Child Opportunity Index 2.0

Technical Documentation

diversitydatakids.org

January 15, 2020

Clemens Noelke, Nancy McArdle, Mikyung Baek, Nick Huntington,
Rebecca Huber, Erin Hardy, Dolores Acevedo-Garcia

Institute for Child, Youth and Family Policy
The Heller School for Social Policy and Management
Brandeis University

Suggested Citation:

Noelke, C., McArdle, N., Baek, M., Huntington, N., Huber, R., Hardy, E., & Acevedo-Garcia, D. (2020). Child Opportunity Index 2.0 Technical Documentation. Retrieved from diversitydatakids.org/research-library/research-brief/how-we-built-it..

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 the aforementioned funding agencies.

Table of contents

Introduction	6
Domains and indicators	7
Data and methods	12
Results	21
Appendix 1.1: Education domain indicators and sources	28
Appendix 1.2: Health and environment domain indicators and sources	30
Appendix 1.3: Social and economic domain indicators and sources	33
Appendix 2: School indicators	35
Appendix 3: Linking census tracts over time	42
References	44

List of abbreviations

ACS	American Community Survey
ASD	Administrative school district
CDC	Centers for Disease Control and Prevention
CCD	NCES Common Core of Data
COI	Child Opportunity Index
ECE	Early childhood education
EPA	Environmental Protection Agency
FRPL	Free or reduced-price lunch
GS	GreatSchools
GSD	Geographic school district
NAEP	National Assessment of Educational Progress
NCHS	National Center for Health Statistics
NCES	National Center for Education Statistics
CRDC	Civil Rights Data Collection
OLS	Ordinary Least Squares
RMSE	Root Mean Squared Error
RWJF	Robert Wood Johnson Foundation
SEDA	Stanford Education Data Archive
SES	Socioeconomic status
USDA	United States Department of Agriculture
USALEEP	U.S. Small-area Life Expectancy Project
WKKF	W.K. Kellogg Foundation

"Children are not the people of tomorrow, but people today... They should be allowed to grow into whoever they were meant to be—The unknown person inside each of them is the hope for the future."

Janusz Korczak, Polish pediatrician, author and Holocaust survivor

Preface

Established in 2014 with support from the W.K. Kellogg Foundation and the Robert Wood Johnson Foundation, diversitydatakids.org set out to fill an urgent need for a rigorous, equity-focused research program with a clear mission to help improve child wellbeing and increase racial and ethnic equity in opportunities for children.

We believe all children deserve the opportunity to thrive. We also believe that when opportunity is shared equitably, everyone benefits. For children, opportunity includes the conditions and resources they need to grow up healthy and learn. This includes the resources available to their families, in the schools they attend and in the neighborhoods where they live.

The launch of the Child Opportunity Index 2.0 marks a new chapter in our mission. The Child Opportunity Index (COI) 2.0 measures neighborhood resources and conditions that matter for children's healthy development. COI 2.0 allows us, for the first time, to compare the level of opportunity that neighborhoods provide for children across the U.S. in a single metric.

We hope that the index provides our thriving community of users with the information they need to make a positive impact through further research, community conversations about equity, and actions to change policy and allocate resources to increase equitable access to opportunity for all children.

Introduction

The neighborhoods where children live, learn and play influence their later life outcomes, including their economic mobility, educational attainment and health. The Child Opportunity Index (COI) 2.0 measures neighborhood resources and conditions that matter for children's healthy development. COI 2.0 allows us, for the first time, to compare the level of opportunity that neighborhoods provide for children across the U.S. in a single metric.

COI 2.0 is unique and distinct from other opportunity indices in its focus on contemporary child-relevant neighborhood features. It offers a summary measure of the quality of neighborhoods children experience every day across the U.S. COI 2.0 includes 29 indicators that measure neighborhood-based opportunities for children including but not limited to access and quality of early childhood education (ECE), high-quality schools, green space, healthy food, toxin-free environments, socioeconomic resources and more. The 29 indicators are grouped into three domains: education, health and environment and social and economic.

COI 2.0 is available for virtually all neighborhoods (census tracts) in the 50 U.S. states and Washington, D.C. for two time points, 2010 and 2015. It is accessible via an interactive web application that allows users to explore the COI 2.0 in their communities and across the U.S., as well as a downloadable database that provides a single, harmonized database of the composite index measures and individual indicators of child opportunity that comprise the index.

COI 2.0 is based on COI 1.0, which was jointly developed with the Kirwan Institute for the Study of Race and Ethnicity at Ohio State University and launched in 2014. Since then, the COI has drawn users from diverse sectors and communities and been used in numerous applications. Among the users are local service providers, community organizations, media, researchers, policymakers, planners and national equity-focused organizations. They have used the COI to increase awareness of equity, promote community discussions, target services and programs, better understand the connections between neighborhoods and health and inform needs assessments, resource allocation and policy development.

COI 2.0 allows users to answer questions such as: Which and where are the metropolitan areas and neighborhoods with the highest and lowest levels of child opportunity? What is the extent of inequality 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? COI 2.0 allows users to explore levels and inequalities of child opportunity across the country, or more specifically in their state, metro area, city or neighborhood.

While COI 2.0 builds on the previous 2014 release of the index (COI 1.0), its construction utilized additional, and in some cases new measures and distinct methodologies. COI 2.0 should therefore not be considered directly comparable to COI 1.0. Table 1 outlines key differences between the two versions. COI 1.0 included information on 19 indicators covering three domains of opportunity.

For COI 2.0, we revisited the indicators and methods for constructing the index. We increased the number of indicators to 29. While COI 1.0 was only available for census tracts in the 100 largest metro areas, COI 2.0 is available for (virtually) all census tracts in the U.S. (>72,000 tracts). While COI 1.0 was available for 2010, COI 2.0 is available for 2010 and 2015. For COI 2.0, individual component indicators and the composite index itself are comparable over time and across neighborhoods, allowing users to study change in opportunity in their communities over time and to compare their community to their city, metro area, state or anywhere in the country. Finally, we modified the methodologies for constructing specific indicators and for combining indicators into three domains and overall scores. In

Table 1. Key differences between COI 1.0 and COI 2.0

COI 1.0	COI 2.0
• 19 indicators	• 29 indicators
• 47,000 census tracts (100 largest metro areas)	• >72,000 census tracts
• 2010 data	• 2010 and 2015 data
• Data comparable within metro areas	• Data comparable within and across metro areas, and over time
• All indicators weighted equally when combined into the index.	• Indicators have individual, varying weights based on how strongly they predict health and economic out-

COI 1.0, each indicator in the index was weighted equally, while in COI 2.0 each indicator has an individual weight based on how strongly the indicator predicts health and economic outcomes, which improves the predictive validity of the index.

In the following sections, we provide an overview of the 29 indicators included in the index for each of the three domains (education, health and environment, social and economic) and outline the scientific rationale for including them. Then, we discuss the methodology used for weighting and combining the indicators into the overall index. We present initial evidence of how well the COI, as a measure of children's neighborhood-based opportunities, predicts later life outcomes (i.e., the predictive validity of the index), by correlating domain and overall index scores with health and economic outcomes. Finally, we provide further information on data sources and methods for constructing the indicators in the Appendices.

Domains and Indicators

Neighborhood factors shape children's access to resources and experiences that promote healthy development, and children's exposure to risks that can hinder development. Neighborhoods are multi-dimensional, influencing child development through numerous causal pathways. Drawing on our previous work[1-4] and related work on neighborhoods and child and human development,[5; 6] we group neighborhood features into three domains, or pathways, through which neighborhood environments 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 and exposure to environmental toxins.

The selection of indicators for each domain is grounded in a comprehensive, cross-disciplinary literature review, and is further informed by our work with users of COI 1.0. Our review has identified key features of neighborhoods relevant to healthy child development, identified the mechanisms through which these features affect children's outcomes and identified high-quality research evidence supporting their inclusion in the index.

Furthermore, we have conducted extensive analyses of the predictive validity of domain and individual component indicators, which directly informed how we constructed the overall index. The weight that each indicator and domain is given when constructing the overall index is a function of how strongly the indicators and domains predict long-term economic and health outcomes.

When selecting measures for a given neighborhood feature thus identified, we obtain direct measures of the specific features when available (rather than a proxy measure) and ensure that we can measure the feature at an appropriate geographic scale. For example, several studies identify proximity to hazardous waste dump sites (Superfund sites) as a detrimental neighborhood factor.[7-9] Research shows that pollutants emitted from these sites exert a negative effect on children's health and educational attainment. However, this effect depends on proximity: children residing within a 2-mile radius of an uncleaned Superfund sites are particularly affected.[8] To capture the appropriate geographic scale for this indicator, we measured exposure to uncleaned Superfund sites as the number of uncleaned Superfund sites within two miles of a census block centroid, and then averaged the number of Superfund sites across blocks within a given census tract, using the block-level child population as weights.

A challenge to measuring neighborhood features relevant for children's wellbeing is that empirically validated and reliable measures are not always available, either because the required data do not exist or are too costly to acquire. Additionally, because COI 2.0 is constructed for neighborhoods across the entire U.S., there are often measures that are not broadly available, i.e., they may only be available for a subset of U.S. neighborhoods. Furthermore, studies yield conflicting and inconclusive results about the causal influence of certain neighborhood factors on child outcomes.[10-12] In these situations, we are left with a choice of either omitting certain neighborhood features or relying on proxy measures.

For instance, in trying to measure the quality of children's neighborhood-based educational opportunities, we lack detailed measures of school- and classroom-level processes in the local schools. We do not have detailed measures of the quality of instruction and instructional classroom environments based, for example, on teacher observations for all schools in the U.S. To still capture the quality of local schools in a meaningful way, we included measures of third grade reading and math proficiency, as well as an indicator of school poverty. Neither measure directly captures specific classroom or school-level processes; rather, they capture the outcomes of student learning (proficiency measures). School poverty, i.e., the percentage of students eligible for free or reduced-price lunches (FRPL), is a proxy measure of the quality of educational resources and contexts that has been shown to be highly predictive of student outcomes.[13-15]

What distinguishes our approach from that of others is the COI's focus on contemporary neighborhood features affecting children. Unlike the Opportunity Atlas[16; 17] indicators, for example, which measure the long-term effect of neighborhood conditions present twenty years ago, the COI focuses on contemporary features of neighborhoods that shape children's experiences today, and that are linked to healthy child development based on previous research. The COI capitalizes on a wide array of neighborhood level measures available in open-source datasets, including the Opportunity Atlas data, which we used as an important set of complementary reference measures to improve and illustrate the predictive validity of the COI. Our empirical analyses show that the overall COI is strongly correlated with measures of intergenerational economic mobility from the Opportunity Atlas and measures of health and life expectancy.

In the following, we briefly review the domains and indicators included in COI 2.0 and the mechanisms through which they affect children's outcomes. We provide references to empirical studies supporting the inclusion of the indicators into the index. Table 2 below lists all indicators and provides brief definitions of each.

Education domain

The neighborhood where a child lives plays a crucial role in shaping their educational experiences throughout childhood. With a predominantly neighborhood-based public school feeder system in the U.S., a child's neighborhood typically determines where they attend primary and often secondary school. Beyond schooling, neighborhood access to early childhood care and educational experiences, broader neighborhood educational contextual factors and local institutions all play a role in shaping a child's educational opportunities. Recent research provides evidence that the effect of neighborhood factors varies depending on child age,[16; 18] informed the inclusion of indicators that measure neighborhood-based educational opportunities throughout the different stages of childhood. COI 2.0 captures key measures related to schooling, neighborhood contextual factors and local institutions (early childhood, elementary and secondary and postsecondary) that together (cumulatively) reflect a child's neighborhood educational opportunities. Indicators are grouped into four subdomains: early childhood education, elementary education, secondary and postsecondary education and educational and social resources.

The early childhood and elementary education subdomains are informed by research indicating that early childhood neighborhood contexts exert larger effects on adult outcomes than neighborhood contexts in middle childhood and adolescence.[16] Experiences in early care and educational settings, access to early care and education programs and peer early education attendance patterns are all factors that influence children's early neighborhood-based early educational opportunities. Indicators of proximity to licensed center-based care and high-quality center-based care are included to represent the presence of early care and education settings that have the potential to promote positive development, with lasting effects into adulthood.[19] Peer preschool enrollment rates are included, as children with higher neighborhood peer enrollment rates are more likely to attend publicly accessible preschool programs.[20]

Third grade math and reading school average proficiency scores are included as a broad, cumulative measure of how children's neighborhood-based early educational opportunities affect their early academic achievement. This measure reflects not only elementary school experiences, but also the broader set of early educational opportunities that children experience through their family settings and through local institutions (e.g., libraries, after school programs, youth/community programs). Research finds that school district socioeconomic status (SES) is highly correlated with third grade test scores, implying that early opportunities that shape learning (e.g., family resources, quality of early care and education, neighborhood conditions) are strongly associated with a community's SES level. School poverty is also included in the education domain, as a marker of the SES composition of neighborhood schools (learning contexts), given the demonstrated links between school SES levels and academic outcomes.[21]

However, recent research[18] suggests that the relationship between community/school district poverty and middle childhood academic outcomes is positive, but much weaker than the strong relationship found between poverty and early academic outcomes.[22] This finding provides evidence that the neighborhood-based factors shaping early learning may be different than the mix of factors shaping later learning and that the quality of early learning experiences in the neighborhood are not determinative of the quality of later childhood learning experiences.[18]

The education domain also seeks to capture neighborhood-level education-related resources including several indicators pertaining to youth and adult educational attainment, which are known to influence child outcomes. Researchers have documented a positive relationship between healthy child development and exposure to adults with higher educational attainment and exposure to community norms that support educational attainment. Living in neighborhoods with a higher educational attainment (i.e., higher percentage of adults enrolled in college or with college degrees) gives students higher expectations, influencing postsecondary pathways (e.g., college attendance).[23-25] Moreover, social networks are one of the primary channels through which job seekers find employment. Neighborhoods with more adults that have college degrees offer social networks that can lead to better jobs.[26]

Table 2. COI 2.0 indicators and domains

	INDICATOR	DESCRIPTION (SOURCE)
EDUCATION	Early childhood education (ECE)	
	ECE centers	Number of ECE centers within a 5-mile radius (own data collection from state and federal sources)
	High-quality ECE centers	Number of NAEYC accredited centers within a 5-mile radius (own data collection from state and federal sources)
	ECE enrollment	Percent 3- and 4-year-olds enrolled in nursery school, preschool or kindergarten (ACS)
	Elementary education	
	Third grade reading proficiency	Percent third graders scoring proficient on standardized reading tests (EDFacts, GS and SEDA)
	Third grade math proficiency	Percent third graders scoring proficient on standardized math tests (EDFacts, GS and SEDA)
	Secondary and postsecondary education	
	High school graduation rate	Percent ninth graders graduating from high school on time (EDFacts and GS)
	Advanced Placement (AP) course enrollment	Ratio of students enrolled in at least one AP course to the number of 11th and 12th graders (CRDC)
	College enrollment in nearby institutions	Percent 18-24 year-olds enrolled in college within 25-mile radius (ACS)
	Educational and social resources	
	School poverty	Percent students in elementary schools eligible for free or reduced-price lunches, reversed (NCES CCD)
	Teacher experience	Percent teachers in their first and second year, reversed (CRDC)
	Adult educational attainment	Percent adults ages 25 and over with a college degree or higher (ACS)
HEALTH & ENVIRONMENT	Healthy environments	
	Access to healthy food	Percent households without a car located further than a half-mile from the nearest super-market, reversed (USDA)
	Access to green space	Percent impenetrable surface areas such as rooftops, roads or parking lots, reversed (CDC)
	Walkability	EPA Walkability Index (EPA)
	Housing vacancy rate	Percent housing units that are vacant, reversed (ACS)
	Toxic exposures	
	Hazardous waste dump sites	Average number of Superfund sites within a 2-mile radius, reversed (EPA)
	Industrial pollutants in air, water or soil	Index of toxic chemicals released by industrial facilities, reversed (EPA)
	Airborne microparticles	Mean estimated microparticle (PM2.5) concentration, reversed (CDC)
	Ozone concentration	Mean estimated 8-hour average ozone concentration, reversed (EPA)
	Extreme heat exposure	Summer days with maximum temperature above 90F, reversed (CDC)
	Health resources	
	Health insurance coverage	Percent individuals ages 0-64 with health insurance coverage (ACS)
SOCIAL & ECONOMIC	Economic opportunities	
	Employment rate	Percent adults ages 25-54 who are employed (ACS)
	Commute duration	Percent workers commuting more than one hour one way, reversed (ACS)
	Economic and social resources	
	Poverty rate[a]	Percent individuals living in households with incomes below 100% of the federal poverty threshold, reversed (ACS)
	Public assistance rate[a]	Percent households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assistance Program, reversed (ACS)
	Homeownership rate[a]	Percent owner-occupied housing units (ACS)
	High-skill employment[a]	Percent individuals ages 16 and over employed in management, business, financial, com-puter, engineering, science, education, legal, community service, health care practitioner, health technology, arts and media occupations (ACS)
	Median household income[a]	Median income of all households (ACS)
	Single-headed households	Percent family households that are single-parent headed, reversed (ACS)

Notes: We reverse some of the indicators when combining them into the index, e.g., the poverty rate, so that more of that indicator always means more opportunity. [a]These five indicators are combined into an economic resource index.

