secondary school district, or to a unified school district. This linkage was performed by school year because district boundaries can change. Similarly, every school is linked – year by year – to either a primary or secondary school district, or to a unified school district. School-district linkages can change over time because schools move or districts change boundaries. Lastly, every block was linked to all schools within a 20-mile radius within the same district in which the block was located.

The school-to-district linkage was performed separately for the two school universes we defined for the school data-based COI component indicators: public elementary schools and public high schools (see Appendix 2). Elementary schools were only linked to either primary or unified districts. A small number of elementary schools could not be linked to either primary or unified districts and were therefore assigned to a state-wide synthetic district and matched to blocks on the basis of proximity alone, i.e., without consideration of the school district to which that block is linked. Similarly, high schools were only linked to secondary or unified districts, or, in a few cases, to a state-wide synthetic district.

Defining nearby schools, inverse distance weighting and aggregation

To construct the set of in-district schools matched to each block, we first deleted all schools with missing outcome data for a given indicator. Each block was then matched to all in-district schools within a 20-mile radius as well as to schools assigned to the synthetic state-wide district if they fell within the 20-mile radius. We then sorted schools in terms of distance from the (population-weighted) block centroid from nearest to farthest. For each block, we defined an initially empty set of schools and algorithmically added schools to the set starting from the nearest school, then the second nearest school, and so forth until two criteria were jointly met: The set included at least three schools for elementary school-based indicators (two schools for high school-based indicators) enrolling at least 1,500 elementary (or high school) students. If there were fewer than the minimum number of schools required, we select however many in-district schools were available within the 20-mile radius. Similarly, if after adding the required number of schools, the student enrollment criterion was not met, we keep adding schools within the 20-mile radius until it was met, though not all districts have a sufficient number of students to fulfill the enrollment criterion. We used these criteria to ensure that block-level estimates were derived from a reasonably large underlying student population drawn from different schools in close proximity to a given block.

After defining a set of in-district schools for each block, we create a weight for each school. Weights are a step function of distance between school and block centroid and are larger for schools nearer to the block centroid. Specifically, we defined the weight for school s in block b and year y as $w_{bsy} = 1/d_{bsy}$ where d_{bsy} is the distance between school s and centroid of block b in year y whenever d_{bsy} was greater than or equal to one mile. If the distance was less than one mile, we set $w_{bsy} = 1$. We top-coded the weights of schools within a one-mile radius to a value of 1 in order to prevent schools in the immediate vicinity of the block centroid from exercising an outsize influence on the block-level statistics. Finally, we rescaled the weights so that they sum up to the number of schools within the block's school set.

When computing the percentage of students that receive free or reduced-price lunches (school poverty indicator) at the block level, the median number of schools in each block's set was 3. The median distance of those schools was 2.2 miles from the population-weighted block centroids. At the high school level, when computing high school graduation at the block level, the median number of schools in each block's school set was 2 and their median distance from the block centroid was 3.6 miles.

After defining the set of in-district schools with non-missing data for each block, and computing the school weights, we multiply those weights with the respective school-level numerator and denominator, e.g., the total number of students receiving free and reduced-price lunches (numerator) and the total number of students enrolled (denominator). We then sum the weighted numerator and weighted denominator across schools for each block's set of schools to obtain a block-level statistic.

APPENDIX 5. CROSSWALKING BETWEEN CENSUS BLOCK VINTAGES

All COI component indicators were harmonized into a single, common format: annual five-year moving-average census block data for both 2010 and 2020 census blocks covering the period from 2012 to 2021. COI component indicators that were sourced from point-level data were directly mapped to either 2010 or 2020 census blocks, but the majority of component indicators were sourced at the census tract-level, including all indicators from the American Community Survey (ACS).

The Census Bureau publishes new census tract data from the ACS every year, but changes the geographic boundaries of the underlying census tracts every ten years with the decennial census. ACS data from 2012 through 2019 is published for census tracts as defined for the 2010 Decennial Census, or “2010 census tracts” for short; 2020 and 2021 census tract-level ACS data is published for 2020 census tracts, i.e., census tracts as defined for the 2020 Decennial Census. Census tracts are defined to have a size between 1,200 and 8,000 people. They are redrawn, split or merged due to population changes to approximately realize an optimum size of 4,000 people. In 2010, the Census Bureau divided the 50 US states plus D.C. into 73,057 census tracts. In 2020, the number of census tracts was 84,414.

Since their geographic boundaries differ, data for 2010 census tracts cannot, in many cases, be compared to data for 2020 census tracts. Because the Census Bureau only releases data for either 2010 or 2020 census tracts, we need to allocate (or transfer) data available for 2010 census tracts to 2020 census tracts and vice versa in order to obtain a consistent time series for 2010 and 2020 census tracts from 2012 to 2021. To allocate data from one census tract vintage to another—for example, to allocate data collected for the year 2021, which is only available for to 2020 census tract boundaries, to 2010 census tract boundaries—we allocated data from census tracts to census blocks, crosswalked census block data from one vintage to another and then aggregated it back up to the tract level (https://www.nhgis.org/geographic-crosswalks). Census blocks are perfectly nested within census tracts.

To crosswalk data for 2010 blocks to 2020 blocks, we used census block relationship files published by the Census Bureau. These relationship files specify, for example, for each 2010 block, the 2020 blocks it intersects with and how much area it shares in common with those 2020 blocks. As a hypothetical example, say 2010 block B has been divided into two blocks in 2020, B1 and B2, where B1 is 40% and B2 is 60% of the former B’s area. The relationship file would then contain two rows for block B, one for each segment it was divided into, and provide the area size of each segment. To crosswalk data from 2010 blocks to 2020 blocks, we allocate, for example, the count of poor households in proportion to the segment area sizes. In this case, if there were 100 poor households in the 2010 block B, we would allocate 40 poor households to B1 (40% of the area) and 60 poor households to B2 (60% of the area). To crosswalk data from 2020 to 2010 blocks, we would—in this example—sum the data from 2010 blocks B1 and B2 to obtain the 2010 block B value.

The specifics of this approach differ depending on the scale of the data being allocated. For count data, e.g., the number of poor households, we first allocate data from the census tract to the census block level using internal estimates of the proportion of the child population within the census tract that resides in a given block. The construction of these estimates is described below. For all other data, i.e., currency (U.S. dollars) or concentration (air pollution metrics), we assign the same census tract value to all constituent census blocks with non-zero child population.

ANNUAL CENSUS BLOCK CHILD POPULATION ESTIMATES

To allocate census tract count data to the census block-level, we designed a protocol that relies on annual five-year moving average block-level child population count estimates. Their derivation is described in this section. As a hypothetical example, say for a given census tract, the ACS estimate of the number of poor households is 400. If this census tract is comprised of 40 blocks, we could allocate the same number of poor households to each of the blocks, i.e., 10 poor households per each of the 40 blocks. This approach would allocate households to blocks regardless of whether the blocks are (or could be) residential environments for children, i.e., industrial areas, parks, golf courses or areas that are very remote and inaccessible by common modes of transportation. Instead of equally distributing households across blocks within a tract, we allocated households in proportion to the census block-level child population. If the block-level child population estimate was zero, e.g., because it was located in an industrial area, we allocated zero poor households to that block. Specifically, to allocate the number of poor households from census tract to census block-level, we computed the proportion of children in a census tract who reside in a given block and multiplied this proportion with the count of poor households.

This procedure requires time-varying block-level estimates of the number of children aged 0-17 as one input. Block-level population data is collected every ten years by the Census Bureau through the Decennial Census. The Census Bureau does not publish block-level population data in intercensal years. We therefore estimated annual five-year average census block child population counts for the period from 2008 to 2021 by combining data from the 2000, 2010 and 2020 Decennial Censuses, the Census Bureau's Intercensal Population Estimates Program (PEP) and the American Community Survey (ACS). We derived these estimates for both 2010 and 2020 census blocks, but focus on the derivation of the 2010 block estimates here.

In brief, we crosswalked child population counts from 2000 and 2020 census blocks to match the 2010 census block geographies, computed the proportion of children within a county that resides in a given 2010 block, and linearly interpolated this proportion between 2000 and 2010, and between 2010 and 2020. We allocated county-level annual population counts using the linearly interpolated proportions as weights. We then computed five-year moving averages of these annual block-level population count estimates and corrected them using census tract-level population estimates from the American Community Survey from the same five-year period.

The Decennial Censuses are the only publicly available source of block-level child population data. Because census blocks are redrawn with every census, we first crosswalk child population counts collected in the 2000 (2020) census for 2000 (2020) census block geographies to 2010 census block geographies using block relationship files published by the Census Bureau (see preceding section). The resulting dataset had the number of children aged 0-17 for all 2010 census blocks for the years 2000, 2010 and 2020. Next, we linearly interpolate the number of children for all years between 2000 and 2010 and between 2010 and 2020. Because we require data up until 2021, we extrapolate the 2020 block-level population counts to 2021. We then subset the resulting dataset to the years 2008 to 2021.

We used the Census Bureau's Population Estimates Program for spatially granular, annual child population counts. PEP data are estimated by combining data from the Decennial Census, the ACS and data on live births, deaths and migration, and are the most accurate source of age-specific, spatially granular population counts. PEP data is not available at the census tract level. We therefore used county-level data. Using the linearly interpolated census block-level population count data, we computed, for each county, the proportion of the county child population that resided in a given block in each year. We then multiplied this annual proportion with annual county level PEP child population counts to obtain an annual estimate of the block-level child population.

