THE UNIVERSITY OF CHICAGO

THE COMMUNITY HUMAN DEVELOPMENT INDEX (CHDI): A NOVEL METRIC FOR PRECISION PUBLIC HEALTH

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRTIZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:
ECOLOGY AND EVOLUTION

BY
SURAJ SHETH

CHICAGO, ILLINOIS
DECEMBER 2022

For my mother, Anu Kris, and my grandmother, Shanthi Krishnaswamy

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to sincerely thank of all of my family, friends, mentors, teachers and advisors who supported me while I was working on my PhD degree. My advisor, Dr. Luis Bettencourt, is an amazing mentor and teacher who encouraged me to pursue novel and interdisciplinary research in the arena of Precision Public Health. His constant positive reinforcement at every step of the journey motivated me to excel in my field of interest. I am deeply grateful to my thesis committee members, Dr. Stefano Allesina, Dr. Habibul Ahsan, and Dr. Trevor Price for their invaluable guidance and constant feedback which inspired me to work across silos to study complex systems related to precision public health, allowing me to gain a deeper understanding of the subject and to evolve as a critical thinker. I also thank Dr. Meredith Cenzer for her time, advice and constant encouragement.

My mother, Anu Kris, is my biggest inspiration and role model. Her brilliance, passion for learning, kindness, optimism, and unconditional love and support have inspired me to become the person I am today. I owe my success to her, and I sincerely thank her from the bottom of my heart. My grandmother, Shanthi Krishnaswamy, has always been there for me, and her intelligence, curiosity and desire to explore the world have served as an example to be the best I can be. I thank my amazing and exceptionally brilliant sister Nikki, an aspiring physician, for inspiring me to succeed. I also thank my father for constantly motivating me to overcome all challenges. A heart-felt thank you to my dog Zoey, who is the embodiment of pure joy, and shows me daily that life is beautiful. She reminds me to enjoy life to the fullest by striking the right balance between work and fun. I am grateful to my uncle, Dr. Dilip Krishnaswamy, and my aunt, Vidya Govindan, who have always supported me, genuinely cared about my well-being, and have been excellent role models.

The University of Chicago, with its emphasis on knowledge building and integration among disciplines, has been a great academic home for me. As a mathematical biologist working in Biomedical Informatics and Precision Public Health, I have really enjoyed my project. I had the unique opportunity to create a framework and a novel metric in the field of Precision Public Health for descriptive, diagnostic, predictive and prescriptive data analytics to design effective targeted interventions to improve health of communities around the world. I am genuinely glad that my academic work can effectively be translated into practice to benefit communities globally. I sincerely thank the leadership and administration of the Interdisciplinary Scientist Training Program at the Pritzker School of Medicine, the Ecology and Evolution Department and the Mansueto Institute for Urban Innovation for jointly supporting me in this endeavor. I look forward to going back to the Pritzker School of Medicine this fall to continue my medical education and earn my medical degree.

I am deeply appreciative of my all of my family and friends who believed in me, cheered me on and stood by me through the years. I am humbly beholden to all the mentors and teachers who supported and encouraged me over the years, who have dedicated their lives to the dissemination of knowledge and have allowed me to stand on their shoulders to see farther than I could ever imagine. In the spirit of Noblesse Oblige, my sincere hope as a future physician-scientist is that my work will continue to greatly improve the health of communities around the globe, bring joy and happiness into people's lives, and leave a lasting, positive impact on the world. With a grateful heart and a sincere intent to pay it forward, I conclude with an excerpt from the poem "Stopping by Woods on a Snowy Evening" by Robert Frost:

The woods are lovely, dark and deep,
But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep.

ABSTRACT

Precision Public Health is fundamentally focused on determining the right intervention for the right population at the right time in order to improve health outcomes. It requires a multifaceted understanding of how both physical and social environmental conditions lead to better outcomes for residential populations over the long term. Here we present the Community Human Development Index (CHDI), a novel precision public health framework and metric to measure the capabilities of populations across spatial scales and over time. We show how CHDI can be constructed and measured across geographic scales in the United States, and show the substantial variation in development that emerges at the neighborhood scale. We then detail the changes in development indicators in the US at the county level over the last 30 years, showing how there have been average improvements across the US, but also rapidly increasing variation. We finally examine the relationship between health risk, vulnerability, and development at the census tract level in the US, and show that the CHDI can be an effective tool in measuring and mitigating health risk and vulnerability in communities.

CHAPTER 1

INTRODUCTION

Precision public health and human ecology seek to understand the impact of the social and built environments on human populations, or people that reside in an area [1]. The capabilities approach, a framework with roots in ethics, economics, and philosophy, developed by Amartya Sen and Martha Nussbaum, approaches human ecology by understanding whether an environment promotes the freedoms and capabilities of its population, giving them the resources they need to pursue their goals in life [2, 3]. In the 1990s, a team at the United Nations (UN) formalized the capabilities approach into a metric called the Human Development Index [4]. The Human Development Index, or HDI, is calculated based on three subindices, measuring life expectancy, education, and income. The HDI is calculated as a geometric mean of these three components [5, 6]. The benefit of the Human Development Index is that it has a strong theoretical foundation in the capabilities approach, and since its adoption by the United Nations, it has been widely used to measure development globally, becoming the gold standard for measuring quality of life [7]. A major drawback, however, is that it is calculated on a national scale, with one value for each nation. This, of course, misses a great deal of heterogeneity within these countries [8, 9]. Over the last three decades, studies have also examined the relationship between HDI, environments, and health outcomes [10, 11, 12, 13].