In contrast, high neighborhood levels of high school non-completers poses a risk factor for children. Neighborhood high school non-completion rates influence children's educational attainment and other developmental outcomes.[27-29] COI 2.0 captures adult educational attainment, teacher experience, college access and high school completion measures.

Health and environment domain

Research from across the social sciences and medicine has established robust links between health and environmental features of neighborhoods and children's health outcomes. These features affect children's health in utero, childhood and adulthood, and through those health effects, likely influence educational achievement and socioeconomic outcomes.[30-32] Neighborhood health and environmental factors that do not directly affect child health can also have indirect effects on children through their impact on parents or primary caregivers. For COI 2.0 we grouped neighborhood health and environmental features into three subdomains: healthy environments, toxic exposures and health resources.

The healthy environments subdomain captures features of the neighborhood environment that are primarily linked to health behaviors and health outcomes, particularly for children. The features include both factors that promote healthy child development and factors that pose risks to healthy child development. These include: (1) access to healthy food options, which has been linked to higher nutritional quality diets, healthier body mass index and increased food security;[33-35] (2) access to green space, linked to increased physical activity, reduced stress and improved mental wellbeing;[32; 36-38] (3) walkability, also linked to increased physical activity[31; 39; 40] and (4) vacant housing which has been linked to reduced feelings of safety and increased crime rates.[41-43]

The toxic exposures subdomain captures the risk of physical exposure to environmental toxins. We included three measures of airborne toxic exposures: airborne microparticles (PM2.5), ozone concentration and an index of industrial pollutants that have been linked to adverse neurodevelopmental and birth outcomes, respiratory and other chronic illness and long-term adverse health and education outcomes.[44-55] Furthermore, this domain includes measures of proximity to hazardous waste dump sites (Superfund sites), areas which affect residents through both air and water release, and increase the risk of adverse birth outcomes as well as impaired long-term health and education outcomes.[7-9; 56] Finally, we have included a measure of extreme heat exposure, which has been linked to adverse birth outcomes, heat stress, heat-related illness and death in children, as well as reduced academic achievement.[57-61]

The health resource subdomain includes neighborhood health insurance coverage rates among the population under age 65 to capture levels of health care access. Health insurance coverage is a marker of health care access since it lowers the costs and increases the demand for health care, but large expansions of health insurance coverage also affect providers, who may increase the provision and quality of services to meet the increased demand.[62-64]

Social and economic domain

The social and economic domain includes two subdomains: economic opportunities and economic and social resources.

The first subdomain of the social and economic domain is economic opportunities. Access to economic opportunities is one of the many factors that influences the economic outcomes of a neighborhood's residents.[65-68] In the economic opportunities subdomain, COI 2.0 captures access to jobs both as the percentage of adults who are employed, as well as how far neighborhood residents have to commute to gauge spatial proximity to employment.

Secondly, the economic and social resources available within neighborhoods are indicative of both access to and perceptions of opportunity, and play a powerful role in shaping individual choices and opportunities for advancement.[65; 69-71] COI 2.0 captures neighborhood economic and social resources through an economic resources index as well as rates of single-headed households. The economic resource index combines several indicators measuring different facets of household wealth and income (poverty rate, public assistance rate, high-skill employment, median household income and homeownership rate). Neighborhoods high in economic resources have more financial resources to invest into amenities that depend on local funding, such as schools, parks and after-school programs, and have greater purchasing power amongst residents, attracting private business and service providers.[27; 69; 72; 73]

Moreover, neighborhood economic resources are closely related to social resources that influence child development and later educational, economic and health outcomes. For example, the employment status of adults in a neighborhood can influence the future employment status of children, both directly (e.g., adults creating neighborhood-based social and economic networks that connect youth to educational and employment opportunities), and indirectly (e.g., adult employment conditions and attitudes can shape or reinforce youth aspirations and decision-making).[70-72]

Additionally, 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).[73-75] Potential explanations for this effect are reduced availability of parental supervision and weakened informal social control, as well as fewer (male) role models.[74; 76-78]

DATA AND METHODS

The COI is comprised of indicators measured on different scales, such as counts, percentages or U.S. dollars. To combine indicators measured on different scales into an index, the raw values of each indicator were standardized using the common approach of z-score transformation (see "Standardization" below).

Next, we combined individual indicators into the three domains (education, health and environment, social and economic). When combining indicators into domains, we used weights that reflect the strength of the association between each indicator and related health and socioeconomic outcomes. The domain scores were then aggregated using the same weighting approach into an overall score.

All component indicators were measured at the census tract level using constant 2010 census tract definitions for the two COI 2.0 time periods (2010 and 2015).[79] Census tracts correspond to the Census Bureau's definition of a neighborhood. They are drawn to cover an area with about 4,000 residents and their boundaries generally follow visible or identifiable local boundaries, such as such as intersections, roadways, streams or other bodies of water and boundaries of administrative entities (e.g., cities, towns and counties). Indicators derived from census tract-level data that used other TIGER/Line vintages more recent than 2010 were crosswalked to 2010 census tract definitions (see Appendix 3).

The COI and all component indicators are available for all census tracts in the 50 U.S. states and Washington, D.C. and two time periods that we label 2010 and 2015. The exact year, or range of years, that a given indicator measures varies from indicator to indicator. Whenever possible we obtained either single-year data for 2010 and for 2015, or multi-year averages with a mid-point in 2010 and 2015 (e.g., American Community Survey (ACS) data from 2008-2012 has a midpoint year of 2010). Two indicators are measured as time-constant variables, i.e., they have identical values in 2010 and 2015, access to greenspace and walkability.

For each indicator, the source data was obtained from identical sources for both years and processed based on constant definitions using identical protocols. Indicator and index values are therefore comparable across census tracts and over time.

Data sources for COI component indicators

The indicators comprising the COI were drawn from numerous public sources, including the Census Bureau, National Center for Health Statistics (NCHS), Department of Education, the Environmental Protection Agency (EPA) and others. The only proprietary data used was a school-level dataset with math and reading achievement scores, as well as high school graduation data licensed from GreatSchools (GS).

Table 2 lists the indicators included and Appendix 1 contains complete indicator definitions, sources and methods. Appendix 2 contains a more detailed discussion of how we calculated census tract-level indicators from school-level data for education indicators.

Data used for calculating weights and validation analyses

To calculate indicator weights and validate the index, we relied on three data sources that include census tract-level measures of adult health and economic outcomes: indicators of intergenerational social mobility from the Opportunity Atlas, health indicators from the RWJF 500 Cities Project and life expectancy data from the Centers for Disease Control and Prevention (CDC).

The Opportunity Atlas includes indicators of intergenerational mobility estimated at the census tract level.[16; 17] Based on Census Bureau data linked to IRS tax records, Chetty et al. estimated the effect of growing up in each U.S. census tract on adult economic outcomes, such as household income or residence in a low poverty neighborhood as an adult. By construction, the measures capture long-term effects of neighborhood conditions that were present in the 1990s. We selected the following indicators for calculating 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

Each of these indicators focuses on the long-term outcomes of individuals, measured at age 35, who grew up in households with an income at the median of the parent household income distribution.

For validation analyses, we selected the following indicators:

- Mean household income rank at age 35 for children with parents at the 25th percentile of the parent income distribution
- Probability of entering the top 20% of the household income distribution at age 35 for children with parents at the 25th percentile of the parent income distribution
- Probability of living in a low poverty census tract at age 35 for children with parents at the 25th percentile of the parent income distribution

These measures focus on the long-term outcomes of individuals at age 35 who grew up in household with an income at the 25th percentile of the parent household income distribution. They more closely reflect opportunities for low income children, were used by Chetty et al. for validation analyses[16] and by a housing mobility intervention targeting low income households to classify neighborhoods in terms of opportunity..[80]

The RWJF-CDC 500 Cities project provides census tract-level estimates of health conditions based on spatially interpolated survey data.[81; 82] We used the following measures of prevalence of the following conditions among the population for calculating weights:

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

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

For validation analyses, we used measures of prevalence of the following conditions among the population ages 18 and older:

- Having ever been diagnosed with or currently having asthma
- Having ever been diagnosed with diabetes
- Having ever been diagnosed with angina or coronary heart disease
- Having smoked in their lifetime and currently smoke every day or some days
- Not engaging in leisure time physical activity in the past month
- Having a body mass index equal to or greater than 30 (obesity)

Finally, for validation analyses only, we used census tract-level data on life expectancy from the CDC U.S. Small-area Life Expectancy Project (USALEEP) for the period from 2010-2015.[83; 84]

Census tracts included

We deleted tracts that were either fully covered by water or had missing data on more than 50% of indicators in any of the three domains. That is, we only reported data for tracts that had at least some land area and non-missing data for half or more of the indicators in each domain. This resulted in a total loss of about 850 tracts in 2010 and 2015, or a loss of about 1.2% of tracts in each year. The final number of census tracts in 2010 (2015) was 72,195 (72,213). The bulk of missing data was concentrated among these excluded tracts; therefore, we have nearly complete information on all indicators among the included tracts.

Outliers

Some variables had skewed distributions (i.e., a small number of cases with values much higher than the majority of the data points), which produced large z-scores that could in individual cases have a disproportionate influence on the resulting domain and overall index scores for a particular neighborhood (see next section). To reduce the impact of extreme outliers, we converted four indicators to a natural log scale: ECE centers, high-quality ECE centers, hazardous waste dump sites and industrial pollutants in air, water or soil. Furthermore, we bottom- and top-coded each indicator at the 1st and 99th percentiles within each period. In consequence, z-scores across all component indicators were bounded in the range of −5 to +5 (see Table 3 below), though these extremes were reached in only very few cases.

Standardization

To combine indicators measured using different units of measurement into an index, some form of standardization is required. Standardization ensured that all indicators were measured on a common scale. Specifically, we performed the common z-score standardization for each indicator, using the following formula:

(1) $z_{ijt} = (x_{ijt} - m_{j,t=2010}) / sd_{j,t=2010}$,

where *i* denotes census tracts, from $i = 1,...,N_t$, where N_t is the total number of census tracts, *j* represents the indicator, from $j = 1, ..., 29$, and *t* represents time period, with $t = 2010, 2015$.

Following the formula above, for each indicator we first calculated the 2010 mean value ($m_{j,t=2010}$) and the standard deviation ($sd_{j,t=2010}$) value, and then subtracted the 2010 mean from the raw tract-level indicator values (x_{ijt}) and divided this difference by the 2010 standard deviation for that indicator. Mean and standard deviation were calculated using the census tract-level child population counts (number of children ages 0-17) from the 2010 Decennial Census as weights.

After standardization, each indicator was measured on the same scale in both time periods, i.e., in 2010 standard deviations. Consequently, indicator z-scores, domain scores and the overall

COI scores can be compared across tracts and over time.

To ensure that higher values always indicate more opportunity, we standardized the directionality of each indicator by multiplying the standardized score of some indicators by -1. Those indicators are labeled as "reversed" in Table 2.

Tables 3 and 4 below report arithmetic means, standard deviations, minima and maxima for the raw (Table 3) and standardized (Table 4) indicator scores for both periods.

Economic resource index

We selected five indicators of neighborhood economic resources, each capturing a different facet of economic resources: neighborhood poverty rate, public assistance rate, homeownership rate, high-skill employment and median household income. Each of these variables are highly intercorrelated. Since they reflect different facets of neighborhood economic resources, rather than adding them separately we decided to join them into a single index combining information on economic resources from all five indicators.

Specifically, we used 2010 data for these five indicators and standardized each indicator using the procedure described above. We then performed principal component analysis, weighting by census tract-level child population data from the 2010 Decennial Census (children ages 0-17). We then used the 2010 component loadings to calculate weighted averages of the five standardized indicators in 2010 and 2015. All five variables are highly intercorrelated (alpha = 0.87) and we found that the first component explains nearly 71% of their total variation.

Calculating domain scores and overall index scores

Next, the standardized indicator and the economic resource index values were aggregated within the three domains to calculate domain scores. For COI 1.0, we assigned each indicator equal weight when constructing domain and overall index scores. From a predictive validity perspective, the equal weights approach implies that each variable is an equally important determinant of children's outcomes. It also implies that a deficit in an influential variable, e.g., neighborhood poverty rate, can be fully canceled out by an equally sized advantage in a less influential variable.

The equal weights approach is optimal when there is fundamental disagreement about the importance of different variables, because it minimizes disagreement between indices constructed using diametrically opposed weighting schemes.[85] In other words, if even diametrically opposed weighting schemes are equally plausible, equal weighting is the least worst solution.

An alternative approach is to specify 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 receives greater weight in the overall index score. However, scientifically accurate empirical weights are difficult to obtain for all indicators. It would require estimates of the average causal effect of each indicator, as defined and operationalized for the index, on the same outcome, for the same time period, estimated using a representative sample of the population.

Our weighting scheme for the COI 2.0 strikes a middle ground between both approaches. We combined empirical weights with unit weights that were constant across indicators. To obtain empirical weights, we used two indicators of intergenerational mobility from the Opportunity Atlas and two health indicators from the 500 Cities Project. We estimated the unconditional association between each COI component indicator and four outcome indicators and then calculated an average association between the indicator and outcomes. These averaged associations were then combined with a domain-specific constant, inflating the empirical weight of weakly associated indicators and shrinking the weight of strongly associated indicators.

Specifically, we standardized each of the four outcomes using their respective national means and standard deviations, weighting each tract by the child population, taken from 2010 Decennial Census. We then regressed each of the standardized outcomes on each of the standardized COI component indicators using the following specification, again using the 2010 child population as weights. The regressions were estimated using 2010 COI indicator values and the respective year and all tracts for which outcome data are available for.