We corrected this block-level child population estimate using ACS data. First, we computed five-year moving averages of the estimated annual block population counts – matching the temporal scale of the ACS. We then aggregate the five-year block-level counts to the census tract level. If the block-based census tract total thus derived exceeded (or fell below) the upper (lower) bound of the 90% margin of error of the ACS child population estimate, we recoded the block-based total to the upper (lower) bound. Whenever we recoded block-based census tract population counts because they fell outside the ACS estimate's lower or upper bound, we reallocated the corrected total to the block level using the block proportion of the population in a given tract as weights.

We thus obtained annual five-year average child population counts at the census block level that incorporates information from three sources: The Census Bureau's intercensal population estimates (PEP), the Decennial Census and the American Community Survey. The block population counts are used to allocate census tract count data to the block level and aggregate block-level composite subdomain, domain and overall z-scores to higher geographic summary levels, including census tracts.

APPENDIX 6. CHANGING 2010 CENSUS TRACT DEFINITIONS

While census tracts are highly stable between Decennial Censuses, a few 2010 census tracts did undergo some changes between the 2010 and 2020 Censuses. This issue only affects component indicators sourced from the ACS between 2012 and 2019. In some cases, both the geographic boundaries and the geographic identifier changed, but in the majority of cases, boundaries stayed intact and only their unique geographic identifier changed. We resolved these issues as follows: We deleted 2010 census tract 36085008900, which was entirely comprised of water and merged with another tract in 2011. A few tracts were assigned a new geographic identifier (FIPS code), the 11-digit variable uniquely identifying each census tract. The change in GEOIDs was almost always due to a renaming of the tract following a renaming of the county it is located in, while leaving the boundaries unchanged. However, in two cases, the geographic boundaries changed, too, though this change is likely to be consequential for only one tract: The boundaries of Los Angeles County census tract number 06037930401 (2010 GEOID) were redrawn, and its comparability over time is therefore limited. Table A6.1 lists all census tracts with changed geographic identifiers and the reason for the change. The column “New GEOID” lists the geographic identifier assigned in the year a change occurred, and “2010 GEOID” lists the 2010 GEOIDs that the new GEOIDs were crosswalked to in order to create an uninterrupted time series for each tract over the entire period.

TABLE A6.1. CHANGES IN 2010 CENSUS TRACT DEFINITIONS AND IDENTIFIERS

Year change occurred	2010 GEOID	New GEOID	Name Change or Reason for Change	Explanation
2011	36053940101	36053030101	9401.01 is now 0301.01	Census tracts renumbered in Madison County, NY
2011	36053940102	36053030102	9401.02 is now 0301.02	
2011	36053940103	36053030103	9401.03 is now 0301.03	
2011	36053940200	36053030200	9402.00 is now 0302.00	
2011	36053940300	36053030300	9403.00 is now 0303.00	
2011	36053940401	36053030401	9404.01 is now 0304.01	
2011	36053940700	36053030402	9407.00 is now 0304.02	
2011	36053940403	36053030403	9404.03 is now 0304.03	
2011	36053940600	36053030600	9406.00 is now 0306.00	
2011	36065940100	36065024700	9401.00 is now 0247.00	Census tracts renumbered in Oneida County, NY.
2011	36065940000	36065024800	9400.00 is now 0248.00	
2011	36065940200	36065024900	Geographic boundary changed	A small portion of 2010 tract 0230.00 was reallocated to 2010 tract 9402.00. The newly formed tract is labeled 0249.00. Because the reallocated area was small, we assume that 0230.00 is comparable and that 9402.00 (2010) and 0249.00 (2011) are comparable over time.
2012	04019002701	04019002704	27.01 is now 27.04	Census tracts renumbered in Pima County, AZ.
2012	04019002903	04019002906	29.03 is now 29.06	
2012	04019410501	04019004118	4105.01 is now 41.18	
2012	04019410502	04019004121	4105.02 is now 41.21	
2012	04019410503	04019004125	4105.03 is now 41.25	
2012	04019470400	04019005200	4704.00 is now 52.00	
2012	04019470500	04019005300	4705.00 is now 53.00	
2012	06037930401	06037137000	Geographic boundary changed	9304.01 (2010) has been combined with part 8002.04 (2010) to form 1370.00 (2012). 9304.01 (2010) and 1370.00 (2012) are not strictly comparable. 8002.04 is also not strictly comparable, because part of its area has been reallocated. Los Angeles County, CA.
2014	51515050100	51019050100	Bedford City merged into Bedford County	Bedford City, VA, changed its legal status and was absorbed into Bedford County, VA.
2015	02270000100	02158000100	County code changed	Wade Hampton Census Area, AK, was renamed as Kusilvak Census Area.
2015	46113940500	46102940500	County code changed	Shannon County, SD, was renamed as Oglala Lakota County and the county code changed to 102 from 113
2015	46113940800	46102940800	County code changed	
2015	46113940900	46102940900	County code changed	

REFERENCES

1. Acevedo-Garcia, D., McArdle, N., Hardy, E. F., Crisan, U. I., Romano, B., Norris, D., . . . Reece, J. (2014). The child opportunity index: Improving collaboration between community development and public health. *Health Affairs, 33*(11), 1948-1957. doi:10.1377/hlthaff.2014.0679.
2. Acevedo-Garcia, D., Noelke, C., McArdle, N., Sofer, N., Hardy, E. F., Weiner, M., . . . Reece, J. (2020). Racial and ethnic inequities in children's neighborhoods: Evidence from the new Child Opportunity Index 2.0. *Health Affairs, 39*(10), 1693-1701. doi:10.1377/hlthaff.2020.00735.
3. Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics, 133*(3). doi:10.1093/qje/qjy007.
4. Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review, 106*(4), 855-902. doi:10.1257/aer.20150572.
5. Leventhal, T., & Dupéré, V. (2019). Neighborhood effects on children's development in experimental and nonexperimental research. *Annual Review of Developmental Psychology, 1*(1). doi:10.1146/annurev-devpsych-121318-085221.
6. Arcaya, M. C., Tucker-Seeley, R. D., Kim, R., Schnake-Mahl, A., So, M., & Subramanian, S. V. (2016). Research on neighborhood effects on health in the United States: A systematic review of study characteristics. *Social Science & Medicine, 168*, 16-29. doi:10.1016/j.socscimed.2016.08.047.
7. Sharkey, P., & Faber, J. W. (2014). Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology, 40*(1). doi:10.1146/annurev-soc-071913-043350.
8. Slopen, N., Cosgrove, C., Acevedo-Garcia, D., Hatzenbuehler, M. L., Shonkoff, J. P., & Noelke, C. (2023). Neighborhood opportunity and mortality among children and adults in their households. *Pediatrics, 151*(4). doi:10.1542/peds.2022-058316.
9. Sharkey, P., & Elwert, F. (2011). The legacy of disadvantage: multigenerational neighborhood effects on cognitive ability. *American Journal of Sociology, 116*(6), 1934-1981. doi:10.1086/660009.
10. Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2003). Assessing “neighborhood effects”: Social processes and new directions in research. *Annual Review of Sociology, 28*(1). doi:10.1146/annurev.soc.28.110601.141114.
11. Chyn, E., & Katz, L. F. (2021). Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives, 35*(4), 197-222. doi:10.1257/jep.35.4.197.
12. Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review, 108*(10), 3028-3056. doi:10.1257/aer.20161352.
13. Galster, G., & Sharkey, P. (2017). Spatial foundations of inequality: A conceptual model and empirical overview. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 3*(2). doi:10.7758/RSF.2017.3.2.01.
14. Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328). doi:10.1126/science.277.5328.918.
15. Wilson, W. J. (1996). *When work disappears: The world of the new urban poor* (1st ed.). New York: Knopf: Distributed by Random House, Inc.
16. Massey, D. S., & Denton, N. A. (1993). *American Apartheid: Segregation and the making of the underclass*. Cambridge, Mass.: Harvard University Press.
17. Acevedo-Garcia, D., Noelke, C., & McArdle, N. (2020). *The geography of child opportunity: Why neighborhoods matter for equity*. Retrieved from website: https://www.diversitydatakids.org/research-library/research-report/geography-child-opportunity-why-neighborhoods-matter-equity.
18. Impact Stories. Retrieved from https://www.diversitydatakids.org/impact-stories.
19. Bibliography of peer-reviewed articles and reports that use the Child Opportunity Index. Retrieved from https://www.diversitydatakids.org/research-library/impact-story/bibliography-peer-reviewed-articles-and-reports-use-child-opportunity.
20. Owens, A. (2016). Inequality in Children’s Contexts. *American Sociological Review, 81*(3), 549-574. doi:10.1177/0003122416642430.
21. Owens, A., Reardon, S., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal, 53*(4), 1159-1197. doi:10.3102/0002831216652722.
22. Bischoff, K. (2008). School district fragmentation and racial residential segregation. *Urban Affairs Review, 44*(2). doi:10.1177/1078087408320651.
23. Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(5), 34-57. doi:10.7758/rsf.2016.2.5.03.

24. Colmer, J., Hardman, I., Shimshack, J., & Voorheis, J. (2020). Disparities in PM2.5 air pollution in the United States. *Science, 369*(6503). doi:10.1126/science.aaz9353.
25. Freemark, Y., Steil, J., & Thelen, K. (2020). Varieties of urbanism: A comparative view of inequality and the dual dimensions of metropolitan fragmentation. *Politics & Society, 48*(2). doi:10.1177/0032329220908966.
26. Faber, J. W. (2020). We built this: Consequences of new deal era intervention in America's racial geography. *American Sociological Review, 85*(5). doi:10.1177/0003122420948464.
27. Noelke, C., Outrich, M., Baek, M., Reece, J., Osypuk, T. L., McArdle, N., . . . Acevedo-Garcia, D. (2022). Connecting past to present: Examining different approaches to linking historical redlining to present day health inequities. *PLOS ONE, 17*(5). doi:10.1371/journal.pone.0267606.
28. Aaronson, D., Faber, J. W., Hartley, D., Mazumder, B., & Sharkey, P. (2021). The long-run effects of the 1930s HOLC "redlining" maps on place-based measures of economic opportunity and socioeconomic success. *Regional Science and Urban Economics, 86*. doi:10.1016/j.regsciurbeco.2020.103622.
29. Aaronson, D., Hartley, D., & Mazumder, B. (2021). The effects of the 1930s HOLC "redlining" maps. *American Economic Journal: Economic Policy, 13*(4). doi:10.1257/pol.20190414.
30. Nardone, A., Rudolph, K. E., Morello-Frosch, R., & Casey, J. A. (2021). Redlines and greenspace: The relationship between historical redlining and 2010 greenspace across the United States. *Environmental Health Perspectives, 129*(1). doi:10.1289/EHP7495.
31. Trounstine, J. (2020). The geography of inequality: How land use regulation produces segregation. *American Political Science Review, 114*(2), 443-455. doi:10.1017/S0003055419000844.
32. Trounstine, J. (2018). *Segregation by design: Local politics and inequality in American cities*. Cambridge: Cambridge University Press.
33. Aaronson, D., Hartley, D., Mazumder, B., & Stinson, M. (2023). The long-run effects of the 1930s redlining maps on children. *Journal of Economic Literature, 61*(3), 846-862. doi:10.1257/jel.20221702.
34. Sharkey, P. (2008). The intergenerational transmission of context. *American Journal of Sociology, 113*(4), 931-969. doi:10.1086/522804.
35. Sharkey, P. (2012). Residential Mobility and the Reproduction of Unequal Neighborhoods. *Cityscape, 14*(3), 9-31.
36. Sampson, R. J. (2016). Individual and community economic mobility in the great recession era: The spatial foundations of persistent inequality. In F. R. B. o. S. Louis (Ed.), *Economic mobility: Research and ideas on strengthening families, communities and the economy*.
37. Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). *The Opportunity Atlas: Mapping the childhood roots of social mobility*. Retrieved from website: https://www.nber.org/papers/w25147.
38. Shanahan, K. H., Subramanian, S. V., Burdick, K. J., Monuteaux, M. C., Lee, L. K., & Fleegler, E. W. (2022). Association of neighborhood conditions and resources for children with life expectancy at birth in the US. *JAMA Network Open, 5*(10), e2235912. doi:10.1001/jamanetworkopen.2022.35912.
39. Area Deprivation Index. Retrieved from https://www.neighborhoodatlas.medicine.wisc.edu/.
40. CDC/ATSDR Social Vulnerability Index (CDC/ATSDR SVI). Retrieved from https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.
41. Hannan, E. L., Wu, Y., Cozzens, K., & Anderson, B. (2023). The Neighborhood Atlas Area Deprivation Index for measuring socioeconomic status: An overemphasis on home value. *Health Affairs, 42*(5). doi:10.1377/hlthaff.2022.01406.
42. Rehkopf, D. H., & Phillips, R. L. (2023). The Neighborhood Atlas Area Deprivation Index and recommendations for area-based deprivation measures. *Health Affairs, 42*(5). doi:10.1377/hlthaff.2023.00282.
43. Petterson, S. (2023). Deciphering the Neighborhood Atlas Area Deprivation Index: The consequences of not standardizing. *Health Affairs Scholar, 1*(5). doi:10.1093/haschl/qxad063.
44. Singh, G. K. (2003). Area deprivation and widening inequalities in US mortality, 1969-1998. *American Journal of Public Health, 93*(7), 1137-1143. doi:10.2105/ajph.93.7.1137.
45. Kind, A. J. H., Jencks, S., Brock, J., Yu, M., Bartels, C., Ehlenbach, W., . . . Smith, M. (2014). Neighborhood socioeconomic disadvantage and 30-day rehospitalization. *Annals of Internal Medicine, 161*(11), 765. doi:10.7326/m13-2946.
46. Duncan, G., Kalil, A., Mogstad, M., & Rege, M. (2022). *Investing in early childhood development in preschool and at home*. Retrieved from website: https://www.nber.org/papers/w29985.
47. Bailey, M. J., Sun, S., & Timpe, B. (2021). Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency. *American Economic Review, 111*(12). doi:10.1257/aer.20181801.
48. Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics, 94*(1-2), 114-128. doi:10.1016/j.jpubeco.2009.11.001.
49. Blau, D. M. (2021). The effects of universal preschool on child and adult outcomes: A review of recent evidence from Europe with implications for the United States. *Early Childhood Research Quarterly, 55*, 52-63. doi:10.1016/j.ecresq.2020.10.009.

50. Cascio, E. U. (2023). Does universal preschool hit the target? *Journal of Human Resources, 58*(1). doi:10.3368/jhr.58.3.0220-10728R1.
51. Cascio, E. U., & Schanzenbach, D. W. (2013). *The impacts of expanding access to high-quality preschool education*. Retrieved from website: https://www.nber.org/papers/w19735.
52. Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2022). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics, 138*(1), 363–411. doi:10.1093/qje/qjac036.
53. Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. *Journal of Human Resources, 40*(3), 533-558.
54. Duer, J. K., & Jenkins, J. (2022). Paying for preschool: Who blends funding in early childhood education? *Educational Policy, 37*(7). doi:10.1177/08959048221103804.
55. Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences, 113*(47). doi:10.1073/pnas.1601135113.
56. Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 5*(2). doi:10.7758/RSF.2019.5.2.03.
57. Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1-40. doi:10.1086/261805.
58. Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics, 131*(1), 157-218. doi:10.1093/qje/qjv036.
59. Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics, 126*(2). doi:10.1093/qje/qjr017.
60. Abdulkadiroğlu, A., Angrist, J. D., Hull, P. D., & Pathak, P. A. (2016). Charters without lotteries: Testing takeovers in New Orleans and Boston. *American Economic Review, 106*(7). doi:10.1257/aer.20150479.
61. Angrist, J. D., Pathak, P. A., & Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics, 5*(4). doi:10.1257/app.5.4.1.
62. Dobbie, W., & Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy, 123*(5). doi:10.1086/682718.
63. Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics, 126*(4), 2063–2115. doi:10.1093/qje/qjr036.
64. Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics, 123*(4). doi:10.1162/qjec.2008.123.4.1373.
65. Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics, 126*(4). doi:10.1093/qje/qjr041.
66. Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review, 104*(9), 2633-2679. doi:10.1257/aer.104.9.2633.
67. Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics, 92*(3), 562-576.
68. Bifulco, R., Fletcher, J. M., & Ross, S. L. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy, 3*(1). doi:10.1257/pol.3.1.25.
69. Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review, 108*(11). doi:10.1257/aer.20160763.
70. Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics, 2*(1). doi:10.1257/app.2.1.211.
71. Angrist, J., Hull, P., Pathak, P. A., & Walters, C. R. (2024). Race and the mismeasure of school quality. *American Economic Review: Insights, 6*(1), 20-37. doi:10.1257/aeri.20220292.
72. Autor, D. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science, 344*(6186), 843-851. doi:10.1126/science.1251868.
73. Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives, 25*(1). doi:10.1257/jep.25.1.159.
74. Crowder, K., & South, S. J. (2003). Neighborhood distress and school dropout: The variable significance of community context. *Social Science Research, 32*(4), 659-698. doi:10.1016/s0049-089x(03)00035-8.
75. Bozick, R., & DeLuca, S. (2011). Not making the transition to college: School, work, and opportunities in the lives of American youth. *Social Science Research, 40*(4). doi:10.1016/j.ssresearch.2011.02.003.
76. Stewart, E. B., Stewart, E. A., & Simons, R. L. (2007). The effect of neighborhood context on the college aspirations of African American adolescents. *American Educational Research Journal, 44*(4). doi:10.3102/0002831207308637.
77. Brumley, L. D., Russell, M. A., & Jaffee, S. R. (2019). College expectations promote college attendance: Evidence from a quasiexperimental sibling study. *Psychological Science, 30*(8), 1186-1194. doi:10.1177/0956797619855385.