(2) $y_{ik} = a_k + r_{jk} X_{i,t=2010} + e_{ik}$,

Where *k* indicates outcomes from *k* = 1,...,4, *j* indicates indicators included in COI 2.0 from *j* = 1,...,29 and *i* indicates census tracts.

We estimated the (Pearson) correlation coefficient, r_{jk} in equation 2, for every combination of indicator *j* and outcome *k*. Then, for each indicator *j*, we averaged the four estimated correlations and computed to obtain $r_{j,av}$, the averaged correlation coefficient for indicator *j*.

The correlations are imperfect measures of the association between neighborhood features and children's outcomes. The outcome measures reflect exposure to past rather than contemporary neighborhood conditions. They are based on aggregate, rather than individual outcome data. Additionally, they are likely confounded by compositional difference across neighborhoods.

Because of these limitations, rather than relying on estimated correlations alone for constructing weights, we shrunk particularly large weights, and inflated particularly small weights, while preserving the ranking of indicators in terms of importance within each domain. This approach acts as a safeguard against biased weights in the tails of the weight distribution and it makes the resulting weights more robust to alternate ways of estimation.

Shrinking large weights was desirable as we suspected that some of the underlying correlations are upwardly biased, and therefore did not want to give too much leverage to any one indicator, given that some of the correlations may be upwardly biased. Similarly, inflating small weights was desirable if confounding drives correlations towards or even below zero, and because we did not want to nullify the impact any one indicator has on the index based on potentially downward biased correlations.

Empirically, we observed that some of the correlation coefficients (r_{jk}) and some of the averaged correlation coefficients ($r_{j,av}$) had a negative sign (see Results section). This was likely due to confounding by neighborhood economic status and/or urban versus suburban location. Further robustness checks indicated that, in each case, the averaged correlation coefficients ($r_{j,av}$) changed sign from negative to positive after we controlled for population density and the economic resource index. When constructing the weights, we set negative average correlation coefficients to zero.

For example, we observed that, unconditionally, the number of ECE centers is negatively associated with three of the four outcomes (see Table 5). This association is likely confounded: compared to suburban areas, urban areas have more ECE centers, fewer economic resources (e.g., higher poverty rates) and perform worse in terms of the outcomes considered. The negative correlation between economic resources and outcomes is likely strong enough to overwhelm any beneficial effects of access to ECE centers and turn the unconditional association between ECE centers and outcomes negative. Consistent with this explanation, we found that for indicators with unconditionally negatively associations with outcomes become positively associated with outcomes once economic resources and population density are conditioned on (i.e., once the economic resource index and a measure of population density (number of children per square kilometer) is added to Equation 1).

To combine the estimated correlations with unit weights, we proceeded as follows. We rescaled the average correlation coefficients ($r_{j,av}$) so that they sum up to the number of indicators within each domain:

(3) $r_j = r_{j,av} \times (D/S)$, where D equals the number of indicators in a given domain and S is the sum of the averaged $r_{j,av}$ in that domain.

We then combined the rescaled, averaged correlation coefficients with a (constant) unity-weight and divide the resulting sum by 2.

(4) $w_j = (r_j + 1)/2$

The resulting weights sum up to the number of indicators in each domain. A constant weight approach was justified if we had no prior knowledge about the magnitude of different weights, or potential biases in estimated weights.

Finally, we rescaled the weights so that they sum up to one within each domain:

(5) $w_{j,rescaled} = w_j / D_d$

Where D is the number of indicators in domain d. The number of indicators per domain differs, and this correction ensures that no single domain has outsized influence by construction due to the number of indicators included.

In tracts with missing data on any of the indicators, we rescaled the weights so that they still sum to the number of indicators with valid data for that tract. For robustness checks, we re-estimated the weights using alternate regression specifications, (a) controlling for population density and the economic resource index and (b) controlling for the aforementioned variables and county fixed effects, which did not substantially alter index scores or predictive validity (see below).

Computing domain and overall scores

After computing the weights, we multiplied each standardized indicator with the respective weights and summed across weighted indicators to calculate domain scores for both periods. We then repeated the same approach to calculate the overall COI scores, regressing the respective outcomes on the domain scores, calculating weights and computing the overall COI score.

We used 2010 national means and standard deviations to standardize the COI 2.0 indicators prior to combining them in an index. The resulting scores were then measured on a scale of standard deviations of the baseline (2010) year. For example, an increase in scores from 0.5 to 1 for a given tract between 2010 and 2010 corresponds to an increase of 0.5 baseline year standard deviations. Or, equivalently in relative terms, scores have doubled, grown by 100% or grown by a factor of 2.

Child Opportunity Levels

In addition to z-scores, another COI 2.0 metric released is the Child Opportunity Level, an ordered, categorical variable, sorting census tracts into 5 ordered categories labeled “very low,” “low,” “moderate,” “high” and “very high.” The cut points (percentiles) were calculated based on the 2015 distributions of the overall index or respective domain scores, weighted by child population (ages 0-17) counts from the 2017 ACS. As a result, each category includes exactly 20% of the U.S. child population.

Specifically, census tracts with scores at or below the 2015 20th percentile were sorted into the “very low” category. Tracts above the 20th and at or below the 40th 2015 percentile were classified as “low opportunity.” Tracts above the 40th and at or below the 60th 2015 percentile were classified as “moderate opportunity,” tracts above the 60th and at or below the 80th 2015 percentile were classified as “high opportunity” and tracts above the 80th 2015 percentile were classified as “very high opportunity.”

State- and metro-normed Child Opportunity Levels

To facilitate analyses highlighting local inequalities, we also published Child Opportunity Levels normed to metro areas and states. For example, for the metro-normed Child Opportunity Levels, we repeated the procedure described in the previous section using data just for tracts located in a given metro area. The resulting variable grouped neighborhoods within a given metro area into five categories, where the bottom 20% of neighborhoods within the metro were assigned to the "very low" opportunity category, the next 20% were assigned to the "low" opportunity category, and so forth. Because we used 2015 percentiles and child population weights, each level of the metro-normed Child Opportunity Levels variable contains exactly 20% of the child population in 2015 for a given metro area. Metro-normed Child Opportunity Levels are comparable across neighborhoods and over time within a given metro, but not across metro areas. For state-normed data, we followed the same procedure and the same interpretations and caveats apply.

Note that we did not use absolute thresholds based on opportunity score values to distinguish areas by levels of opportunity. For instance, if we used absolute thresholds we would classify areas as high opportunity if they had opportunity scores above a certain numeric value threshold that was deemed "high" (e.g., opportunity scores that are 2 or more standard deviation above average). Instead, the opportunity levels were based on ranking a given set of census tracts from lowest to highest and then sorting them into different categories. Because this approach was based on rankings and percentiles based on a specific set of tracts, the opportunity level of a given tract depends on the set of tracts included in the rankings. While they tend to be highly correlated, this is why nationally and metro- or state-standardized opportunity levels generally differ, because the former uses all tracts in the U.S. to define percentiles for opportunity levels, while the latter only uses tracts in a given metro area or state.

Choosing between state, metro and nationally standardized Child Opportunity Levels

Users who are only interested in exploring inequalities within a given state or metro area are encouraged to use the state- or metro-normed Child Opportunity Levels. States and metro areas across the U.S. differ in their levels of neighborhood opportunity. For example, for any given metro area, the metro-normed levels better capture the inequality within the metro than the state or nationally normed index. For metro areas that have high opportunity levels compared to other metro areas nation-wide, 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 when referenced to all tracts nationally. Nevertheless, even for users interested in exploring a specific metro area, the nationally normed levels may provide important contextual information because they can be compared across other neighborhoods in the U.S. while the metro-normed levels cannot. The same reasoning and caveats apply to using state versus nationally normed data.

Child Opportunity Scores

The National Child Opportunity Score ranks all >72,000 census tracts on a single metric from zero to 100. For each period, we ranked all tracts based on the COI overall score, and then divided tracts into 100 groups, from 1 to 100. We used population weights to calculate the exact cut points (percentiles) so that each of the groups includes 1% of children. The bottom 1% of tracts were assigned a score for 1, the next 1% were assigned a score of 2, and so forth, until the top 1% of neighborhoods, which were assigned a score of 100. Child Opportunity Scores are released as metro, state and nationally normed versions.

Table 3. Descriptive statistics for component indicators before Z-score transformation

INDICATOR		2010[c]				2015[c]			
		Mean	SD	Min	Max	Mean	SD	Min	Max
EDUCATION	ECE centers[a]	3.5	1.5	0.0	6.6	3.6	1.5	0.0	6.7
	High-quality ECE centers[a]	-3.8	7.1	-13.8	4.0	-3.5	7.0	-13.8	4.1
	ECE enrollment	49.2	25.5	0.0	100.0	48.1	25.2	0.0	100.0
	Third grade reading proficiency	203.3	37.7	100.3	310.5	204.7	62.0	62.1	376.5
	Third grade math proficiency	225.1	33.9	123.5	306.6	223.0	64.2	67.6	391.9
	High school graduation rate	75.3	15.8	26.6	98.0	79.3	14.7	32.3	98.0
	AP course enrollment	0.3	0.2	0.0	1.1	0.4	0.2	0.0	1.0
	College enrollment in nearby institutions	40.5	10.4	14.5	72.3	40.5	10.3	13.7	70.6
	School poverty	52.7	25.7	1.9	97.2	56.0	25.8	4.4	100.0
	Teacher experience	9.7	8.2	0.0	41.3	12.9	9.2	0.0	51.7
	Adult educational attainment	26.7	17.7	2.8	77.8	29.1	18.4	3.4	79.9
HEALTH & ENVIRONMENT	Access to healthy food	4.0	4.7	0.0	26.4	4.0	4.6	0.0	25.9
	Access to green space[d]	26.5	22.8	0.2	84.9	26.5	22.7	0.2	84.9
	Walkability [d]	8.9	3.7	2.4	16.9	8.9	3.6	2.4	16.9
	Housing vacancy rate	8.1	5.6	0.0	28.5	7.6	5.4	0.0	27.6
	Hazardous waste dump sites[a]	-12.9	3.2	-13.8	0.0	-13.1	3.0	-13.8	0.0
	Industrial pollutants in air, water or soil[a]	6.6	2.8	-4.9	11.7	6.3	2.9	-5.1	11.5
	Airborne microparticles	9.7	1.8	5.5	13.5	9.6	1.8	5.0	14.1
	Ozone concentration	39.9	4.0	27.9	49.6	38.1	4.0	29.9	52.4
	Extreme heat exposure	48.3	45.6	0.0	142.3	42.9	39.2	0.0	134.7
	Health insurance coverage	82.8	10.4	51.6	98.5	87.6	8.4	60.4	99.2
SOCIAL & ECONOMIC	Employment rate	74.9	9.7	40.7	90.8	76.5	9.4	42.4	91.7
	Commute duration	8.4	6.9	0.0	34.6	9.1	7.4	0.1	36.7
	Poverty rate[b]	15.4	11.8	0.7	53.4	15.1	11.3	1.1	51.6
	Public assistance rate[b]	13.2	11.2	0	51.4	14.4	11.9	0.3	54.2
	Homeownership rate[b]	66.6	21.5	4.5	97.4	64.6	21.7	4.4	96.8
	High-skill employment[b]	33.7	14.5	7	71.7	35.2	14.9	7.9	73.2
	Median household income[b]	62,876	29,502	18,817	166,316	64,311	30,554	19,250	171,923
	Single-headed households	34.4	18.2	3.9	86.0	34.8	18.2	4.6	86.4

Source: Child Opportunity Index 2.0 Database, diversitydatakids.org.
Notes: SD = standard deviations. Min = Minimum. Max = Maximum. Statistics have been calculated using the US child population (ages 0-17) from the ACS as weights. Prior to calculation of these statistics, each indicator has been top- and bottom-coded using the procedure described above. [a]These indicators have been converted to natural logs. [b]These indicators are combined into an economic resource index. [c]The exact year, or range of years, that a given indicator is measured for varies from indicator to indicator. Whenever possible we have obtained either single-year data for 2010 and for 2015, or multi-year averages with a mid-point in 2010 and 2015. [d]These indicators are time-constant. Descriptive statistics vary across years because of the weights used vary across years.

Table 4. Descriptive statistics for component indicators after Z-score transformation

INDICATOR		2010[c]				2015[c]			
		Mean	SD	Min	Max	Mean	SD	Min	Max
EDUCATION	ECE centers	0.0	1.0	-2.3	2.0	0.1	1.0	-2.3	2.1
	High-quality ECE centers	0.0	1.0	-1.4	1.1	0.0	1.0	-1.4	1.1
	ECE enrollment	0.0	1.0	-1.9	2.0	0.0	1.0	-1.9	2.0
	Third grade reading proficiency	0.0	1.0	-2.7	2.8	0.0	1.6	-3.7	4.6
	Third grade math proficiency	0.0	1.0	-3.0	2.4	-0.1	1.9	-4.6	4.9
	High school graduation rate	0.0	1.0	-3.1	1.4	0.3	0.9	-2.7	1.4
	AP course enrollment	0.0	1.0	-1.5	3.3	0.2	1.0	-1.5	3.1
	College enrollment in nearby institutions	0.0	1.0	-2.5	3.1	0.0	1.0	-2.6	2.9
	School poverty	0.0	1.0	-1.7	2.0	-0.1	1.0	-1.8	1.9
	Teacher experience	0.0	1.0	-3.9	1.2	-0.4	1.1	-5.1	1.2
	Adult educational attainment	0.0	1.0	-1.4	2.9	0.1	1.0	-1.3	3.0
HEALTH & ENVIRONMENT	Access to healthy food	0.0	1.0	-4.8	0.9	0.0	1.0	-4.7	0.9
	Access to green space[b]	0.0	1.0	-2.6	1.2	0.0	1.0	-2.6	1.2
	Walkability[b]	0.0	1.0	-1.8	2.2	0.0	1.0	-1.8	2.2
	Housing vacancy rate	0.0	1.0	-3.6	1.4	0.1	0.9	-3.4	1.4
	Hazardous waste dump sites	0.0	1.0	-0.3	4.0	0.0	0.9	-0.3	4.0
	Industrial pollutants in air, water or soil	0.0	1.0	-4.0	1.8	-0.1	1.0	-4.1	1.7
	Airborne microparticles	0.0	1.0	-2.1	2.3	0.1	1.0	-2.4	2.6
	Ozone concentration	0.0	1.0	-2.4	3.0	0.4	1.0	-3.1	2.5
	Extreme heat exposure	0.0	1.0	-2.1	1.1	0.1	0.9	-1.9	1.1
	Health insurance coverage	0.0	1.0	-3.0	1.5	0.5	0.8	-2.1	1.6
SOCIAL & ECONOMIC	Employment rate	0.0	1.0	-3.5	1.6	0.2	1.0	-3.3	1.7
	Commute duration	0.0	1.0	-3.8	1.2	-0.1	1.1	-4.1	1.2
	Economic resource index	0.0	1.0	-3.4	2.4	0.0	1.0	-3.4	2.4
	Single-headed households	0.0	1.0	-2.8	1.7	0.0	1.0	-2.9	1.6

Source: Child Opportunity Index 2.0 Database, diversitydatakids.org.

Notes: SD = standard deviations. Min = Minimum. Max = Maximum. Statistics have been calculated using the US child population (ages 0-17) from the ACS as weights. [a]The exact year, or range of years, that a given indicator is measured for varies from indicator to indicator. Whenever possible we have obtained either single-year data for 2010 and for 2015, or multi-year averages with a mid-point in 2010 and 2015. [b]These indicators are time-constant. Descriptive statistics vary across years because of the weights used vary across years.