78. Smith, J., Hurwitz, M., & Avery, C. (2016). Giving college credit where it is due: Advanced placement exam scores and college outcomes. *Journal of Labor Economics, 35*(1), 67-147. doi:10.1086/687568.
79. Shaw, E. J., Marini, J. P., & Mattern, K. D. (2012). Exploring the utility of Advanced Placement participation and performance in college admission decisions. *Educational and Psychological Measurement, 73*(2), 229-253. doi:10.1177/0013164412454291.
80. Jackson, C. K. (2010). A little now for a lot later: A look at a Texas Advanced Placement incentive program. *Journal of Human Resources, 45*(3), 591-639.
81. Card, D. (1993). *Using geographic variation in college proximity to estimate the return to schooling*. Retrieved from website: https://www.nber.org/papers/w4483.
82. Kling, J. R. (2001). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business & Economic Statistics, 19*(3), 358-364.
83. Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly Journal of Economics, 118*(4), 1495-1532. doi:10.1162/003355303322552856.
84. Fletcher, J., & Noghanibehambari, H. (2024). The effects of education on mortality: Evidence using college expansions. *Health Economics, 33*(3). doi:10.1002/hec.4787.
85. Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review, 94*(1), 155-189. doi:10.1257/000282804322970751.
86. Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics, 91*(11-12), 2213-2229. doi:10.1016/j.jpubeco.2007.02.002.
87. Rauscher, E. (2014). Hidden gains: Effects of early U.S. compulsory schooling laws on attendance and attainment by social background. *Educational Evaluation and Policy Analysis, 36*(4), 501-518.
88. Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics, 106*(4), 979-1014. doi:10.2307/2937954.
89. Caldas, S. J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research, 90*(5), 269-277. doi:10.1080/00220671.1997.10544583.
90. Catsambis, S., & Beveridge, A. A. (2001). Does neighborhood matter? Family, neighborhood, and school influences on eighth-grade mathematics achievement. *Sociological Focus, 34*(4), 435-457. doi:10.1080/00380237.2001.10571212.
91. Portes, A., & Hao, L. (2004). The schooling of children of immigrants: contextual effects on the educational attainment of the second generation. *Proceedings of the National Academy of Sciences, 101*(33), 11920-11927. doi:10.1073/pnas.0403418101.
92. Jargowsky, P. A., & El Komi, M. (2011). Before or after the bell? School context and neighborhood effects on student achievement. In H. B. Newburger, E. L. Birch, & S. M. Wachter (Eds.), *How Place Matters in Modern America* (pp. 50-72): University of Pennsylvania Press.
93. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.
94. Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics, 6*(2), 127-151. doi:10.1257/app.6.2.127.
95. Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics, 6*(1), 801-825. doi:10.1146/annurev-economics-080213-040845.
96. Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives, 24*(3), 97-117.
97. Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics, 130*, 105-119. doi:10.1016/j.jpubeco.2015.02.008.
98. Ressler, R. W. (2020). What village? Opportunities and supports for parental involvement outside of the family context. *Children and Youth Services Review, 108*. doi:10.1016/j.childyouth.2019.104575.
99. Ressler, R. W., Paxton, P., Velasco, K., Pivnick, L., Weiss, I., & Eichstaedt, J. C. (2021). Nonprofits: A public policy tool for the promotion of community subjective well-being. *Journal of Public Administration Research and Theory, 31*(4), 822-838. doi:10.1093/jopart/muab010.
100. Mayer, D. J. (2023). Understanding social disorganization and the nonprofit infrastructure: An ecological study of child maltreatment rates. *Social Currents, 10*(5). doi:10.1177/23294965231159317.
101. Hoffman, J. S., Shandas, V., & Pendleton, N. (2020). The effects of historical housing policies on resident exposure to intra-urban heat: A study of 108 US urban areas. *Climate, 8*(1), 12. doi:10.3390/cli8010012.
102. Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives, 33*(1), 185-208.
103. Bell, S. L., Foley, R., Houghton, F., Maddrell, A., & Williams, A. M. (2018). From therapeutic landscapes to healthy spaces, places and practices: A scoping review. *Social Science & Medicine, 196*, 123-130. doi:10.1016/j.socscimed.2017.11.035.

104. Cushing, L., Morello-Frosch, R., Wander, M., & Pastor, M. (2015). The haves, the have-nots, and the health of everyone: the relationship between social inequality and environmental quality. *Annual Review of Public Health, 36*, 193-209. doi:10.1146/annurev-publhealth-031914-122646.
105. Schell, C. J., Dyson, K., Fuentes, T. L., Des Roches, S., Harris, N. C., Miller, D. S., . . . Lambert, M. R. (2020). The ecological and evolutionary consequences of systemic racism in urban environments. *Science, 369*(6510). doi:10.1126/science.aay4497.
106. Berberian, A. G., Gonzalez, D. J. X., & Cushing, L. J. (2022). Racial disparities in climate change-related health effects in the United States. *Current Environmental Health Reports, 9*(3), 451-464. doi:10.1007/s40572-022-00360-w.
107. Moore, L. V., & Diez Roux, A. V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health, 96*(2), 325-331. doi:10.2105/AJPH.2004.058040.
108. Sanchez-Vaznaugh, E. V., Weverka, A., Matsuzaki, M., & Sanchez, B. N. (2019). Changes in Fast Food Outlet Availability Near Schools: Unequal Patterns by Income, Race/Ethnicity, and Urbanicity. *American Journal of Preventive Medicine, 57*(3), 338-345. doi:10.1016/j.amepre.2019.04.023.
109. Morland, K., Wing, S., Diez Roux, A., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine, 22*(1), 23-29. doi:10.1016/s0749-3797(01)00403-2.
110. Chay, K. Y., & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics, 118*(3). doi:10.1162/00335530360698513.
111. Currie, J., & Walker, R. (2011). Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics, 3*(1). doi:10.1257/app.3.1.65.
112. Currie, J., Ray, S. H., & Neidell, M. (2011). Quasi-experimental studies suggest that lowering air pollution levels benefits infants' and children's health. *Health Affairs, 30*(12), 2391-2399. doi:10.1377/hlthaff.2011.0212.
113. Heissel, J. A., Persico, C., & Simon, D. (2022). Does pollution drive achievement? The effect of traffic pollution on academic performance. *Journal of Human Resources, 57*(3), 747-776. doi:10.3368/jhr.57.3.1218-9903r2.
114. Isen, A., Rossin-Slater, M., & Walker, W. R. (2017). Every breath you take—every dollar you'll make: The long-term consequences of the Clean Air Act of 1970. *Journal of Political Economy, 125*(3), 848-902. doi:10.1086/691465.
115. Mohai, P., Kweon, B. S., Lee, S., & Ard, K. (2011). Air pollution around schools is linked to poorer student health and academic performance. *Health Affairs, 30*(5), 852-862. doi:10.1377/hlthaff.2011.0077.
116. Moore, K., Neugebauer, R., Lurmann, F., Hall, J., Brajer, V., Alcorn, S., & Tager, I. (2008). Ambient ozone concentrations cause increased hospitalizations for asthma in children: an 18-year study in Southern California. *Environmental Health Perspectives, 116*(8), 1063-1070. doi:10.1289/ehp.10497.
117. Salam, M. T., Millstein, J., Li, Y. F., Lurmann, F. W., Margolis, H. G., & Gilliland, F. D. (2005). Birth outcomes and prenatal exposure to ozone, carbon monoxide, and particulate matter: results from the Children's Health Study. *Environmental Health Perspectives, 113*(11), 1638-1644. doi:10.1289/ehp.8111.
118. Yuan, A., Halabicky, O., Rao, H., & Liu, J. (2023). Lifetime air pollution exposure, cognitive deficits, and brain imaging outcomes: A systematic review. *Neurotoxicology, 96*, 69-80. doi:10.1016/j.neuro.2023.03.006.
119. Castagna, A., Mascheroni, E., Fustinoni, S., & Montirosso, R. (2022). Air pollution and neurodevelopmental skills in preschool- and school-aged children: A systematic review. *Neuroscience & Biobehavioral Reviews, 136*, 104623. doi:10.1016/j.neubiorev.2022.104623.
120. Dominski, F. H., Lorenzetti Branco, J. H., Buonanno, G., Stabile, L., Gameiro da Silva, M., & Andrade, A. (2021). Effects of air pollution on health: A mapping review of systematic reviews and meta-analyses. *Environmental Research, 201*, 111487. doi:10.1016/j.envres.2021.111487.
121. Khalili, R., Bartell, S. M., Hu, X., Liu, Y., Chang, H. H., Belanoff, C., . . . Vieira, V. M. (2018). Early-life exposure to PM(2.5) and risk of acute asthma clinical encounters among children in Massachusetts: a case-crossover analysis. *Environmental Health, 17*(1), 20. doi:10.1186/s12940-018-0361-6.
122. Strickland, M. J., Hao, H., Hu, X., Chang, H. H., Darrow, L. A., & Liu, Y. (2016). Pediatric emergency visits and short-term changes in PM2.5 concentrations in the U.S. State of Georgia. *Environmental Health Perspectives, 124*(5), 690-696. doi:10.1289/ehp.1509856.
123. Aguilar-Gomez, S., Dwyer, H., Zivin, J. S. G., & Neidell, M. (2022). *This is air: The "non-health" effects of air pollution*. Retrieved from website: https://www.nber.org/papers/w29848.
124. Agarwal, N., Banternghansa, C., & Bui, L. T. (2010). Toxic exposure in America: estimating fetal and infant health outcomes from 14 years of TRI reporting. *Journal of Health Economics, 29*(4), 557-574. doi:10.1016/j.jhealeco.2010.04.002.
125. Liu, J., & Lewis, G. (2014). Environmental toxicity and poor cognitive outcomes in children and adults. *Journal of environmental health, 76*(6), 130-138.
126. Hill, D. T., Jandev, V., Petroni, M., Atallah-Yunes, N., Bendinskas, K., Brann, L. S., . . . Collins, M. B. (2023). Airborne levels of cadmium are correlated with urinary cadmium concentrations among young children living in the New York state city of Syracuse, USA. *Environmental Research, 223*, 115450. doi:10.1016/j.envres.2023.115450.