RESULTS

The following sections summarize analyses illustrating the weighting strategy used to construct the COI 2.0 domain and overall scores, followed by analyses of the predictive validity of the overall index and its domains.

Weights

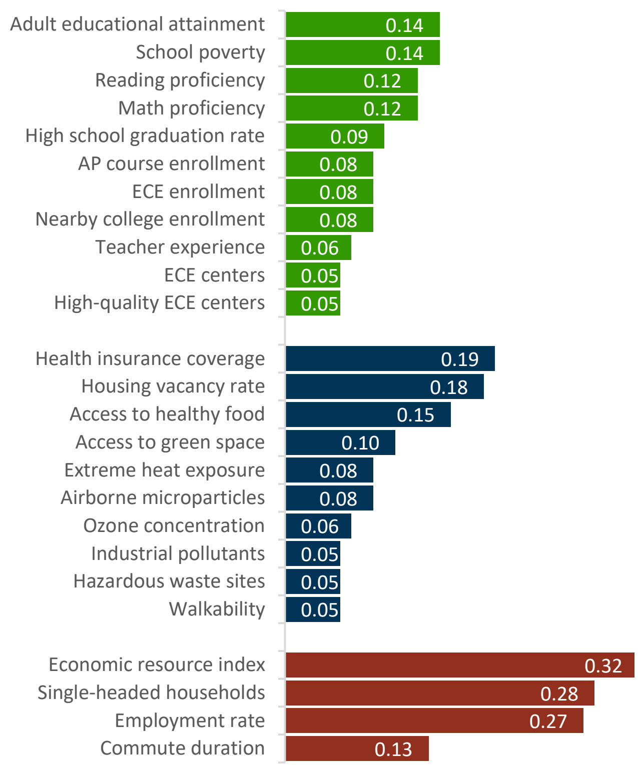
The weights used to aggregate indicators into domain scores are based on estimates of the bivariate association between each indicator and four different outcomes measured at the census tract level.

Table 5 reports the bivariate association (Pearson correlation coefficients) for all pairwise correlations between indicators and outcomes, the averaged correlation coefficient across the four outcomes, and the weights used to compute the domain scores. The median correlation coefficient across the 100 estimates is 0.27, and they range from a minimum of -0.19 to a maximum of 0.87. After averaging across correlation coefficients for each indicator, we observe a median averaged correlation coefficient of 0.25, a minimum of -0.10 and a maximum of 0.78.

Four of the 25 averaged correlation coefficients are (slightly) below zero. Negative correlations likely reflect confounding by economic status. For example, compared to suburban areas, urban areas are empirically more likely to have more ECE centers, fewer economic resources and they tend to perform worse on the outcomes considered here. The negative effect of economic resources can be strong enough to overwhelm smaller positive empirical effects. Consistent with this explanation, we observe that the average correlation changes sign from negative to positive when we estimate the correlations conditional on the economic resource index and population density. When constructing the weights, we set the negative averaged correlation coefficients ($r_{j,av}$ from above) to zero.

Figure 1 visualizes, for each domain, the distribution of indicator weights. Weights sum up to one within each domain. The weights of each domain score in the final index are then determined using the same approach we used for determining the weights of each indicator. Because it includes relatively fewer indicators, each indicator in the social and economic domain carries somewhat greater weight in the final index score compared to indicators from the other two domains. This is justified from a predictive validity standpoint, as each of the indicators is highly predictive of outcomes with the exception of commute duration (see Table 5). Within the education domain, adult

Figure 1. Indicator weights



Source: Child Opportunity Index 2.0 Database, diversitydatakids.org.

educational attainment and school poverty are most strongly predictive of the different outcomes, followed by the indicators for third grade reading and math proficiency. In the health and environment domain, health insurance coverage, vacant housing and access to healthy food have the largest effects on outcomes. In the social and economic domain, the economic resource index, single-headed households and the employment rate have the large effects. Compared to all other indicators, the economic resource index is most strongly associated with the different outcomes, highlighting the important role that socioeconomic gradients play in predicting inequality of opportunity across neighborhoods.

Table 5. Indicator-outcome correlations, average correlations across outcomes ($r_{j,av}$) and final weights

	INDICATOR	Household income rank	In low poverty neighborhood	Mental health	Physical health	Average columns 1-4, $r_{j,av}$	Final weight
EDUCATION	ECE centers	-0.14	0.09	-0.16	-0.20	-0.10	0.05
	High-quality ECE centers	-0.04	0.18	-0.06	-0.08	0.00	0.05
	ECE enrollment	0.20	0.24	0.30	0.26	0.25	0.08
	Third grade reading proficiency	0.54	0.49	0.61	0.58	0.56	0.12
	Third grade math proficiency	0.53	0.47	0.58	0.55	0.53	0.12
	High school graduation rate	0.36	0.24	0.30	0.29	0.30	0.09
	AP course enrollment	0.19	0.28	0.33	0.31	0.28	0.08
	College enrollment in nearby institutions	0.15	0.28	0.19	0.24	0.22	0.08
	School poverty	0.64	0.67	0.64	0.62	0.64	0.14
	Teacher experience	0.18	0.13	0.11	0.10	0.13	0.06
	Adult educational attainment	0.55	0.59	0.80	0.79	0.68	0.14
HEALTH & ENVIRONMENT	Access to healthy food	0.37	0.35	0.39	0.41	0.38	0.15
	Access to green space	0.21	0.06	0.28	0.28	0.21	0.10
	Walkability	-0.09	0.05	-0.12	-0.12	-0.07	0.05
	Housing vacancy rate	0.50	0.48	0.52	0.48	0.49	0.18
	Hazardous waste dump sites	-0.04	0.01	-0.07	-0.07	-0.04	0.05
	Industrial pollutants in air, water or soil	-0.05	0.10	-0.13	-0.14	-0.06	0.05
	Airborne microparticles	0.14	0.07	0.14	0.13	0.12	0.08
	Ozone concentration	0.08	0.06	0.05	0.00	0.05	0.06
	Extreme heat exposure	0.17	0.27	0.01	0.02	0.12	0.08
	Health insurance coverage	0.49	0.58	0.56	0.58	0.55	0.19
SOCIAL & ECONOMIC	Employment rate	0.52	0.57	0.66	0.66	0.60	0.27
	Commute duration	-0.02	-0.07	0.08	0.11	0.02	0.13
	Economic resource index	0.69	0.73	0.88	0.84	0.78	0.32
	Single-headed households	0.65	0.54	0.68	0.65	0.63	0.28

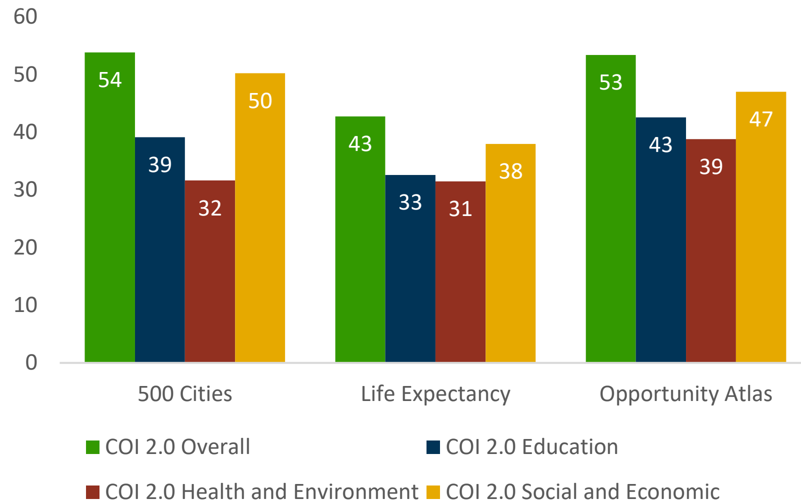
Source: Child Opportunity Index 2.0 Database, diversitydatakids.org.

Validity

Predictive validity

We now examine how well the COI 2.0, as a measure of children's neighborhood-based opportunities, predicts later life outcomes (i.e., the predictive validity of the COI 2.0 overall and its domain scores). For this analysis, we utilize additional neighborhood-level outcome data that was not used for constructing weights and, while the weights were calculated using 2010 COI data, we perform the analysis using 2015 COI data. The datasets and indicators for the validation analyses are further described in the data and methods section above.

Figure 2. Average percent variance explained by COI overall and domain scores



Sources: Child Opportunity Index 2.0 Database, Opportunity Atlas, CDC-RWJF 500 Cities Project and CDC USALEEP data.
Notes: The 500 Cities health outcomes include separate indicators for the percentage of adults with obesity, diabetes, coronary heart disease, smoking, limited physical activity and asthma. Opportunity Atlas outcomes: Household income rank, household in top 20% of household income distribution and household resides in low poverty neighborhood. All indicators are measured at the census tract level.

The datasets used for validation analyses are described in the Data and Methods section. We included the following outcomes in the analyses: (a) from the RWJF-CDC 500 Cities project, we included the percent adults ages 18 and older in a given neighborhood with the following characteristics: obesity, diabetes, smoking, coronary heart disease, limited physical activity and asthma; (b) from the Opportunity Atlas, we included household income rank, household income in the top 20% of the income distribution and residing in a low poverty neighborhood, each measured in adulthood for children with parents at the 25th percentile of the parent income distribution and (c) we used CDC data on life expectancy that is not part of either database and was not used for constructing the weights.

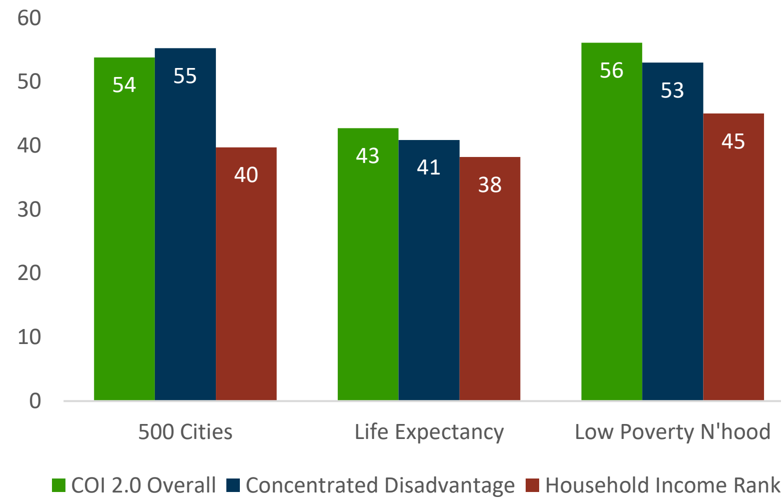
We estimated the percentage variance in the respective outcomes explained by the COI 2.0 overall and its domain scores. To simplify reporting, we average the percentage variance explained across outcomes for the different sets of outcomes from the Opportunity Atlas and the 500 Cities data.

Overall, COI 2.0 explains 43% of the variation in life expectancy, on average 50% of the variation in the Opportunity Atlas indicators and on average 54% of the variation in the 500 Cities health indicators (Figure 2).

The health and environment domain explains about a third of the variance across the different outcomes, while the social and economic domain explains between 38-50% of the variance. The overall index explains more variation than any of its constituent domains, indicating that the overall index is a stronger predictor of outcomes than any of the individual domain indices.

Next, we examine the predictive validity of the overall index in comparison to two other prominent indicators of neighborhood conditions used by sociologists and economists. The first is the index of Concentrated Disadvantage, which is a composite index used by sociologists to measure overall neighborhood conditions.[72; 86; 87] It is calculated from six socioeconomic status measures taken from the ACS: the neighborhood poverty rate, unemployment rate, public assistance rate, female single-headed households and two indicators of educational attainment (percent with less than a high school degree, percent with a college degree or more). The second measure is an indicator of intergenerational economic mobility taken from the Opportunity Atlas: the household income rank in adulthood of children with parents at the 25th percentile of the parent income distribution. This indicator has, for example, been used to classify neighborhoods for a large residential mobility program in Seattle, WA.[80]

Figure 3. Average percent variance explained by COI 2.0 overall score and two alternative metrics of neighborhood quality/opportunity



Sources: Child Opportunity Index 2.0 Database, Opportunity Atlas, CDC-RWJF 500 Cities Project and CDC USALEEP data.
Notes: The 500 Cities health outcomes include separate indicators for the percentage of adults with obesity, diabetes, coronary heart disease, smoking, limited physical activity and asthma. Opportunity Atlas outcomes: Household income rank, household in top 20% of household income distribution and household resides in low poverty neighborhood. All indicators are measured at the census tract level.

Because we are now examining household income rank as a predictor, we remove it from the set of outcome variables taken from the Opportunity Atlas in the validation analysis. We similarly remove household income in the top 20% of the income distribution. Therefore, the only variable left from the initial set of Opportunity Atlas indicators used for the predictive validity analyses is the probability to reside in a low poverty neighborhood.

COI 2.0 compares favorably to the other two indices. That concentrated disadvantage and the COI 2.0 are more highly correlated with the 500 Cities outcome data than household income rank is likely due to the fact that the algorithm used to construct the tract-level health estimates in the 500 Cities data includes census tract poverty rates as a predictor variable.[81] This induces a mechanical correlation between 500 Cities data and both the COI and Concentrated Disadvantage Index, which include the poverty rate and other variables highly correlated with poverty.

Content validity

While all three indicators are strong predictors of long-term outcomes, the COI 2.0 is the only metric that specifically focuses on neighborhood features that matter for children's experiences. The index of Concentrated Disadvantage has strong predictive validity, but lacks the content validity of an index specifically designed to capture the multitude of neighborhood features relevant to children. For example, it omits exposure to environmental toxins, which may or may not explain a large percentage of the variation in long-term outcomes. However, they are certainly harmful to children whenever they are exposed and, therefore, we capture them in the Child Opportunity Index.

Household income rank, taken from the Opportunity Atlas, captures neighborhood conditions in place 20 years ago that are predictive of household income in adulthood. As a measure of opportunity, it lacks a focus on contemporary conditions and leaves unspecified what it is about neighborhoods that generates high incomes in later life.

The Concentrated Disadvantage and the Opportunity Atlas indices do lend themselves to summative comparisons, e.g., which neighborhood is best from an outcome perspective. However, they include very limited information about the specific features of neighborhood that matter for children and lack contemporary information to guide place-based strategies, such as how well schools perform or whether neighborhoods suffer from environmental hazards.

Robustness of weights

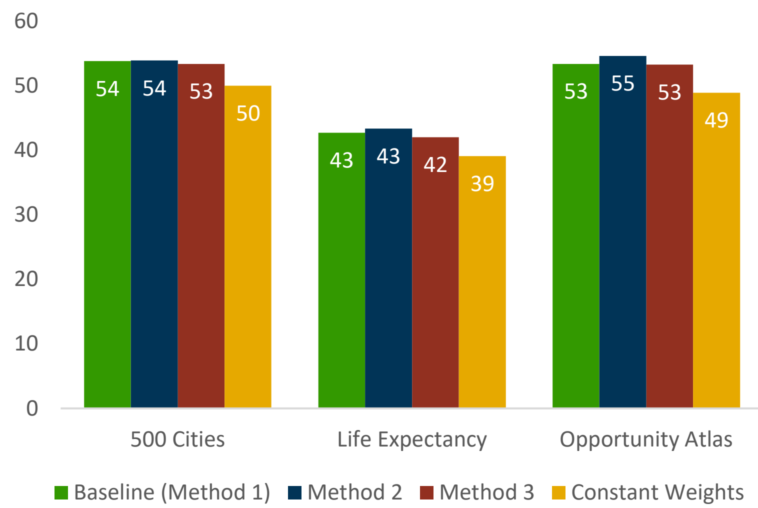
To test whether the COI 2.0 is sensitive to how the weights are estimated, we assembled the index using four different approaches: (1) the baseline approach described above, i.e., weights are based on bivariate correlations; (2) same as baseline, but the underlying correlations are estimated conditional on the economic resource index and population density; (3) same as (2) but additionally controlling for county fixed effects and (4) using constant weights for each indicator that sum to one within domains. The aggregate COI scores calculated using any of these approaches are very similar. They are correlated 0.98 or higher with the COI using our preferred (baseline) method. The weights obtained using these different approaches are also highly correlated with our preferred set of weights, with correlation coefficients of 0.74 or higher. These analyses indicate that using different specifications to estimate the weights underlying the domain and overall scores does not have a substantial impact on the overall index scores.