127. Hill, D. T., Petroni, M., Larsen, D. A., Bendinskas, K., Heffernan, K., Atallah-Yunes, N., . . . Gump, B. B. (2021). Linking metal (Pb, Hg, Cd) industrial air pollution risk to blood metal levels and cardiovascular functioning and structure among children in Syracuse, NY. *Environmental Research, 193*, 110557. doi:10.1016/j.envres.2020.110557.
128. Boyle, J., Ward, M. H., Cerhan, J. R., Rothman, N., & Wheeler, D. C. (2023). Modeling historic environmental pollutant exposures and non-Hodgkin lymphoma risk. *Environmental Research, 224*, 115506. doi:10.1016/j.envres.2023.115506.
129. Gatzke-Kopp, L. M., Warkentien, S., Willoughby, M., Fowler, C., Folch, D. C., & Blair, C. (2021). Proximity to sources of airborne lead is associated with reductions in Children's executive function in the first four years of life. *Health & Place, 68*, 102517. doi:10.1016/j.healthplace.2021.102517.
130. Currie, J., Greenstone, M., & Moretti, E. (2011). Superfund cleanups and infant health. *American Economic Review, 101*(3), 435-441. doi:10.1257/aer.101.3.435.
131. Persico, C., Figlio, D., & Roth, J. (2019). The developmental consequences of Superfund sites. *Journal of Labor Economics, 38*(4), 1055-1097. doi:10.1086/706807.
132. Vrijheid, M. (2000). Health effects of residence near hazardous waste landfill sites: a review of epidemiologic literature. *Environmental Health Perspectives, 108 Suppl 1*(Suppl 1), 101-112. doi:10.1289/ehp.00108s1101.
133. Kiaghadi, A., Rifai, H. S., & Dawson, C. N. (2021). The presence of Superfund sites as a determinant of life expectancy in the United States. *Nature Communications, 12*(1), 1947. doi:10.1038/s41467-021-22249-2.
134. Klemick, H., Mason, H., & Sullivan, K. (2020). Superfund Cleanups and Children's Lead Exposure. *Journal of Environmental Economics and Management, 100*. doi:10.1016/j.jeem.2019.102289.
135. Dolk, H., Vrijheid, M., Armstrong, B., Abramsky, L., Bianchi, F., Garne, E., . . . Tenconi, R. (1998). Risk of congenital anomalies near hazardous-waste landfill sites in Europe: the EUROHAZCON study. *Lancet, 352*(9126), 423-427. doi:10.1016/s0140-6736(98)01352-x.
136. Yin, Y., He, L., Wennberg, P. O., & Frankenberg, C. (2023). Unequal exposure to heatwaves in Los Angeles: Impact of uneven green spaces. *Science Advances, 9*(17), eade8501. doi:10.1126/sciadv.ade8501.
137. Hsu, A., Sheriff, G., Chakraborty, T., & Manya, D. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications, 12*(1), 2721. doi:10.1038/s41467-021-22799-5.
138. Jhun, I., Mata, D. A., Nordio, F., Lee, M., Schwartz, J., & Zanobetti, A. (2017). Ambient temperature and sudden infant death syndrome in the United States United States. *Epidemiology, 28*(5), 728-734. doi:10.1097/EDE.0000000000000703.
139. Basu, R., Rau, R., Pearson, D., & Malig, B. (2018). Temperature and term low birth weight in California. *American Journal of Epidemiology, 187*(11), 2306-2314. doi:10.1093/aje/kwy116.
140. Park, R. J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy, 12*(2). doi:10.1257/pol.20180612.
141. Park, R. J., Behrer, A. P., & Goodman, J. (2021). Learning is inhibited by heat exposure, both internationally and within the United States. *Nature Human Behavior, 5*(1), 19-27. doi:10.1038/s41562-020-00959-9.
142. Deschenes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics, 46*, 606-619. doi:10.1016/j.eneco.2013.10.013.
143. Deschenes, O., Greenstone, M., & Guryan, J. (2009). Climate change and birth weight. *American Economic Review, 99*(2), 211-217. doi:10.1257/aer.99.2.211.
144. Strand, L. B., Barnett, A. G., & Tong, S. (2011). The influence of season and ambient temperature on birth outcomes: a review of the epidemiological literature. *Environmental Research, 111*(3), 451-462. doi:10.1016/j.envres.2011.01.023.
145. Graff Zivin, J., Hsiang, S. M., & Neidell, M. (2017). Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists, 5*(1), 77-105. doi:10.1086/694177.
146. Hartig, T., & Kahn, P. H., Jr. (2016). Living in cities, naturally. *Science, 352*(6288), 938-940. doi:10.1126/science.aaf3759.
147. Hartig, T., Mitchell, R., de Vries, S., & Frumkin, H. (2014). Nature and health. *Annual Review of Public Health, 35*, 207-228. doi:10.1146/annurev-publhealth-032013-182443.
148. James, P., Hart, J. E., Banay, R. F., & Laden, F. (2016). Exposure to greenness and mortality in a nationwide prospective cohort study of women. *Environmental Health Perspectives, 124*(9), 1344-1352. doi:10.1289/ehp.1510363.
149. Sprague, N. L., Bancalari, P., Karim, W., & Siddiq, S. (2022). Growing up green: A systematic review of the influence of greenspace on youth development and health outcomes. *Journal of Exposure Science and Environmental Epidemiology, 32*(5), 660-681. doi:10.1038/s41370-022-00445-6.
150. Makram, O. M., Pan, A., Maddock, J. E., & Kash, B. A. (2024). Nature and mental health in urban Texas: A NatureScore-based study. *International Journal of Environmental Research and Public Health, 21*(2). doi:10.3390/ijerph21020168.
151. Browning, M., Hanley, J. R., Bailey, C. R., Beatley, T., Gailey, S., Hipp, J. A., . . . Minson, C. T. (2024). Quantifying nature: Introducing NatureScore(TM) and NatureDose(TM) as health analysis and promotion tools. *American Journal of Health Promotion, 38*(1), 126-134. doi:10.1177/08901171231210806b.

152. Klompmaker, J. O., Hart, J. E., Bailey, C. R., Browning, M., Casey, J. A., Hanley, J. R., . . . James, P. (2023). Racial, ethnic, and socioeconomic disparities in multiple measures of blue and green spaces in the United States. *Environmental Health Perspectives, 131*(1), 17007. doi:10.1289/EHP11164.
153. Makhlouf, M. H. E., Motairek, I., Chen, Z., Nasir, K., Deo, S. V., Rajagopalan, S., & Al-Kindi, S. G. (2023). Neighborhood walkability and cardiovascular risk in the United States. *Current Problems in Cardiology, 48*(3), 101533. doi:10.1016/j.cpcardiol.2022.101533.
154. Duncan, D. T., Sharifi, M., Melly, S. J., Marshall, R., Sequist, T. D., Rifas-Shiman, S. L., & Taveras, E. M. (2014). Characteristics of walkable built environments and BMI z-scores in children: Evidence from a large electronic health record database. *Environmental Health Perspectives, 122*(12), 1359-1365. doi:10.1289/ehp.1307704.
155. Kowaleski-Jones, L., Zick, C., Smith, K. R., Brown, B., Hanson, H., & Fan, J. (2018). Walkable neighborhoods and obesity: Evaluating effects with a propensity score approach. *SSM Population Health, 6*, 9-15. doi:10.1016/j.ssmph.2017.11.005.
156. Yang, S., Chen, X., Wang, L., Wu, T., Fei, T., Xiao, Q., . . . Jia, P. (2021). Walkability indices and childhood obesity: A review of epidemiologic evidence. *Obesity Reviews, 22 Suppl 1*(Suppl 1), e13096. doi:10.1111/obr.13096.
157. Leyden, K. M. (2003). Social capital and the built environment: the importance of walkable neighborhoods. *American Journal of Public Health, 93*(9), 1546-1551. doi:10.2105/ajph.93.9.1546.
158. Dubowitz, T., Ghosh-Dastidar, M., Cohen, D. A., Beckman, R., Steiner, E. D., Hunter, G. P., . . . Collins, R. L. (2015). Diet and perceptions change with supermarket introduction in a food desert, but not because of supermarket use. *Health Affairs, 34*(11), 1858-1868. doi:10.1377/hlthaff.2015.0667.
159. Thomsen, M. R., Nayga Jr., R. M., Alviola IV, P. A., & Rouse, H. L. (2016). The effect of food deserts on the body mass index of elementary school children. *American Journal of Agricultural Economics, 98*(1), 1-18. doi:https://doi.org/10.1093/ajae/aav039.
160. Cobb, L. K., Appel, L. J., Franco, M., Jones-Smith, J. C., Nur, A., & Anderson, C. A. (2015). The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. *Obesity (Silver Spring), 23*(7), 1331-1344. doi:10.1002/oby.21118.
161. Rummo, P. E., Algur, Y., McAlexander, T., Judd, S. E., Lopez, P. M., Adhikari, S., . . . Elbel, B. (2021). Comparing competing geospatial measures to capture the relationship between the neighborhood food environment and diet. *Annals of Epidemiology, 61*, 1-7. doi:10.1016/j.annepidem.2021.05.005.
162. Jia, P., Xue, H., Cheng, X., & Wang, Y. (2019). Effects of school neighborhood food environments on childhood obesity at multiple scales: A longitudinal kindergarten cohort study in the USA. *BMC Medicine, 17*(1), 99. doi:10.1186/s12916-019-1329-2.
163. Rogus, S., Athens, J., Cantor, J., & Elbel, B. (2018). Measuring micro-level effects of a new supermarket: Do residents within 0.5 mile have improved dietary behaviors? *Journal of the Academy of Nutrition and Dietetics, 118*(6), 1037-1046. doi:10.1016/j.jand.2017.06.360.
164. Currie, J., DellaVigna, S., Moretti, E., & Pathania, V. (2010). The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy, 2*(3), 32-63. doi:10.1257/pol.2.3.32.
165. Zick, C. D., Curtis, D. S., Meeks, H., Smith, K. R., Brown, B. B., Kole, K., & Kowaleski-Jones, L. (2023). The changing food environment and neighborhood prevalence of type 2 diabetes. *SSM Population Health, 21*, 101338. doi:10.1016/j.ssmph.2023.101338.
166. Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics, 134*(4), 1793-1844. doi:10.1093/qje/qjz015.
167. Ver Ploeg, M., & Rahkovsky, I. (2016). Recent evidence on the effects of food store access on food choice and diet quality. Retrieved from https://www.ers.usda.gov/amber-waves/2016/may/recent-evidence-on-the-effects-of-food-store-access-on-food-choice-and-diet-quality/.
168. Sharkey, P. (2010). The acute effect of local homicides on children's cognitive performance. *Proceedings of the National Academy of Sciences, 107*(26), 11733-11738. doi:10.1073/pnas.1000690107.
169. Sharkey, P. T., Tirado-Strayer, N., Papachristos, A. V., & Raver, C. C. (2012). The effect of local violence on children's attention and impulse control. *American Journal of Public Health, 102*(12), 2287-2293. doi:10.2105/AJPH.2012.300789.
170. Burdick-Will, J. (2013). School violent crime and academic achievement in Chicago. *Sociology of Education, 86*(4). doi:10.1177/0038040713494225.
171. Burdick-Will, J. (2016). Neighborhood violent crime and academic growth in Chicago: Lasting effects of early exposure. *Social Forces, 95*(1), 133-158. doi:10.1093/sf/sow041.
172. Manduca, R., & Sampson, R. J. (2019). Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proceedings of the National Academy of Sciences, 116*(16), 7772-7777. doi:10.1073/pnas.1820464116.
173. Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics, 133*(3), 1163-1228. doi:10.1093/qje/qjy006.
174. Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics, 129*(4), 1553-1623. doi:10.1093/qje/qju022.