Finally, we examined how well COI indices that are constructed using different weights perform in terms of validity relative to the baseline method chosen for COI 2.0. We calculated the overall index using four different approaches: (1) the baseline approach chosen for COI 2.0 and described above, i.e., weights are based on bivariate correlations and then inflated/shrunk; (2) same as baseline, but weights are estimated conditional on the economic resource index and population density; (3) same as (2) but additionally controlling for county fixed effects and (4) using constant weights for each indicator that sum to one within domains (see Figure 4).

We observe that the indices constructed using weights derived from empirical data perform about 10% better than the index constructed using constant weights within each domain (Figure 4). This gain contributes to the favorable performance of COI 2.0 in relation to the other metrics examined in Figure 3.

There is virtually no difference in the predictive validity of the index calculated using different weights (methods 1, 2 and 3). Using more complex estimates of the association between the respective indicators and outcomes that condition on other factors does not change the overall index in a meaningful way. Therefore, we felt it justified to rely on method 1, which did improve the overall predictive validity of the index compared with method 4 (the simplest and most transparent approach), but was more transparent, i.e., bivariate correlations for constructing the weights, than methods 2 and 3.

Figure 4. Average percent variance explained by COI 2.0 overall score constructed using different weights



Sources: Child Opportunity Index 2.0 Database, Opportunity Atlas, CDC-RWJF 500 Cities Project and CDC USALEEP data.
Notes: The 500 Cities health outcomes include separate indicators for the percentage of adults with obesity, diabetes, coronary heart disease, smoking, limited physical activity and asthma. Opportunity Atlas outcomes: Household income rank, household in top 20% of household income distribution and household resides in low poverty neighborhood. All indicators are measured at the census tract level. Method 1 is the baseline approach chosen for COI 2.0, i.e., weights are constructed as shrunk, averaged bivariate correlations. Method (2) is as Method 1, but weights are estimated conditional on the economic resource index and population density. Method 3 is as Method 2, but additionally controlling for county fixed effects, and Method 4 uses constant weights for each indicator that sum to one within domains.

Strengths and limitations

The Child Opportunity Index 2.0 is a composite index based on 29 neighborhood level indicators covering three domains: education, health and environment, and social and economic. The index and its components are available for two time points 2010 and 2015 for virtually all census tracts in the U.S.

While the index compares favorably to other metrics for the purposes for which it was designed, it also has certain limitations. First, it lacks indicators on certain neighborhood features that previous research has identified as relevant for children but for which we were unable to gather comparable data at the census tract level. These include measure capturing neighborhood-level prevalence of violent crime,[43; 86] neighborhood social capital and collective efficacy,[88; 89] transportation costs and provider-side measures of access to health care (e.g., density of primary care physicians or pediatricians). Available measures we have identified were either too costly to acquire or not comparable over time. It is, however, possible to augment and recalculate the COI 2.0 with local measures of these indicators and we anticipate that the COI 2.0 will be used in this way to advance and improve the field's measures of neighborhood features. For example, for the Health Chicago 2.0 Report, we worked with the Chicago Department of Public Health to incorporate a measure of violent crime into the index.[90]

Furthermore, there are many challenges involved in calculating census tract-level reading and math proficiency estimates. While care has been taken to ensure comparability, not all potential threats to comparability can be addressed (see Appendix 2). However, it is encouraging that the measures we constructed are highly correlated with different set of outcomes (Figure 1 and Table 5).

The COI component indicators contain considerably more untapped information that could be used to improve predictive validity. For example, the weights we use to combine indicators into domain and aggregate scores are constant across all tracts and over time. The predictive performance of the index could be improved by adopting a more flexible method to estimate weights. For example, we could allow weights to have non-linear or interactive effects, and allows them to differ across neighborhoods. Exploratory analyses using cross-validation and Lasso models indicate that using more flexible weights improves the percent variance explained by the overall index considerably. However, adopting this approach will first require a better understanding of why different indicators vary across in their effects, e.g., across different states or metro areas.

Finally, we have relied on some of the best available evidence to date to determine which indicators to include and how to measure them, and we have used high quality, population representative data and high-quality neighborhood level outcome data to construct the weights. Nevertheless, the validity of the index could be further improved through better measurement of the specific features generating neighborhood opportunity and better estimates of their causal effects to address bias in the weights used to combine the indicators into an index. These are challenges we seek to address in future work.

Unlike other neighborhood metrics, COI 2.0 specifically focuses on the neighborhood conditions affecting children. Unlike the Opportunity Atlas indicators in particular, which capture the long-term effects of neighborhoods as they were 15-20 years ago, COI 2.0 focuses on contemporary neighborhood features linked to healthy child development by previous research.

COI 2.0 is strengthened by relying on data from the Opportunity Atlas and the 500 Cities Project to determine how much each indicator should count in the overall index and to illustrate its predictive validity. Given past uses of COI 1.0, we expect that COI 2.0 will be broadly used across many different sectors that require up-to-date information on neighborhood characteristics affecting child development, such as measuring neighborhood opportunity, quantifying inequity across neighborhoods and by race/ethnicity, strategic planning resource allocation, understanding the causes of neighborhood inequity and informing place-based and mobility interventions.

Appendix 1.1: Education domain indicators and sources

Early childhood education

Early childhood education centers

- Description: The number of ECE centers within a 5-mile radius.
- Definition: We counted the number of ECE centers within a 5-mile radius of each census block's centroid and averaged these counts across all blocks within a census tracts using the number of children ages 0-17 (2010 Decennial Census) as weights.
- Years: 2015-2019 (see notes).
- Scale: Averaged counts.
- Source: Child care licensing agency of each U.S. state, 2012-2013 and 2017-2019. National Center for Education Statistics (NCES) Common Core of Data (CCD), 2009-2010 and 2015-2016 school years. National Association for the Education of Young Children (NAEYC) Accredited Program Database, 2012-2013 and 2017-2019.
- Source geography: Center latitude and longitude.
- Notes: This indicator is based on three sources. For each U.S. state, a list of all licensed, center-based ECE service providers was created based on three data sources: state licensing agencies overseeing all early education and care programs, the NCES CCD for public school-based preschool programs and the NAEYC database, which included some additional centers not included in either of the previous two sources. We geocoded each center's address to obtained their latitude and longitude and removed duplicate entries. We then counted the number of centers within a 5-mile radius of each census block and averaged these counts across all blocks within a census tract using the number of children ages 0-17 (2010 Decennial Census) as weights.

High-quality early childhood education centers

- Description: The number of NAEYC accredited centers within a 5-mile radius.
- Definition: High-quality providers are defined as those programs that have received national accreditation from NAEYC. For this indicator, we counted the number of such centers within a 5-mile radius of each census block centroid and averaged these counts across all blocks within a census tracts using the number of children ages 0-17 (2010 Decennial Census) as weights.
- Years: 2012-2013 and 2017-2019 (see notes).
- Scale: Count.
- Source: NAEYC Accredited Program Database, 2012-2013 and 2017-2019.
- Source geography: Center latitude and longitude.
- Notes: See previous indicator.

Early childhood education enrollment

- Description: Percent 3- and 4-year-olds enrolled in nursery school, preschool or kindergarten.
- Definition: The number of children ages 3-4 years enrolled in school divided by the number of children ages 3-4 years, times 100.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Source: ACS 5-Year Summary Files, Table B14003, 2012 and 2017.[91]
- Source geography: Census tract.

Elementary education

Third grade reading proficiency

- Description: Percent third graders scoring proficient on standardized reading tests.
- Definition: The percentage of third graders tested who score proficient in standardized reading tests.
- Years: School years 2010/11 and 2015/16 for most states. See Appendix 2 for further details.
- Scale: Percent.
- Source: EDFacts, GS and Stanford Education Data Archive (SEDA).
- Source geography: Schools (point data).
- Notes: Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

Third grade math proficiency

- Description: Percent third graders scoring proficient on standardized math tests.
- Definition: The percentage of third graders tested who score proficient in standardized math tests.
- Years: School years 2010/11 and 2015/16 for most states. See Appendix 2 for further details.
- Scale: Percent.

- Source: EDFacts, GS and SEDA.
- Source geography: Schools (point data).
- Notes: Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

Secondary and postsecondary education

High school graduation rate

- Description: Percent ninth graders graduating from high school on time.
- Definition: 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.
- Years: School years 2010/11 and 2015/16.
- Scale: Percent.
- Source: EDFacts, GS and SEDA.
- Source geography: Schools (point data).
- Notes: Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

Advanced Placement (AP) course enrollment

- Description: Ratio of students enrolled in at least one AP course to the number of 11th and 12th graders.
- Definition: The number of students enrolled in at least one AP course divided by the number of students enrolled in grades 11 and 12.
- Years: School years 2011/12 and 2015/16.
- Scale: Ratio.
- Source: NCES OCRCD.
- Source geography: Schools (point data).
- Notes: Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

College enrollment in nearby institutions

- Description: Percent 18-24 year-olds enrolled in college within 25-mile radius.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Definition: The number of individuals ages 18-24 years enrolled in college or graduate school divided by the number of individuals ages 18-24 years, times 100, averaged across census tracts within a 25-mile radius of the home census tract's centroid.
- Source: ACS 5-Year Summary Files, Table B14001, 2012 and 2017.[92]
- Source geography: Census tract.
- Notes: To construct this indicator, we took a spatial average of the percent 18-24 year-olds enrolled in college across all tracts with centroids 25 miles or less away from the home tract's centroid. We weighted the percentage enrolled in each tract by the inverse of the distance between a given tract's centroid and the home tract centroid. Specifically, for each tract, we calculated a weight as $w_t = 1/d_t$, where d_t is the distance between tract t and the home tract. w_t was set to one for the home tract, because $1/d_t$ is not defined for a distance of zero. And, w_t was set to one for tracts within 1 mile of the home tract to prevent very close tracts from exercising an outsize influence on the weighted estimate. Lastly, w_t was rescaled so that it summed up to the total number of tracts within the 25-mile radius. To calculate the indicator, we multiplied the weight w_t with the percentage enrolled in tract t, and then averaged across products.

Educational and social resources

School poverty

- Description: Percent students in elementary schools eligible for FRPL.
- Definition: The number of students in grades 1 through 5 who are eligible for FRPL divided by the total number of student enrolled in grades 1 through 5, times 100.
- Years: School years 2010/11 and 2015/16.
- Scale: Percent.
- Source: NCES CCD.
- Source geography: Schools (point data).
- Notes: 2015/16 for Massachusetts was imputed. Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

Teacher experience

- Description: Percent teachers in their first and second year.
- Definition: The number of full-time teachers in their first or second year of teaching divided the number of full-time teachers.
- Years: School years 2011/12 and 2015/16.
- Scale: Percent.
- Source: NCES OCRCD.
- Source geography: Schools (point data).
- Notes: The 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.[93] Appendix 2 includes a detailed description of the sources and methods used to construct census tract-level estimates from school-level data.

Adult educational attainment

- Description: Percent adults ages 25 and over with a college degree or higher.
- Definition: The number of adults ages 25 years and older who have completed a Bachelor's degree or higher divided by the number of adults ages 25 years and older, times 100.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Source: ACS 5-Year Summary Files, 2012 and 2017, Table B15002.[94]
- Source geography: Census tract.

Appendix 1.2: Health and Environment Domain Indicators and Sources

Healthy environments

Access to healthy food

- Description: Percent households without a car located further than a half-mile from the nearest supermarket.
- Years: 2010, 2015
- Scale: Percent.
- Definition: The percentage of all housing units within the tract that lack a vehicle for transportation and are further than a half-mile away from the nearest supermarket.
- Source: USDA Food Access Research Atlas, United States Department of Agriculture Economic Research Service: Washington, D.C. Downloaded from https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/ on 12/06/18.[95]
- Source geography: Census tract.
- Notes: The data are taken from the USDA Food Access Research Atlas. The USDA used the following data and methods to create the indicators. Store data come from the 2010 and 2015 extractions from the Trade Dimensions TDLinx store directory. Supermarkets, supercenters and large grocery stores are included, while membership based warehouse stores (e.g., Costco) are excluded. Also excluded are drug stores, dollar stores and convenience stores. Data on vehicle access come from the 2006-2010 and 2010-2014 ACS. Data on housing units (vehicle availability) are downcast from the block (block group) level to half-kilometer grid cells. For each cell, the distance between geographic center of the cell and nearest supermarket is calculated. Then, the share of households without vehicles that are further than a half-mile from the nearest supermarket within a census tract (2010 definition) is computed.[96] Sources, definitions and methods are comparable across time.

Access to green space

- Description: Percent of impervious surface areas such as rooftops, roads or parking lots.
- Year: 2011. This indicator is available for one time period only and is therefore constant across the two periods for COI 2.0.
- Scale: Percent.
- Definition: Impervious surfaces are covered by impenetrable, artificial materials, such as brick, concrete and asphalt and includes structures such as roads, pavement, parking lots, buildings and roof tops. Access to green space is then defined as the inverse of the percentage of census tract covered by impervious surfaces. Specifically, we standardized the indicator (see Methods) and then multiplied the resulting z-scores by -1.
- Source CDC, https://ephtracking.cdc.gov/showIndicatorPages.action?selected-

ContentAreaAbbreviation=22&selectedIndicatorId=134&selectedMeasureId=, downloaded on 03/15/2019.[97]

- Source geography: Census tract.
- Notes: The source dataset is published by the CDC, based on data from the National Land Cover Database (NLCD). The NLCD data include total impervious area estimates for 30-meter pixels based on satellite imagery and other sources that were then aggregated to the census tract level by the CDC. The source data do not cover Alaska and Hawaii. Census tract data for these states are missing.

Walkability

- Description: EPA Walkability Index.
- Years: 2010-2012 (see notes). This indicator is available for one time period only and is therefore constant across the two periods for COI 2.0.
- Scale: Index units ranging from 1 (least walkable) to 20 (most walkable).
- Definition: The walkability of neighborhoods based on different features of the built environment and commuting mode choice that influence the choice to walk as a mode of transportation.
- Source: United States Environmental Protection Agency, Washington, D.C. Downloaded from ftp://new-ftp.epa.gov/EPADataCommons/OP/Natl_WI_SHP.zip on 02/06/19.[98]
- Source geography: Block group.
- Notes: For details on the construction of the index, see Thomas and Zeller[99] and Ramsey and Bell[100] for further information on sources from the EPA Smart Location Database. The walkability index was developed by the EPA and uses 2010 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 employment and house-hold mix receive a weight of 1/6, respectively. Source variables were gathered for somewhat different time points that represent conditions over the period from 2010 to 2012. We treat walkability as a time-constant variable, imputing the 2010-2012 value for both periods included in the COI. The index is published at the block group level. We aggregate block group data to the to the census tract level using the proportion of the block group group's land area of the total tract land area.