175. Sharkey, P., Torrats-Espinosa, G., & Takyar, D. (2017). Community and the crime decline: The causal effect of local nonprofits on violent crime. *American Sociological Review, 82*(6). doi:10.1177/0003122417736289.
176. Branas, C. C., Cheney, R. A., MacDonald, J. M., Tam, V. W., Jackson, T. D., & Ten Have, T. R. (2011). A difference-in-differences analysis of health, safety, and greening vacant urban space. *American Journal of Epidemiology, 174*(11), 1296-1306. doi:10.1093/aje/kwr273.
177. Branas, C. C., South, E., Kondo, M. C., Hohl, B. C., Bourgois, P., Wiebe, D. J., & MacDonald, J. M. (2018). Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proceedings of the National Academy of Sciences, 115*(12), 2946-2951. doi:10.1073/pnas.1718503115.
178. Kondo, M. C., Andreyeva, E., South, E. C., MacDonald, J. M., & Branas, C. C. (2018). Neighborhood interventions to reduce violence. *Annual Review of Public Health, 39*, 253-271. doi:10.1146/annurev-publhealth-040617-014600.
179. Moyer, R., MacDonald, J. M., Ridgeway, G., & Branas, C. C. (2019). Effect of remediating blighted vacant land on shootings: A citywide cluster randomized trial. *American Journal of Public Health, 109*(1), 140-144. doi:10.2105/AJPH.2018.304752.
180. South, E. C., Hohl, B. C., Kondo, M. C., MacDonald, J. M., & Branas, C. C. (2018). Effect of greening vacant land on mental health of community-dwelling adults: A cluster randomized trial. *JAMA Network Open, 1*(3), e180298. doi:10.1001/jamanetworkopen.2018.0298.
181. South, E. C., MacDonald, J. M., Tam, V. W., Ridgeway, G., & Branas, C. C. (2023). Effect of abandoned housing interventions on gun violence, perceptions of safety, and substance use in black neighborhoods: A citywide cluster randomized trial. *JAMA Internal Medicine, 183*(1), 31-39. doi:10.1001/jamainternmed.2022.5460.
182. Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of Medicare. *The Quarterly Journal of Economics, 122*(1), 1-37. doi:10.1162/qjec.122.1.1.
183. Sommers, B. D., Maylone, B., Blendon, R. J., Orav, E. J., & Epstein, A. M. (2017). Three-year impacts of the affordable care act: Improved medical care and health among low-income adults. *Health Affairs, 36*(6), 1119-1128. doi:10.1377/hlthaff.2017.0293.
184. Wishner, J. B., & Burton, R. A. (2017). How Have Providers Responded to the Increased Demand for Health Care Under the Affordable Care Act? Retrieved from https://www.urban.org/research/publication/how-have-providers-responded-increased-demand-health-care-under-affordable-care-act.
185. Escarce, J. J., Wozniak, G. D., Tsipas, S., Pane, J. D., Brotherton, S. E., & Yu, H. (2021). Effects of the Affordable Care Act Medicaid expansion on the distribution of new general internists across states. *Medical Care, 59*(7), 653-660. doi:10.1097/MLR.0000000000001523.
186. Escarce, J. J., Wozniak, G. D., Tsipas, S., Pane, J. D., Ma, Y., Brotherton, S. E., & Yu, H. (2022). The Affordable Care Act Medicaid expansion, social disadvantage, and the practice location choices of new general internists. *Medical Care, 60*(5), 342-350. doi:10.1097/MLR.0000000000001703.
187. Haslam, A., Nesbit, R., & Christensen, R. K. (2019). The dynamic impact of nonprofit organizations: Are health-related nonprofit organizations associated with improvements in obesity at the community level? *, 10*(3). doi:doi:10.1515/npf-2018-0040.
188. Sampson, R. J., Sharkey, P., & Raudenbush, S. W. (2008). Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proceedings of the National Academy of Sciences, 105*(3), 845-852. doi:10.1073/pnas.0710189104.
189. Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review, 76*(5), 713-736. doi:10.1177/0003122411420816.
190. Burdick-Will, J., Ludwig, J., Raudenbush, S. W., Sampson, R. J., Sanbonmatsu, L., & Sharkey, P. (2011). Converging Evidence for Neighborhood Effects on Children's Test Scores: An Experimental, Quasi-Experimental, and Observational Comparison. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity?* (pp. 255-276): Russell Sage Foundation.
191. Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review, 103*(6). doi:10.1257/aer.103.6.2121.
192. Autor, D., Dorn, D., & Hanson, G. (2019). When work disappears: Manufacturing decline and the falling marriage market value of young men. *American Economic Review: Insights, 1*(2). doi:10.1257/aeri.20180010.
193. Krieger, N., Waterman, P. D., Spasojevic, J., Li, W., Maduro, G., & Van Wye, G. (2016). Public health monitoring of privilege and deprivation with the Index of Concentration at the Extremes. *American Journal of Public Health, 106*(2). doi:10.2105/AJPH.2015.302955.
194. Krieger, N., Kim, R., Feldman, J., & Waterman, P. D. (2018). Using the Index of Concentration at the Extremes at multiple geographical levels to monitor health inequities in an era of growing spatial social polarization: Massachusetts, USA (2010-14). *International Journal of Epidemiology, 47*(3), 788-819. doi:10.1093/ije/dyy004.
195. Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., . . . Cutler, D. (2016). The association netween income and life expectancy in the United States, 2001-2014. *JAMA, 315*(16), 1750-1766. doi:10.1001/jama.2016.4226.
196. Egen, O., Beatty, K., Blackley, D. J., Brown, K., & Wykoff, R. (2017). Health and social conditions of the poorest versus wealthiest counties in the United States. *American Journal of Public Health, 107*(1), 130-135. doi:10.2105/AJPH.2016.303515.

197. Dunn, J. R. (2020). Housing and healthy child development: Known and potential impacts of interventions. *Annual Review of Public Health, 41*, 381-396. doi:10.1146/annurev-publhealth-040119-094050.
198. Thomson, H., Thomas, S., Sellstrom, E., & Petticrew, M. (2013). Housing improvements for health and associated socio-economic outcomes. *Cochrane Database of Systematic Reviews*(2), CD008657. doi:10.1002/14651858.CD008657.pub2.
199. Li, Y., Spoer, B. R., Lampe, T. M., Hsieh, P. Y., Nelson, I. S., Vierse, A., . . . Gourevitch, M. N. (2023). Racial/ethnic and income disparities in neighborhood-level broadband access in 905 US cities, 2017-2021. *Public Health, 217*, 205-211. doi:10.1016/j.puhe.2023.02.001.
200. Solari, C. D., & Mare, R. D. (2012). Housing crowding effects on children's wellbeing. *Social Science Research, 41*(2), 464-476. doi:10.1016/j.ssresearch.2011.09.012.
201. Conley, D. (2001). A Room With a View or a room of one's own? Housing and social stratification. *Sociological Forum, 16*(2), 263-280. doi:10.1023/A:1011052701810.
202. Lopoo, L. M., & London, A. S. (2016). Household crowding during childhood and long-term education outcomes. *Demography, 53*(3), 699-721. doi:10.1007/s13524-016-0467-9.
203. Gove, W. R., Hughes, M., & Galle, O. R. (1979). Overcrowding in the Home: An empirical investigation of its possible pathological consequences. *American Sociological Review, 44*(1), 59-80. doi:10.2307/2094818.
204. McLaughlin, J. M., Khan, F., Pugh, S., Angulo, F. J., Schmitt, H. J., Isturiz, R. E., . . . Swerdlow, D. L. (2021). County-level predictors of Coronavirus Disease 2019 (COVID-19) cases and deaths in the United States: What happened, and where do we go from here? *Clinical Infectious Diseases, 73*(7), e1814-e1821. doi:10.1093/cid/ciaa1729.
205. Dasgupta, S., Bowen, V. B., Leidner, A., Fletcher, K., Musial, T., Rose, C., . . . Oster, A. M. (2020). Association between social vulnerability and a county's risk for becoming a COVID-19 Hotspot - United States, June 1-July 25, 2020. *Morbidity and Mortality Weekly Report, 69*(42), 1535-1541. doi:10.15585/mmwr.mm6942a3.
206. Acevedo-Garcia, D. (2001). ZIP code-level risk factors for tuberculosis: neighborhood environment and residential segregation in New Jersey, 1985-1992. *American Journal of Public Health, 91*(5), 734-741. doi:10.2105/ajph.91.5.734.
207. Stevenson, B. (2008). *The internet and job search*. Retrieved from website: https://www.nber.org/papers/w13886.
208. Zuo, G. W. (2021). Wired and hired: employment effects of subsidized broadband internet for low-income Americans. *American Economic Journal: Economic Policy, 13*(3), 447-482. doi:10.1257/pol.20190648.
209. Dettling, L. J. (2017). Broadband in the labor market:The impact of residential high-speed internet on married women's labor force participation. *ILR Review, 70*(2), 451-482. doi:10.1177/0019793916644721.
210. Dettling, L. J., Goodman, S., & Smith, J. (2018). Every little bit counts: The impact of high-speed internet on the transition to college. *The Review of Economics and Statistics, 100*(2), 260-273. doi:10.1162/REST_a_00712.
211. Caldarulo, M., Mossberger, K., & Howell, A. (2023). Community-wide broadband adoption and student academic achievement. *Telecommunications Policy, 47*(1), 102445. doi:https://doi.org/10.1016/j.telpol.2022.102445.
212. Hampton, K. N., Robertson, C. T., Fernandez, L., Shin, I., & Bauer, J. M. (2021). How variation in internet access, digital skills, and media use are related to rural student outcomes: GPA, SAT, and educational aspirations. *Telematics and Informatics, 63*, 101666. doi:https://doi.org/10.1016/j.tele.2021.101666.
213. Fahle, E. M., Kane, T. J., Patterson, T., Reardon, S. F., Staiger, D. O., & Stuart, E. A. (2023). *School district and community factors associated with learning loss during the*
COVID-19 pandemic. Retrieved from website: https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/explaining_covid_losses_5.23.pdf.
214. Hampton, K. N., Hales, G., & Bauer, J. M. (2023). *Broadband and student performance gaps after the COVID-19 pandemic*. Retrieved from website: https://ssrn.com/abstract=4574526.
215. Bauerly, B. C., McCord, R. F., Hulkower, R., & Pepin, D. (2019). Broadband access as a public health issue: The role of law in expanding broadband access and connecting underserved communities for better health outcomes. *Journal of Law, Medicine & Ethics, 47*(S2), 39-42. doi:10.1177/1073110519857314.
216. Graves, J. M., Abshire, D. A., Amiri, S., & Mackelprang, J. L. (2021). Disparities in technology and broadband internet access across rurality: Implications for health and education. *Family & Community Health, 44*(4), 257-265. doi:10.1097/fch.0000000000000306.
217. van Parys, J., & Brown, Z. Y. (2023). *Broadband internet access and health outcomes: Patient and provider responses in Medicare*. Retrieved from website: https://www.nber.org/papers/w31579.
218. Putnam, R. D. (2020). *Bowling alone: the collapse and revival of American community*. New York: Simon & Schuster.
219. Portes, A. (2000). The two meanings of social capital. *Sociological Forum, 15*(1), 1-12.
220. Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, S95-S120. doi:10.1086/228943.
221. Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., . . . Wernerfelt, N. (2022). Social capital I: Measurement and associations with economic mobility. *Nature, 608*(7921), 108-121. doi:10.1038/s41586-022-04996-4.

222. Lim, S., Min, B. H., & Berlan, D. G. (2023). The nonprofit role in building community social capital: A moderated mediation model of organizational learning, innovation, and shared Mission for social capital creation. *Nonprofit and Voluntary Sector Quarterly, 53*(1), 210-235. doi:10.1177/08997640221146965.
223. Small, M. L., & Allard, S. W. (2013). *Reconsidering the urban disadvantaged: The role of systems, institutions, and organizations*. Los Angeles: SAGE Publications.
224. Cheng, Y., & Hung, C. (2023). Exploring the moderators of the relationship between nonprofit sector size and its societal impact: A meta-analysis. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*. doi:10.1007/s11266-023-00604-9.
225. Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic ppportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics, 135*(2), 711-783. doi:10.1093/qje/qjz042.
226. Cross, C. J. (2020). Racial/ethnic differences in the association between family structure and children's education. *Journal of Marriage and Family, 82*(2), 691-712. doi:https://doi.org/10.1111/jomf.12625.
227. Cleveland, H. H., & Gilson, M. (2004). The effects of neighborhood proportion of single-parent families and mother-adolescent relationships on adolescents' number of sexual partners. *Journal of Youth and Adolescence, 33*(4), 319-329. doi:10.1023/B:JOYO.0000032640.25593.9f.
228. Sampson, R. J. (1987). Urban black violence: The effect of male joblessness and family disruption. *American Journal of Sociology, 93*(2), 348-382. doi:10.1086/228748.
229. Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94*(4), 774-802. doi:10.1086/229068.
230. Rothstein, R. (2017). *The color of law: A forgotten history of how our government segregated America* (First edition. ed.). New York; London: Liveright Publishing Corporation.
231. Quillian, L., Lee, J. J., & Honoré, B. (2020). Racial discrimination in the U.S. housing and mortgage lending markets: A quantitative review of trends, 1976–2016. *Race and Social Problems, 12*(1), 13-28. doi:10.1007/s12552-019-09276-x.
232. Miller, P., Podvysotska, T., Betancur, L., & Votruba-Drzal, E. (2021). Wealth and child development: Differences in associations by family income and developmental stage. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 7*(3), 154. doi:10.7758/RSF.2021.7.3.07.
233. Thomas, H., Mann, A., & Meschede, T. (2018). Race and location: The role neighborhoods play in family wealth and well-being. *The American Journal of Economics and Sociology, 77*(3-4), 1077-1111. doi:https://doi.org/10.1111/ajes.12239.
234. Pfeffer, F. T., & Schoeni, R. F. (2016). How wealth inequality shapes our future. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*(6), 2-22. doi:10.7758/rsf.2016.2.6.01.
235. Pfeffer, F. T., & Killewald, A. (2016). Intergenerational correlations in wealth. In F. R. B. o. S. Louis (Ed.), *Economic mobility: Research and ideas on strengthening families, communities and the economy*.
236. Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics, 114*(2), 577-599. doi:10.1162/003355399556070.
237. Black, S. E., & Machin, S. (2011). Housing valuations of school performance. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 485-519): Elsevier.
238. Machin, S. (2011). Houses and schools: Valuation of school quality through the housing market. *Labour Economics, 18*(6), 723-729. doi:https://doi.org/10.1016/j.labeco.2011.05.005.
239. Holt, J. R., & Borsuk, M. E. (2020). Using Zillow data to value green space amenities at the neighborhood scale. *Urban Forestry & Urban Greening, 56*, 126794. doi:https://doi.org/10.1016/j.ufug.2020.126794.
240. Crompton, J. L., & Nicholls, S. (2020). Impact on property values of distance to parks and open spaces: An update of U.S. studies in the new millennium. *Journal of Leisure Research, 51*(2), 127-146. doi:10.1080/00222216.2019.1637704.
241. Chay, Kenneth Y., & Greenstone, M. (2005). Does air quality matter? Evidence from the housing market. *Journal of Political Economy, 113*(2), 376-424. doi:10.1086/427462.
242. Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings. *American Economic Review, 105*(2), 678-709. doi:10.1257/aer.20121656.
243. Lang, C., VanCeylon, J., & Ando, A. W. (2023). Distribution of capitalized benefits from land conservation. *Proceedings of the National Academy of Sciences, 120*(18), e2215262120. doi:10.1073/pnas.2215262120.
244. Card, D., Hallock, K. F., & Moretti, E. (2010). The geography of giving: The effect of corporate headquarters on local charities. *Journal of Public Economics, 94*(3), 222-234. doi:https://doi.org/10.1016/j.jpubeco.2009.11.010.
245. Dyer, Z., Alcusky, M. J., Galea, S., & Ash, A. (2023). Measuring the enduring imprint of structural racism on American neighborhoods. *Health Affairs, 42*(10), 1374-1382. doi:10.1377/hlthaff.2023.00659.
246. Derenoncourt, E. (2022). Can you move to opportunity? Evidence from the great migration. *American Economic Review, 112*(2), 369-408. doi:10.1257/aer.20200002.
247. Boustan, L. P. (2010). Was postwar suburbanization “white flight”? Evidence from the black migration. *The Quarterly Journal of Economics, 125*(1), 417-443. doi:10.1162/qjec.2010.125.1.417.