Housing vacancy rate

- Description: Percent of housing units that are vacant.
- Years: 2008-2012, 2013-2017
- Scale: Percent.
- Definition: The number of vacant housing units excluding housing units for seasonal, recreational and occasional use divided number of housing units, times 100.
- Source: ACS 5-Year Summary Files, 2012 and 2017, Table B25002.[101; 102]
- Source geography: Census tract.
- 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.

Toxic exposures

Hazardous waste dump sites

- Description: Average number of Superfund sites within a 2-mile radius.
- Years: 2010, 2015
- Scale: Count.
- Definition: We linked each 2010 census block to all uncleaned Superfund sites as of June 30, 2010 and June 30, 2015 and counted, for each block, the number of uncleaned Superfund sites within a 2-mile radius of the block centroid. To obtain a measure of exposure at the census tract level, we averaged the number of Superfund sites across blocks within a tract, using as weights the proportion of the tract child population residing in a

given block, taken from 2010 Decennial Census data.

- Source: Data on the location of Superfund sites are taken from the Superfund National Priorities List (NPL) compiled by the EPA, https://epa.maps.arcgis.com/apps/webappviewer/index.html?id=33ce-bcdfdd1b4c3a8b51d416956c41f1, downloaded on 02/03/2019.[103]
- Source geography: Point data.

Industrial pollutants in air, water or soil

- Description: Index of toxic chemicals released by industrial facilities.
- Years: 2010, 2015.
- Scale: See definition.
- Definition: For this indicator, we aggregated the census block group level EPA RSEI (Risk-Screening Environmental Indicators) score (variable TOX-CONC) to the census tract level. The RSEI Score is calculated as toxicity weight multiplied by the exposed population multiplied by the estimated dose based on emission data of over 600 toxic chemicals.
- Source: EPA, https://www.epa.gov/rsei, downloaded on 10/30/2018.[104]
- Source geography: Census block group.
- Notes: The 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 EPA used the following data and methods to create the indicator. 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 locating facilities and people geographically, and attributes characteristics of the physical environment, such as meteorology, to areas surrounding the facilities once located. To locate the facilities and attribute corresponding data, the model describes the U.S. and territories on a 810m by 810m grid system. For each cell in the grid, a location "address" is assigned based on latitude and longitude. In order to estimate potential exposure, TRI facilities and the U.S. population must be geographically located on the model grid. TRI facilities are located using the facilities' latitude/longitude coordinates. To locate population, the model uses U.S. Decennial Census data at the block level. These data are used to create detailed age-sex-defined population groups for each of the census blocks in the U.S. Because the census block boundaries change between Decennial Census years, each set of census block level data is first transposed onto the model grid (which is unchanging) using an area weighted method. Once populations for 1990, 2000, and 2010 are placed on the grid system, the model uses a linear interpolation for each grid cell to create annual estimates of the population sizes for each year between 1990 and 2000 and between 2000 and 2010. The straight-line plot between 1990 and 2000 is extrapolated backward to estimate population for 1988-89 and the straight-line plot between 2000 and 2010 is extrapolated forward to estimate population for the years after 2010. Once facilities and people are located on the model's grid system, three main components (quantity of chemicals released/transferred, adjustments for chronic human health toxicity and adjustments for exposure potential and population size) are used to compute risk-related impacts in the model.

Airborne microparticles

- Description: Mean estimated microparticle (PM 2.5) concentration in µg/m^3.
- Years: 2010, 2014
- Scale: µg/m^3
- Definition: Microparticle exposure is defined as the mean estimated microparticle (PM 2.5) concentration in µg/m^3. Specifically, we standardized the indicator (see Methods) and then multiplied the resulting z-scores by -1.
- Source: CDC, National Environmental Public Health Tracking Network, https://data.cdc.gov/Environmental-Health-Toxicology/Daily-Census-Tract-Level-PM2-5-Concentrations-2011/fcqm-xrf4, downloaded on 02/12/2019.[105]
- Source geography: Census tract.
- Notes: The CDC used output from a Bayesian space-time downscaling fusion model called Downscaler (DS). DS combines PM2.5 monitoring data from the EPA Air Quality System (AQS) repository of ambient air quality data (e.g., National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS)) and simulated PM2.5 data from the deterministic prediction model, Models-3/Community Multiscale Air Quality (CMAQ) for each census tract within the contiguous U.S. for each day of the modeling year. The source data

does not cover Alaska and Hawaii. Census tract data for these states are missing.

Ozone concentration

- Description: Mean estimated 8-hour average ozone concentration.
- Years: 2011, 2014.
- Scale: Parts per billion.
- Definition: Ozone concentration is defined as the mean estimated 8-hour average ozone concentration in parts per billion (ppb) within 3 meters of the surface of the earth. We standardized the indicator (see Methods) and then multiplied the resulting z-scores by -1.
- Source: CDC, National Environmental Public Health Tracking Network, https://data.cdc.gov/Environmental-Health-Toxicology/Daily-Census-Tract-Level-Ozone-Concentrations-2011/372p-dx3h, downloaded on 02/12/2019.[106]
- Source geography: Census tract.
- Notes: The CDC used the output from a Bayesian space-time downscaling fusion model called Downscaler (DS). DS combines ozone monitoring data from the EPA AQS repository of ambient air quality data (e.g., NAMS/SLAMS) and simulated ozone data from the deterministic prediction model, Models3/CMAQ for each of the U.S. Census tracts within the contiguous U.S. for each day of the modeling year. The source data does not cover Alaska and Hawaii. Census tract data for these states are missing.

Extreme heat exposure

- Description: Summer days with maximum temperature above 90 degrees Fahrenheit.
- Years: 2009-2011, 2014-2016
- Scale: Count.
- Definition: We used annual data on the number of days with temperatures above 90 degrees Fahrenheit for the months of May through September in a given census tract and then averaged these annual counts of days over three-year periods.
- Source: CDC, National Environmental Public Health Tracking Network, based on temperature data from the North American Land Data Assimilation System (NLDAS), http://www.cdc.gov/ephtracking, downloaded on 02/06/2019.[107]
- Source geography: Census tract.
- Notes: The CDC used the NLDAS as the primary source of temperature data. The NLDAS contain gridded climate variables measured with hourly frequency for a 1/8th-degree grid (approximately 14x14 km, 103,936 grid cells) that covers the contiguous U.S. Grid-level climate variables are downcast to the census block level by allocating each block to the grid cell that contains the block centroid. Using block level population as weights, block level data are aggregated to the census tract level. The NLDAS does not cover Alaska and Hawaii. Census tract data for these states are missing.

Health resources

Health insurance coverage

- Description: Percent individuals ages 0-64 with health insurance coverage.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Definition: The number of individuals ages 0-64 years with health insurance coverage divided by the number of individuals ages 0-64, times 100.
- Source: ACS 5-Year Summary Files, Table B27001, 2012 and 2017.[108]
- Source geography: Census tract.

APPENDIX 1.3: SOCIAL AND ECONOMIC DOMAIN INDICATORS AND SOURCES

Economic opportunities

Employment rate

- Description: Percent adults ages 25-54 who are employed.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Definition: The number of adults ages 25-54 years who are employed in the civilian labor force divided by the number of adults ages 25-54 years, times 100.
- Source: ACS 5-Year Summary Files, 2012 and 2017, Table B23001.[109]
- Source geography: Census tract.

Commute duration

- Description: Percent workers commuting more than one hour one way.

- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Definition: The number of workers ages 16 years and older who did not work at home with a mean travel time from home to work of 60 minutes or longer divided by the number of workers ages 16 years and older who did not work at home, times 100.
- Source: ACS 5-Year Summary Files, 2012 and 2017, Table B08303.[110]
- Source geography: Census tract.

Economic and social resources

Economic resource index

- Description: Index combining poverty rate, public assistance rate, homeownership rate, high-skill employment and median household income.
- Years: 2008-2012, 2013-2017.
- Scale: Z-score.
- Definition: The indicator is constructed as the first principal component of five indicators of neighborhood economic resources: poverty rate, public assistance rate, homeownership rate, high-skill employment and median household income. The neighborhood poverty rate is defined as the number of individuals all ages who live in families/households with incomes below 100% of the federal poverty threshold divided by the number of individuals of all ages for whom poverty status could be determined, times 100. The public assistance rates is defined as the number of households receiving cash public assistance or Food Stamps/Supplemental Nutrition Assistance Program (SNAP) divided by the number of households, times 100. The homeownership rates is defined as the number of housing units that are owner occupied divided by the number of occupied housing units in the tract, times 100. High-skill employment is defined as the number of individuals ages 16 years and over who are employed in management, business, financial, computer, engineering, science, education, legal, community service, health care practitioner, health technology, arts and media occupations divided by the number of individuals ages 16 years and over, times 100. Median household income is the median income across all households in a census tract, deflated in constant 2017 U.S. dollars using Consumer Price Index (CPI) for urban consumers provided by the Bureau of Labor Statistics.[111]
- Source: ACS 5-Year Summary Files, 2012 and 2017, Tables B17001,[112] B19058,[113] B25003,[114] C24010,[115] and B19013.[116]
- Source geography: Census tract.
- Notes: Principal component analysis was performed on the five indicators (standardized using the procedure outlined above) measured in 2008-12, weighing by the census tract population of children ages 0-17 from the Decennial Census. All five variables were highly intercorrelated (alpha = 0.87) and the first component explained 71% of their total variation. We obtained component loadings and used these to construct predicted component scores, by multiplying indicator z-scores in 2008-2012 and 2013-2017 with component loadings and summing across the resulting products.

Single-headed households

- Description: Percent family households headed by a single parent.
- Years: 2008-2012, 2013-2017.
- Scale: Percent.
- Definition: The number of single parent (male householder and no wife present or female householder and no husband present) family households with children ages 0-17 years related to the householder divided by the number of family households with children ages 0-17 years related to the householder, times 100.
- Source: ACS 5-Year Summary Files, 2012 and 2017, Table B17010.[117]
- Source geography: Census tract.

Appendix 2: School indicators

COI 2.0 includes six indicators calculated from school-level data, aggregated to the census tract level:

- The percentage of public-school students in grades one through five receiving FRPL.
- The percentage of third grade public school students scoring proficient on standardized math tests.
- The percentage of third grade public school students scoring proficient on standardized reading tests.
- The percentage of ninth grade public school students graduating from high school on time.
- The percentage of 11th through 12th grade public school students taking at least one AP class.
- The percentage of public school teachers with two years or less experience.

These indicators are based on some common datasets and procedures that we describe in the following sections. We utilized school and school district data from the following sources:

- National Center for Education Statistics Common Core of Data: The CCD is an annual database of all public elementary and secondary schools and school districts. From the CCD we drew a comprehensive list of public schools and the following information for each school: total and grade-specific enrollment counts for all students, total enrollment counts by race/ethnicity, the number of students eligible for FRPL, their local education administrative district identifier (LEAID) and school latitude and longitude.
- National Center for Education Statistics School District Boundary Files: The NCES School District Boundary Files allow us to assign schools and census blocks to a common, spatially defined geographic school district.
- The latitude and longitude for census block centroids are taken from TIGER/Line 2010 shapefiles, obtained from the U.S. Census Bureau.[79]
- Stanford Education Data Archive (SEDA) Version 2.1: The SEDA data files contain school district-level data on math and reading proficiency that is comparable across states and over time. We also use a SEDA school crosswalk file to exclude schools that we consider outside our universe of schools.[118; 119]
- U.S. Department of Education EDFacts Data Files: We used EDFacts data files that include grade-specific, school-level and school district-level data on math and reading proficiency as well as adjusted cohort high school graduation rates. The data are coarsened into intervals of varying size depending on the underlying number of students for which data are reported.
- GreatSchools (GS) data: A proprietary data source that includes uncoarsened data on student proficiency levels and adjusted cohort high school graduation rates based on web-scraped school-level data.
- U.S. Department of Education Office for Civil Rights Data Collection: 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.

For each of the data sources, we obtained and processed school and/or school-district level data for the school-years 2010/11 to 2015/16.

Universe of schools

Our universe of schools is all public schools geographically close to where children live. We start with the list of all public primary and secondary schools in the CCD. We omit schools located outside the 50 states and Washington, D.C. and then further drop schools matching the following criteria:

- Missing data on latitude or longitude, and schools with assigned coordinates of zero degrees latitude and zero degrees longitude.
- Schools for which kindergarten is the highest grade.
- Schools for which adult education is the lowest grade, or for which adult education is the lowest grade and the highest grade is "ungraded."
- Schools with zero or missing total enrollment.
- Schools that do not have at least one student enrolled in grades 1-5 and grades 9-12. These are the grade ranges that our indicators are defined for, and we hence omit schools solely enrolling students in grades 6-8 from our universe.
- We only include schools that are either "operational at the time of the last report and are currently operational" or that have been "opened since the time of the last report."
- We delete schools and school districts that are designated as virtual schools/districts or districts serving predominantly special needs students using information included in the SEDA school crosswalk file (version 2.1).[119]

The resulting dataset includes about 80,000 schools per school year for the school years 2010/11 through 2015/16.

Linking census blocks to school districts

Each school is assigned to a local administrative school district (ASD) governed by a local education agency. Some ASDs do not have a spatially defined catchment area, e.g., charter school or other special school districts. Each school is also located within a geographic school district (GSD), a geographic catchment area defined in the NCES School District Boundary shapefiles.

Based on the longitude and latitude of schools and census block centroids, we link every school in our universe and every block with a non-zero child population in the 2010 Decennial Census to a GSD. For 8% of schools, ASD and GSD identifiers differ, i.e., the ASD identifier in the CCD does not have a corresponding spatial catchment area in the NCES School District Boundary Files. We assign these schools to a state-specific synthetic district including all such schools in a given state.

ASD and GSD identifiers can differ for several reasons. Charter schools and other special administrative districts may not have a spatial catchment area. A school may relocate (sometimes only temporarily) and thereby is assigned to another GSD even if its ASD (and its student intake population) do not change.

For each GSD, we calculate pairwise distances between each block assigned to that GSD and a set of schools. Specifically, we used the geospheres library in R to calculate the geodesic distance between point data (e.g., block centroids and schools) using their respective latitude and longitude coordinates on the WGS84 ellipsoid.[120]

The set of schools is defined as all schools assigned to that GSD, i.e., all schools with identical ASD and GSD identifiers, and all schools belonging to the synthetic district in the state in which the block is located. We then delete all school-block pairs further than 30 miles apart. For the three largest school districts, we lower this threshold to 20 miles in order to limit the number of rows of the resulting dataset.

In states where elementary and secondary school districts exist (as opposed to a single unified school district), a block may be linked to both an elementary or secondary school district. This results in duplicate block-school pairs, one for the elementary and one for the secondary district. When calculating indicators, e.g., at the elementary school level we include only those block-school pairs corresponding with the elementary district and follow a similar procedure for secondary school-based indicators.

The resulting dataset includes all blocks with a non-zero child population that have at least one school within 30 miles of the block centroid. For each block, the data include one row for each school the block is linked to, i.e., all schools that have the same GSD and are within 30 miles of the block centroid and all "arealess" schools in its home state within a 30-mile radius.

The resulting dataset has approximately 340 million rows per school year. It is further subset for specific indicators, based on grade-specific enrollment, distance between block centroid and school, presence of and other characteristics, such as having non-missing data and other characteristics.

Defining nearby schools

Our ultimate goal is to create robust estimates of school-based indicators for each census tract (neighborhood) that reflect the schools that children in that tract most likely attend. To most accurately match children with schools, we begin at the census block level and later aggregate to the tract level (discussed below). For each block, we create a subset of schools which are nearest to the block centroid. We try 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 apply inverse distance weighting to place a greater emphasis on the nearest schools and their characteristics.

For each indicator-related grade (such as third grade for math proficiency), we create a subset of public schools by deleting those with zero enrollment as well as those missing data necessary to calculate the indicator. Each of the schools in this subset are linked to a number of census blocks, based on distance to the block centroid. For each block, we then rank its linked schools based on distance (with nearest schools ranked highest) and calculate the running sum of students enrolled in a given grade across schools, adding schools sequentially from nearest to farthest school.

We next define an empty set of schools for each block. We add to the set the three nearest schools. If there are fewer than three schools located within 30 miles, we select however many are available (one or two). Unless it leaves the set empty, we delete schools further away than 20 miles. This leaves us with some blocks whose only school may be further than 20 miles away. If the running sum of students across the schools thus selected is less than 300, we add schools (from nearest to farthest) until the running sum of students either exceeds a threshold of 300 students (the threshold may not be reached in some cases) or the last school added is further away than 20 miles.

In effect, for each block, we create a subset of the three nearest schools but we include fewer than three if the second (and/or third) school is further than 20 miles from the block centroid. And, we increase the subset of schools beyond three if it is necessary to achieve a combined enrollment of at least 300 students.

Inverse distance weights

Once we have created a set of schools for each block, we create a weight for each school in the block that reflects its distance to the block centroid. These weights are a step function of distance between school and block centroid and are larger for schools nearer to the centroid.

Specifically, we define the weight for school s in block b as $w_{b,s} = 1/d_{b,s}$ where $d_{b,s}$ is the distance between school *s* and centroid of block *b* if the distance is greater than one mile. If the distance is one mile or less, we set $w_{b,s} = 1$. We top-code the weights of schools within one mile to a value of 1 in order to prevent situations where schools in the immediate vicinity of the block centroid exercise an outsize influence on the block-level statistics. Finally, we rescale the weights so that they sum up to the number of schools within the set of remaining schools.

For example, when calculating the school poverty indicator for school year 2010/11, the median number of schools per block in the final set is 3, the mean is 2.7, the minimum is 1 and the maximum is 8. The median centroid-school distance is 1.4 miles and the unweighted average distance is 3 miles (2.4 miles using the inverse distance weights).

Aggregating to the census tract level

Once we have created a subset of relevant schools with associated distance-based weights and variables for each block, we can calculate block-level statistics from school-level data by aggregating variables across the schools linked with each block.

For example, our school poverty indicator is calculated using information on the number of students eligible for FRPL (numerator) and the total number of students enrolled (denominator) for a given school. We weight both the numerator and denominator for each school by the rescaled distance-based weight and then sum these weighted numerators and denominators across the schools linked with each block. We then calculate the block-level indicator by (in this case) dividing the weighted numerator sum by the weighted denominator sum (and multiply by 100 to obtain the percent of students eligible for FRPL or "school poverty").

Finally, using 2010 Decennnial Census child population counts as weights, we aggregate the block-level statistics to the census tract level.

School poverty

Our school poverty indicator is defined as the percentage of students in grades one through five eligible for FRPL, based on the NCES CCD. About 2% of schools in our universe had missing information on the count of FRPL-eligible students, including all Massachusetts schools in 2015/16. We imputed missing counts using state-specific regressions that included school fixed effects, school-specific linear trends and variables counting the total number of students by race/ethnicity. R-squared statistics from the state-specific regressions varied between 0.95 and 0.99. We imputed the dependent variable, if, for a given school with missing data, the (school-specific) standard deviation of the residuals divided by the (school-specific) mean of the outcome was less than 0.3. This reduced the percent of schools with missing data on FRPL students to 0.5%, and in Massachusetts in 2015/16 to 9.8% (down from 100%). We then calculated school-poverty rates first at the census block level and then aggregated to the census tract level using the procedure described in Appendix 3.

Reading and math proficiency

Threats to validity

There are many challenges to calculating census tract-level indicators of student math and reading proficiency that are comparable across states and over time. We therefore recommend that users exercise caution when comparing reading and math proficiency data across states.

While care has been taken to ensure comparability, not all potential threats to comparability and accuracy can be addressed. We are unable to adjust school-level estimates if multiple tests were administered in a given state and year or adjust for other school-level processes that may influence test results, such as selective test taking.

However, we rely on the extensive efforts by Reardon et al. who have identified threats to comparability of reading and math proficiency data across all U.S. school districts.[118] Based on their work, we exclude school districts and sometimes states in certain years with non-comparable data, which limits the coverage of the reading and math score indicators.

Another threat to comparability are missing data and coarsened proficiency estimates. For a subset of schools, proficiency estimates are aggregated into intervals (e.g., 80-84% proficient). We imputed exact proficiency estimates in those instances and also imputed school-level proficiency estimates in a subset of cases.

Sources

For these indicators, we combined data from three sources:

- Stanford Education Data Archive (SEDA) Version 2.1: The SEDA data files contain school-district level grade-specific data on math and reading proficiency that is comparable across states and over time. We use the data file with estimates in NAEP scale score points, a metric that is comparable across districts, states and years.[119]
- U.S. Department of Education EDFacts Data Files: The EDFacts files include grade-specific, school-level and school district level data on the percentage of third grade students who tested proficient in math and or reading on state-specific standardized tests.
- Great Schools (GS) data: A proprietary data source that includes uncoarsened data on student proficiency levels with lower coverage of schools in certain years than EDFacts data.

Schools included

We take all schools in our universe with non-zero and non-missing third grade enrollment from our universe of schools defined using CCD and SEDA data (see Appendix 3) for school years 2010/11 through 2015/16. We attach to these data the percentage of student scoring proficient in math and reading from EDFacts and GS data using school-level NCES identifier that is common across datasets.

Because we rely on SEDA data to map state-specific proficiency scores onto a nationally-comparable scale, we drop districts excluded from the SEDA data. SEDA excludes districts for which data quality is considered to be low. These exclusions result in the loss of 17% of schools from our dataset, which corresponds to 12% of third grade student enrollment. Statistics reported in the following paragraphs exclude these schools and, unless noted otherwise, use third grade enrollment counts as weights.

Imputation of coarsened and missing data

Among the remaining schools, data on the percentage of students scoring proficient are coarsened into intervals for the majority of schools. For example, 60% of the non-missing school-level math proficiency estimates are coarsened into intervals that are up to 4 percentage points wide (e.g., 80-84%) and another 30% of the non-missing observations are coarsened into intervals that are between 5 and 10 percentage points wide.

To obtain more precise estimates, we impute the coarsened EDFacts proficiency estimates with uncoarsened GS proficiency estimates, if the GS estimates fall within the EDFacts proficiency score interval. In the resulting dataset, math proficiency scores are uncoarsened or coarsened into intervals maximally up to 10 percentage points wide for 88% of schools. Proficiency estimates for 6% of schools are coarsened into intervals wider than 10 percentage points and 6% of schools have missing proficiency data.

Next, we impute both the missing and remaining coarsened proficiency data using Ordinary Least Squares (OLS) Regression. The outcome to be predicted in the regression analyses is the (uncoarsened) percentage proficient. For schools with coarsened data, first impute the midpoint of the proficiency interval, e.g., 82% for a proficiency score of 80-84%. We use the following variables as predictors: racial/ethnic composition of the student population and school poverty from the CCD data, and district-level math and reading proficiency from EDFacts. Missing covariate data was imputed with the district median across schools, and in cases where the district estimate is missing (or the district median cannot be calculated because data on all schools are missing), we impute the state-wide median.

We then run state-specific regressions of proficiency estimates (midpoints of intervals, in case school data are coarsened) on school fixed effects, year fixed effects, school racial/ethnic composition, school poverty, district-level proficiency estimates and interactions between district level proficiency and school-level variables (poverty and racial ethnic composition).

Each observation is weighted by the inverse of the width of the proficiency estimate interval, e.g., 1/5 for an interval of width 5 (e.g., 80-84%). This weighting is performed so that the resulting predictions are more strongly informed by covariate values of those observations for which the outcome is known with greater precision. Weights are rescaled so that they sum to the number of schools in each state and year.

The state-specific regressions yield R^2 statistics ranging from 0.69 to 0.98 (average 0.91) and root mean squared errors (RMSE) statistics ranging from 4 to 10 (average 6). We then impute the outcome variable for all remaining coarsened estimates with the predicted outcome from the state-specific regressions, if the predicted outcome fell within the coarsened interval and the district level RMSE was less than 10. We imputed the midpoint of the interval if the predicted value was outside the coarsened range or the district level RMSE was greater than 10. And we imputed missing proficiency estimates with the predicted outcome if the district-level RMSE was less than 10. In the resulting dataset, 1.1% of schools still have missing data.

Achieving comparability across states and over time

Finally, we rescale the school-level proficiency estimates to make them comparable across states using subject-specific state and year-specific multipliers calculated from SEDA and EDFacts data.

From SEDA, we obtain district-level third grade proficiency estimates on the NAEP scale. The district level data are comparable across states and over time. We average proficiency estimates across districts within states and years, using the total number of students enrolled in third grade as weights. Similarly, we compute average subject-specific proficiency estimates by state and year from our imputed school-level EDFacts dataset, using the number of students enrolled in third grade as weights.

We combine both state-year level datasets to calculate a state and year-specific multiplier. For state *s* and year *t*, the multiplier $m_{s,t}$ is defined as $m_{s,t} = SN_{s,t} / EF_{s,t}$, where SN is the average proficiency level on the NAEP-scale in state *s* and year *t* from SEDA data and $EF_{s,t}$ is the average percentage proficient in state s and year t from EDFacts data. We then multiply each school-level estimate with the multiplier $m_{s,t}$ for a given state and year to obtain a school-level proficiency estimate on the NAEP scale.

Additional notes

Because of the deletion of school districts due to data quality, we lack observations for some states in our reference school years, 2010/11 and 2015/16. For the 2010 indicator, we substituted 2011/12 data for New Hampshire (math proficiency estimates) and Wyoming and 2012/13 data for Colorado. For 2015, we substituted 2014/15 data for Alaska, New York, North Dakota, Oklahoma and West Virginia (reading proficiency estimates) and 2013/14 data for Montana, Nevada and Washington.

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 that graduate from high school with a regular diploma in four years or less.

Sources

For this indicator, we combined data from two sources:

- U.S. Department of Education EDFacts Data Files: From EDFacts, we obtain school and school-district level data on adjusted cohort graduation rates. The data are coarsened into intervals of varying sizes depending on the underlying number of students.

- Great Schools (GS) data: A proprietary data source from GS that includes uncoarsened data on cohort graduation rates, but has lower coverage than EDFacts in certain states and years.

Schools included

We take all schools in our universe with non-zero and non-missing enrollment in any grades from 9 to 12 from our universe of schools defined using CCD and SEDA data (see Appendix 3) for school years 2010/11 through 2015/16. We attach to these data the adjusted cohort graduation rates from EDFacts and GS data using school-level NCES identifier that is common across datasets.

Imputation of coarsened and missing data

To obtain more precise estimates, we impute the coarsened EDFacts graduation rates with uncoarsened GS rates, if the GS estimates fall within the EDFacts graduation rate interval. In the resulting dataset, graduation rates are uncoarsened or coarsened into intervals maximally 10 percentage points wide for 75% of schools. Graduation rates for 10% of schools are coarsened into intervals wider than 10 percentage points and 15% of schools have missing graduation data.

Next, we impute both the missing and remaining coarsened graduation rates using OLS Regression. The outcome to be predicted in the regression analyses is the (uncoarsened) percentage proficient. For schools with coarsened data, we first impute the midpoint of the graduation rate interval, e.g., 82 for a graduation rate of 80-84%.

We use the following variables as predictors: racial/ethnic composition of the student population and school poverty from the CCD data and district-level graduation rates from EDFacts. Missing covariate data were imputed with the district median across schools, and in cases where the district estimate is missing (or the district median cannot be calculated because data on all schools are missing), we impute the state-wide median.

We then run state-specific regressions of graduation rates (midpoints of intervals for coarsened estimates) on school fixed effects, year fixed effects, school-level racial/ethnic composition variables, school poverty, district-level graduation rates and interactions between district level graduation rates and school-level variables (poverty and racial ethnic composition).

Each observation is weighted by the inverse of the width of the coarsened graduation rate interval, e.g., 1/5 for an interval of width 5 (80-84%). This weighting is performed so that the resulting predictions are more strongly informed by covariate values of those observations for which the outcome is known with greater precision. Weights are rescaled so that they sum to the number of schools in each state and year.

We then impute the outcome variable for all coarsened estimates with the predicted outcome from the state-specific regressions if the predicted outcome fell within the coarsened range and if the district level RMSE of the regressions was less than 10. We imputed the midpoint of the interval if the predicted value was outside the coarsened range or the district level RMSE was greater than 10 and we imputed missing with predicted graduation rates if the district-level RMSE was less than 10. In the resulting dataset, 4.5% of schools still have missing data.

Appendix 3: Linking census tracts over time

In this section, we provide further detail on the census tracts included and comparability of census tracts over time. 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 (GEOID) the 11-digit variable uniquely identifying each census tract. We cross-

Table A3.1 Changes in Census tract definitions and identifiers (GEOIDs)

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 definition changed	A small portion of 2010 tract 0230.00 was reallocated to 2010 tract 9402.00. The newly formed tract (from 2010 0230.00 and 2010 9402.00) is labeled 0249.00. Because the reallocated area was very small, we assume that 0230.00 is comparable over time 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 definition 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	County equivalent Bedford City merged into Bedford County	Bedford City, Virginia changed its legal status to town, ending its independent city status (county equivalent), and was absorbed as a municipality within Bedford County, Virginia.
2015	02270000100	02158000100	County code changed	Wade Hampton Census Area, Alaska, was renamed as Kusilvak Census Area.
2015	46113940500	46102940500	County code changed	Shannon County, South Dakota, 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	

walked these new geographic identifiers to 2010 geographic identifiers. The change in GEOIDs was almost always due to a renaming of the tract, 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 (CA) county census tract number 06037930401 (2010 GEOID) were redrawn, and its comparability over time is therefore limited.

Table A3.1 lists all 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 for comparability over time.