248. NatureQuant. (2020). *Delivering technology to assess and promote nature exposure*. Retrieved from website: https://www.naturequant.com/.
249. Child Opportunity Index 2.0 database. (2020). Retrieved from https://data.diversitydatakids.org/dataset/coi20-child-opportunity-index-2-0-database.
250. Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American Journal of Epidemiology, 179*(8). doi:10.1093/aje/kwu018.
251. Zhang, X., Holt, J. B., Yun, S., Lu, H., Greenlund, K. J., & Croft, J. B. (2015). Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. *American Journal of Epidemiology, 182*(2). doi:10.1093/aje/kwv002.
252. Arias, E., Escobedo, L. A., Kennedy, J., Fu, C., & Cisewki, J. (2018). U.S. Small-area Life Expectancy Estimates Project: Methodology and results summary. *Vital Health Stat 2*(181), 1-40.
253. Kind, A. J., Jencks, S., Brock, J., Yu, M., Bartels, C., Ehlenbach, W., . . . Smith, M. (2014). Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. *Ann Intern Med, 161*(11), 765-774. doi:10.7326/M13-2946.
254. EPA releases annual air report, highlighting trends through 2022. (2023). Retrieved from https://www.epa.gov/newsreleases/epa-releases-annual-air-report-highlighting-trends-through-2022.
255. *Health coverage under the Affordable Care Act: Current enrollment trends and state estimates*. (2023). Retrieved from website: https://aspe.hhs.gov/reports/current-health-coverage-under-affordable-care-act.
256. Public High School Graduation Rates. (2023). Retrieved from https://nces.ed.gov/programs/coe/indicator/coi/high-school-graduation-rates.
257. Autor, D., Dube, A., & McGrew, A. (2023). *The unexpected compression: Competition at work in the low wage labor market*. Retrieved from website: https://dx.doi.org/10.3386/w31010.
258. Climate Change Indicators, Heat Waves, Figure 3. Retrieved from https://www.epa.gov/climate-indicators/climate-change-indicators-heat-waves.
259. Osypuk, T. L., Galea, S., McArdle, N., & Acevedo-Garcia, D. (2009). Quantifying separate and unequal: Racial-ethnic distributions of neighborhood poverty in metropolitan America. *Urban Affairs Review, 45*(1), 25-65. doi:10.1177/1078087408331119.
260. Rollings, K. A., Noppert, G. A., Griggs, J. J., Melendez, R. A., & Clarke, P. J. (2023). Comparison of two area-level socioeconomic deprivation indices: Implications for public health research, practice, and policy. *PLOS ONE, 18*(10), e0292281. doi:10.1371/journal.pone.0292281.
261. Aris, I. M., Perng, W., Dabelea, D., Padula, A. M., Alshawabkeh, A., Velez-Vega, C. M., . . . Environmental Influences on Child Health, O. (2023). Neighborhood opportunity and vulnerability and incident asthma among children. *JAMA Pediatrics, 177*(10), 1055-1064. doi:10.1001/jamapediatrics.2023.3133.
262. Aris, I. M., Perng, W., Dabelea, D., Padula, A. M., Alshawabkeh, A., Velez-Vega, C. M., . . . Program Collaborators for Environmental Influences on Child Health, O. (2022). Associations of neighborhood opportunity and social vulnerability with trajectories of childhood body mass index and obesity among us children. *JAMA Network Open, 5*(12), e2247957. doi:10.1001/jamanetworkopen.2022.47957.
263. Beyer, L., Keen, R., Ertel, K. A., Okuzono, S. S., Pintro, K., Delaney, S., & Slopen, N. (2024). Comparing two measures of neighborhood quality and internalizing and externalizing behaviors in the adolescent brain cognitive development study. *Social Psychiatry and Psychiatric Epidemiology*. doi:10.1007/s00127-024-02614-4.
264. Software as a Medical Device (SAMD): Clinical vvaluation. (2017). Retrieved from https://www.fda.gov/media/100714/download.
265. Acevedo-Garcia, D., Noelke, C., Ressler, R. W., & Shafer, L. (2023). Improving the infrastructure for neighborhood indices to advance health equity. *Health Affairs Forefront*. doi:10.1377/forefront.20230830.428660.
266. Lavoie, A., Sparks, K., Kasperski, S., Himes-Cornell, A., Hoelting, K., & Maguire, C. (2018). Ground-truthing social vulnerability indices of Alaska fishing communities. *Coastal Management, 46*(5). doi:10.1080/08920753.2018.1498710.
267. Oulahen, G., Mortsch, L., Tang, K., & Harford, D. (2015). Unequal vulnerability to flood hazards: "Ground truthing" a social vulnerability index of five municipalities in metro Vancouver, Canada. *Annals of the Association of American Geographers, 105*(3), 473-495. doi:10.1080/00045608.2015.1012634.
268. Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2022). *Stanford Education Data Archive (Version 4.1)*. Retrieved from website: https://purl.stanford.edu/xv742vh9296.
269. Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2021). Stanford Education Data Archive: Technical documentation (version 4.1) [Press release]. Retrieved from http://purl.stanford.edu/db586ns4974.
270. U.S. Department of Education Civil Rights Data Collection (CRDC). Retrieved from https://www2.ed.gov/about/offices/list/ocr/data.html.
271. Four-Year Adjusted-Cohort Graduation Rates. Retrieved from https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html.

272. National Center for Charitable Statistics (NCCS) IRS Business Master File, Exempt Organizations. Retrieved from https://datacatalog.urban.org/dataset/national-center-charitable-statistics-ncc-irs-business-master-file.
273. Master List of CRDC Definitions, 2017-18.
274. U.S. Department of Education Common Core of Data (CCD). Retrieved from https://nces.ed.gov/ccd/.
275. Fused Air Quality Surface Using Downscaling (FAQSD) output files. Retrieved from https://www.epa.gov/hesc/rsig-related-downloadable-data-files.
276. Reff, A., Phillips, S., Eyth, A., Godfrey, J., Vukovich, J., & Mintz, D. (2023). *Bayesian Space-time Downscaling Fusion Model (Downscaler) - Derived Estimates of Air Quality for 2020*. Retrieved from website: https://ofmpub.epa.gov/rsig/rsigserver?data/FAQSD/docs/2020_DS_Annual_Report.pdf.
277. EPA Risk-Screening Environmental Indicators (RSEI) methodology, version 2.3.10. (2022). Retrieved from https://www.epa.gov/rsei/risk-screening-environmental-indicators-rsei-methodology-0.
278. Superfund: National Priorities List (NPL). Retrieved from https://www.epa.gov/superfund/superfund-national-priorities-list-npl.
279. NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002 (NLDAS_FORA0125_H). Retrieved from https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_002/summary?keywords=NLDAS.
280. National Walkability Index methodology and user guide. (2021). Retrieved from https://www.epa.gov/smartgrowth/national-walkability-index-user-guide-and-methodology.
281. U.S. Environmental Protection Agency's and U.S. General Services Administration Smart Location Database, version 3. Retrieved from https://www.epa.gov/smartgrowth/smart-location-mapping.
282. American Community Survey and Puerto Rico Community Survey 2021 Subject Definitions. (2021). Retrieved from https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html.
283. Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., . . . Wernerfelt, N. (2022). Social capital II: Determinants of economic connectedness. *Nature, 608*(7921), 122-134. doi:10.1038/s41586-022-04997-3.
284. Din, A., & Wilson, R. (2020). Crosswalking ZIP codes to census geographies: Geoprocessing the U.S. Department of Housing & Urban Development's ZIP code crosswalk files. *Cityscape, 22*(1), 293-314.
285. HUD USPS ZIP Code Crosswalk Files. Retrieved from https://www.huduser.gov/portal/datasets/usps_crosswalk.html.
286. Community Eligibility Provision. (2023). Retrieved from https://www.fns.usda.gov/cn/community-eligibility-provision.
287. EDFacts State Assessments in Mathematics and Reading/Language Arts Data Files. Retrieved from https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html.
288. Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
289. Buuren, S. v. (2018). *Flexible imputation of missing data*: https://stefvanbuuren.name/fimd/.
290. Allard, S. W. (2017). *Places in need: The changing geography of poverty*. New York: Russell Sage Foundation.
291. Small, M. L. (2009). *Unanticipated gains: Origins of network inequality in everyday life*. New York; Oxford: Oxford University Press.
292. Nelson, I. A. (2016). *Why afterschool matters*. New Brunswick, NJ: Rutgers University Press.
293. DiMaggio, P. J., & Powell, W. W. (2000). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. In J. A. C. Baum & F. Dobbin (Eds.), *Economics Meets Sociology in Strategic Management* (Vol. 17, pp. 143-166): Emerald Group Publishing Limited.
294. Exempt Purposes - Internal Revenue Code Section 501(c)(3). (2023). Retrieved from https://www.irs.gov/charities-non-profits/charitable-organizations/exempt-purposes-internal-revenue-code-section-501c3.