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 (Project Hope), 33*(11), 1948-1957. doi:10.1377/hlthaff.2014.0679.
2. Acevedo-Garcia, D., McArdle, N., Hardy, E. F., Crisan, U. I., Romano, B., Norris, D., . . . Reece, J. (2014). Child Opportunity Index mapping. *diversitydatakids.org and the Kirwan Institute for the Study of Race and Ethnicity*, 1-52.
3. Acevedo-Garcia, D., Hardy, E. F., McArdle, N., Crisan, U. I., Romano, B., Norris, D., . . . Reece, J. (2016). The Child Opportunity Index: Measuring and mapping neighborhood-based opportunities for U.S. children. *diversitydatakids.org and the Kirwan Institute for the Study of Race and Ethnicity*, 1-56.
4. Acevedo-Garcia, D., McArdle, N., Hardy, E., Dillman, K.-N., Reece, J., Crisan, U. I., . . . Osypuk, T. L. (2016). Neighborhood opportunity and location affordability for low-income renter families. *Housing Policy Debate, 26*(4-5), 607-645. doi:10.1080/10511482.2016.1198410.
5. O'Hare, W. P., Mather, M., Dupuis, G., Land, K. C., Lamb, V. L., & Fu, Q. (2013). Analyzing differences in child well-being among U.S. states. *Child Indicators Research, 6*(2), 401-431. doi:10.1007/s12187-012-9173-3.
6. Lee, B. J. (2014). Mapping domains and indicators of children's well-being. In A. Ben-Arieh, F. Casas, I. Frønes, & J. E. Korbin (Eds.), *Handbook of child well-being: Theories, methods and policies in global perspective* (pp. 2797-2805). Dordrecht: Springer Netherlands.
7. Currie, J., Greenstone, M., & Moretti, E. (2011). Superfund cleanups and infant health. *American Economic Review, 101*(3), 435-441. doi:10.3386/w16844.
8. Persico, C., Figlio, D., & Roth, J. (2016). Inequality before birth: The developmental consequences of environmental toxicants. doi:10.3386/w22263.
9. Vrijheid, M. (2000). Health effects of residence near hazardous waste landfill sites: A review of epidemiologic literature. *Environmental Health Perspectives, 108*(Suppl 1), 101-112. doi:10.1289/ehp.00108s1101.
10. Leventhal, T., & Dupéré, V. (2019). Neighborhood effects on children's development in experimental and nonexperimental research. *Annual Review of Developmental Psychology, 1*(1), 149-176. doi:10.1146/annurev-devpsych-121318-085221.
11. Sharkey, P., & Elwert, F. (2012). The legacy of disadvantage: Multigenerational neighborhood effects on cognitive ability. *American Journal of Sociology, 116*(6), 1934-1981.
12. Pebley, A. R., & Sastry, N. (2003). *Neighborhoods, poverty and children's well-being: A review*. Retrieved from website: https://www.rand.org/pubs/drafts/DRU3001.html.
13. 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.
14. 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.
15. 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 of the*

United States of America, 101(33), 11920. doi:10.1073/pnas.0403418101.
16. Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). The Opportunity Atlas: Mapping the childhood roots of social mobility. *National Bureau of Economic Research*(25147), 1-95. doi:10.3386/w25147.
17. Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). Data. *Opportunity Insights*.
18. Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *The Russell Sage Foundation Journal of the Social Sciences, 2*(5), 34. doi:10.7758/rsf.2016.2.5.03.
19. Magnuson, K., & Duncan, G. J. (2016). Can early childhood interventions decrease inequality of economic opportunity? *The Russell Sage Foundation journal of the social sciences: RSF, 2*(2), 123-141. doi:10.7758/RSF.2016.2.2.05.
20. Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics, 92*(3), 562-576. doi:10.3386/w14277.
21. Mickelson, R. A. (2018). *Is there systematic meaningful evidence of school poverty thresholds?* (14). Retrieved from website: https://school-diversity.org/wp-content/uploads/2018/09/NCSD_Brief14.pdf.
22. Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *The Russell Sage Foundation Journal of the Social Sciences, 5*(2), 40-68. doi:10.7758/rsf.2019.5.2.03.
23. Orfield, G., & Lee, C. (2004). *"Brown" at 50: King's dream or "Plessy's" nightmare?* : Harvard Education Publishing Group, 8 Story Street, 1st Floor, Cambridge, MA 02138.
24. 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), 1249-1262. doi:10.1016/j.ssresearch.2011.02.003.
25. Jargowsky, P. A., & El Komi, M. (2009). Before or after the bell? School context and neighborhood effects on student achievement. *National Center for Analysis of Longitudinal Data in Education Research*(28).
26. Chapple, K. (2006). Overcoming mismatch: Beyond dispersal, mobility, and development strategies. *Journal of the American Planning Association, 72*(3), 322-336. doi:10.1080/01944360608976754.
27. Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337. doi:10.1037/0033-2909.126.2.309.
28. Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *The Journal of Human Resources, 33*(4), 915-946. doi:10.2307/146403.
29. Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school? *Psychological Methods, 1*(3), 249-260. doi:10.1037/1082-989X.1.3.249.
30. Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives, 33*(1), 185-208. doi:10.1257/jep.33.1.185.
31. 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.
32. Hartig, T., Mitchell, R., de Vries, S., &

Frumkin, H. (2014). Nature and health. *Annual Review of Public Health, 35*(1), 207-228. doi:10.1146/annurev-publhealth-032013-182443.
33. 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 (Project Hope), 34*(11), 1858-1868. doi:10.1377/hlthaff.2015.0667.
34. Thomsen, M. R., Nayga, R. M., Alviola, P. A., & Rouse, H. L. (2016). The effect of food deserts on the Body Mass Index of elementary schoolchildren. *American Journal of Agricultural Economics, 98*(1), 1-18. doi:10.1093/ajae/aav039.
35. Ver Ploeg, M., & Rahkovsky, I. (2016). *Recent evidence on the effects of food store access on food choice and diet quality*. Retrieved from website: https://www.ers.usda.gov/amber-waves/2016/may/recent-evidence-on-the-effects-of-food-store-access-on-food-choice-and-diet-quality/.
36. Hartig, T., & Kahn, P. H. (2016). Living in cities, naturally. *Science, 352*(6288), 938-940. doi:10.1126/science.aaf3759.
37. 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*. doi:10.1289/ehp.1510363.
38. Larson, L. R., Barger, B., Ogletree, S., Torquati, J., Rosenberg, S., Gaither, C. J., . . . Schutte, A. (2018). Gray space and green space proximity associated with higher anxiety in youth with autism. *Health & Place, 53*, 94-102. doi:10.1016/j.healthplace.2018.07.006.
39. 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*. doi:10.1289/ehp.1307704.
40. 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.
41. 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.
42. Laurito, A., Lacoe, J., Schwartz, A. E., Sharkey, P., & Ellen, I. G. (2019). School climate and the impact of neighborhood crime on test scores. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 5*(2), 141-166. doi:10.7758/RSF.2019.5.2.08.
43. 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.
44. Agarwal, N., Banternghansa, C., & Bui, L. T. M. (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.
45. 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), 1121-1167. doi:10.1162/00335530360698513.
46. Gaffin, J. M., Kanchongkittiphon, W., & Phipatanakul, W. (2014). Perinatal and early childhood environmental factors influencing allergic asthma immunopathogenesis. *International*

Immunopharmacology, 22(1), 21-30. doi:10.1016/j.intimp.2014.06.005.

47. Liu, J., & Lewis, G. (2014). Environmental toxicity and poor cognitive outcomes in children and adults. *Journal of environmental health, 76*(6), 130-138.
48. 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.
49. 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.
50. Naujokas, M. F., Anderson, B., Graziano, J. H., Thompson, C., & Suk, W. A. (2013). The broad scope of health effects from chronic arsenic exposure: Update on a worldside public health problem. *Environmental Health Perspectives (Online); Research Triangle Park, 121*(3), 295. doi:10.1289/ehp.1205875.
51. Ogneva-Himmelberger, Y., Dahlberg, T., Kelly, K., & Simas, T. A. M. (2015). Using Geographic Information Science to explore associations between air pollution, environmental amenities, and preterm births. *AIMS public health, 2*(3), 469-486. doi:10.3934/publichealth.2015.3.469.
52. Pope, C. A. (2000). Epidemiology of fine particulate air pollution and human health: Biologic mechanisms and who's at risk? *Environmental Health Perspectives, 108*(Suppl 4), 713-723. doi:10.1289/ehp.108-1637679.
53. 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.
54. Shier, V., Nicosia, N., Shih, R., & Datar, A. (2019). Ambient air pollution and children's cognitive outcomes. *Population and Environment*. doi:10.1007/s11111-019-0313-2.
55. Stillerman, K. P., Mattison, D. R., Giudice, L. C., & Woodruff, T. J. (2008). Environmental exposures and adverse pregnancy outcomes: A review of the science. *Reproductive Sciences, 15*(7), 631-650. doi:10.1177/1933719108322436.
56. 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. *The Lancet, 352*(9126), 423-427. doi:10.1016/S0140-6736(98)01352-X.
57. 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.
58. 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.
59. Goodman, J., Hurwitz, M., Park, J., & Smith, J. (2018). Heat and learning. doi:10.3386/w24639.
60. 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.
61. Zivin, J. G., & Shrader, J. (2016). Temperature extremes, health, and human capital. *Future of Children, 26*, 31-50.
62. 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.

63. Sommers, B. D., Maylone, B., Blendon, R. J., Orav, J. E., & 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). doi:10.1377/hlthaff.2017.0293.

64. 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 website: https://www.urban.org/sites/default/files/publication/94396/2001576-how-have-providers-responded-to-the-increased-demand-for-health-care-under-the-affordble-care-act_0.pdf.

65. Wilson, W. J. (1997). *When work disappears: The world of the new urban poor*. New York, NY: Vintage.

66. Mouw, T. (2007). Spatial mismatch hypothesis. *The Blackwell Encyclopedia of Sociology*. doi:10.1002/9781405165518.wbeoss215.

67. Mouw, T. (2000). Job relocation and the racial gap in unemployment in Detroit and Chicago, 1980 to 1990. *American Sociological Review, 65*(5), 730-753. doi:10.2307/2657544.

68. Andersson, F., Haltiwanger, J. C., Kutzbach, M. J., Pollakowski, H. O., & Weinberg, D. H. (2014). Job displacement and the duration of joblessness: The role of spatial mismatch. doi:10.3386/w20066.

69. Wilson, W. J. (2012). *The truly disadvantaged: The inner city, the underclass, and public policy* (2nd edition ed.). Chicago, IL: The University of Chicago Press.

70. Galster, G., & Sharkey, P. (2017). Spatial foundations of inequality: A conceptual model and empirical overview. *The Russell Sage Foundation Journal of the Social Sciences, 3*(2), 1-33. doi:10.7758/rsf.2017.3.2.01.

71. Massey, D. S., & Denton, N. (1993). American apartheid: Segregation and the making of the underclass. In D. B. Grusky (Ed.), *Social Stratification: Class, Race, and Gender in Sociological Perspective*: Westview Press.

72. 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.

73. Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "Neighborhood Effects": Social processes and new directions in research. *Annual Review of Sociology, 28*(1), 443-478. doi:10.1146/annurev.soc.28.110601.141114.

74. Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2018). Race and economic opportunity in the United States: An intergenerational perspective. *National Bureau of Economic Research*(24441), 1-105. doi:10.3386/w24441.

75. Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics, 129*(4), 1553-1623. doi:10.1093/qje/qju022.

76. 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.

77. 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.

78. 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.
79. U.S. Census Bureau. (2019). TIGER/Line Geodatabases.
80. Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L. F., & Palmer, C. (2019). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. doi:10.3386/w26164.
81. 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), 1025-1033. doi:10.1093/aje/kwu018.
82. 500 Cities Project: Local data for better health. (2017). *Centers for Disease Control and Prevention*.
83. Arias, E., Escobedo, L. A., Kennedy, J., Fu, C., & Cisewski, J. (2018). *U.S. Small-area Life Expectancy Estimates project: Methodology and results summary* (181). Retrieved from website: https://www.cdc.gov/nchs/data/series/sr_02/sr02_181.pdf.
84. National Center for Health, S. (2018). United States Small-area Life Expectancy Estimates Project. *Centers for Disease Control and Prevention*.
85. Hagerty, M. R., & Land, K. C. (2007). Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights. *Sociological Methods & Research, 35*(4), 455-496. doi:10.1177/0049124106292354.
86. 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? Rising inequality, schools, and children's life chances*. New York, NY: Russell Sage Foundation.
87. 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*. doi:10.1073/pnas.0710189104.
88. Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, S95-S120. doi:10.1086/228943.
89. Sampson, R. J., Morenoff, J. D., & Earls, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review, 64*(5), 633-660. doi:10.2307/2657367.
90. Dircksen, J. C., & Prachand, N. G. (2016). *Healthy Chicago 2.0: Partnering to improve health equity*. Retrieved from website: https://www.chicago.gov/content/dam/city/depts/cdph/CDPH/Healthy%20Chicago/HC2.0Upd4152016.pdf.
91. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B14003*.
92. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B14001*.
93. U.S. Department of Education Office for Civil Rights. (2012). *Civil Rights Data Collection: User guide*. Retrieved from website: https://ocrdata.ed.gov/downloads/UserGuide.pdf.
94. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B15002*.
95. *USDA Food Access Research Atlas*. (2014). Retrieved from website: https://www.ers.usda.gov/webdocs/DataFiles/80591/documentation.pdf?v=0.
96. Ver Ploeg, M., Breneman, V., Farrigan, T., Hamrick, K., Hopkins, D., Kaufman, P., . . . Kim, S. (2009). *Access to affordable and*

nutritious food: Measuring and understanding food deserts and their consequences. Retrieved from website: https://www.ers.usda.gov/webdocs/publications/42711/12716_ap036_1_.pdf.
97. *National Environmental Public Health Tracking indicators and data: Land cover*. (2019). Retrieved from website: https://ephtracking.cdc.gov/showIndicatorPages.action?selectedContentAreaAbbreviation=22&selectedIndicatorId=134&selectedMeasureId=.
98. Thomas, J. V. (2015). Walkability Index. Retrieved from https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=%7B251AFDD9-23A7-4068-9B27-A3048A7E6012%7D.
99. Thomas, J., & Zeller, L. (2017). *National Walkability Index user guide and methodology*. Retrieved from website: https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=%7B251AFDD9-23A7-4068-9B27-A3048A7E6012%7D.
100. Ramsey, K., & Bell, A. (2014). *Smart Location Database: Version 2.0 user guide*. Retrieved from website: https://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf.
101. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B25002*.
102. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B25004*.
103. U.S. Environmental Protection Agency. (2018). Superfund National Priorities List.
104. *Risk-Screening Environmental Indicators (RSEI) model*. (2014). Retrieved from website: https://www.epa.gov/rsei.
105. Kassinger, C. (2018). *Daily Census tract-level PM2.5 concentrations, 2011-2014*. Retrieved from website: https://data.cdc.gov/Environmental-Health-Toxicology/Daily-Census-Tract-Level-PM2-5-Concentrations-2011/fcqm-xrf4.
106. Kassinger, C. (2018). *Daily Census tract-level ozone concentrations, 2011-2014*. Retrieved from website: https://data.cdc.gov/Environmental-Health-Toxicology/Daily-Census-Tract-Level-Ozone-Concentrations-2011/372p-dx3h.
107. *National Environmental Public Health Tracking indicators and data: Climate change*. (2019). Retrieved from website: https://ephtracking.cdc.gov/showIndicatorPages.action?selectedContentAreaAbbreviation=22&selectedIndicatorId=134&selectedMeasureId=.
108. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B27001*.
109. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B23001*.
110. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B08303*.
111. U.S. Department of Labor Bureau of Labor Statistics. (2019). *CPI-All Urban Consumers*. Retrieved from website: https://data.bls.gov/timeseries/CUUR0000SA0.
112. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B17001*.
113. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B19058*.
114. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B25003*.
115. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table C24010*.
116. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B19013*.
117. U.S. Census Bureau. American Community Survey 5-Year Summary Files. *Table B17010*.

118. Reardon, S., Kalogrides, D., Ho, A., Shear, B., Shores, K., & Fahle, E. (2016). Stanford Education Data Archive 3.0. *Stanford University*.
119. Fahle, E., Shear, B., Kalogrides, D., Reardon, S., DiSalvo, R., & Ho, A. (2018). *Stanford Education Data Archive: Technical documentation* (2.1). Retrieved from website: https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v21.pdf.
120. Hijmans, R. J., Williams, E., & Vennes, C. (2019). *Package 'geosphere'*. Retrieved from website: https://cran.r-project.org/web/packages/geosphere/geosphere.pdf.