This interplay between environments and population health outcomes has taken on increased relevance in recent years due to the COVID-19 pandemic and the disparities observed in health outcomes across communities [14, 15, 16]. In response to the challenges posed by the pandemic, the United Nations Development Programme, or UNDP, named three strategic directions for the organization moving forward [17]. The first, "Leaving No One Behind," is an equity based approach focused on understanding the heterogeneity, or differences, that exist between communities across geographic spatial scales. The second, "Structural Transformation," focuses on understanding the changes that communities undergo over time, and how decision-makers can intervene to improve population health outcomes [18]. And the third, "Building Resilience", centers on understanding how risk and vulnerability vary from community to community, and require customized solutions at a local scale [19]. These new strategic directions reflect a shift from a national focus to one that emphasizes the importance of implementing solutions on both the national and community scales [20].

In order to achieve the UNDP goals, we require a framework that has very strong theoretical foundations. It also must be useful across spatial scales, working at the community level as well as the national level, to promote equitable outcomes and facilitate multi-level analysis. The framework must also work across countries, and allow for a common baseline to make comparisons across nations. It must be measurable across time if it is to be relevant to understanding change. And finally, it must accurately assess vulnerability, and account for how resilient communities are to major disruptions.

In this dissertation, I introduce a novel precision public health metric called the Community Human Development Index (CHDI), which has all of these characteristics. This framework uses actionable data and multilevel analysis at the state, county, city, and neighborhood scales to drive progress towards localizing health-related United Nations Sustainable Development Goals. I utilize the CHDI to systematically characterize development for neighborhoods in the United states, across geographic scales and over time, and examine the relationship between health risk and development at the census tract level. In the following chapters, any reference to HDI at a subnational scale is referring to the Community Human Development Index (CHDI).

1.1 Overview of Dissertation

1.1.1 Chapter 2: The Community Human Development Index (CHDI): Localizing Sustainable Development Goals Across Scales

The Community Human Development Index (CHDI) is a multidimensional measure designed to assess human achievement across scales at the state, county, city, and neighborhood levels to advance the localization of the United Nations Sustainable Development Goals (UNSDGs). Currently, one of the most widely used metrics is the Human Development Index (HDI), a composite index of income, education, and life expectancy calculated at a national level. However, national averages obscure substantial disparities between states, cities and smaller neighborhoods in large urban areas. The CHDI model is designed to gain a better understanding of heterogeneities in human development within cities and across communities on a subnational scale. As a case study, we calculated the CHDI for nearly every census tract in the state of Illinois in the United States. The analysis revealed significant disparities in local levels of development within Illinois. It also highlighted differences between communities within cities, including the city of Chicago. These results demonstrate that multidimensional indices created to track sustainable development at the national level can be disaggregated, calculated, analyzed and compared across scales. The CHDI model can be utilized by researchers and policymakers to identify, visualize and address local inequalities, and design and implement evidence based customized solutions at the community level. [1]

1. The results from chapter 2 were published previously [21]. Much of the chapter text appears as in the published manuscript.

1.1.2 Chapter 3: Measuring and Comparing the Community Human Development of US Cities and Neighborhoods

The impact of the environment on an individual's outcomes remains an area of study for precision public health and for urban sustainable development policy. However, no single integrative framework has emerged to conceptualize and measure human development across scales, with approaches based on human capabilities applying internationally and neighborhood effects locally. Here, we bring these two approaches together by constructing and calculating the Community Human Development Index (CHDI) of US cities and neighborhoods, creating a large dataset characterizing over 70,000 local communities. We demonstrate how to create this index, which allows direct comparison of neighborhoods, cities and nations and any other intermediate scale. We use this metric to analyze patterns of community human development in US cities and neighborhoods. We find a systematic effect of city size associated with higher average community human development but also an even clearer trend towards higher inequality across neighborhoods. We show that increases in CHDI at the local level are associated with the systematic and simultaneous reduction in magnitude and risk of a large and diverse set of neighborhood effects associated with disadvantage, including measures of health, status, opportunity and poverty. We expect that the convergence of the capabilities approach to community human development with studies of neighborhood effects will produce a powerful interdisciplinary synthesis of both theory and practice on a theme of critical societal importance and urgency. [2]

2. The results from chapter 3 were published as a preprint previously [22]. Much of the chapter text appears as in the preprint.

1.1.3 Chapter 4: Community Human Development Indicators in US Counties: An Analysis of the Divergence of Life Expectancies Across the United States Over Time

Measuring community human development over time is critical to determining how changes in environments and policies impact populations and create growth or exacerbate inequality. Previously, the majority of longitudinal development studies of human development have been done at the national scale, leaving major gaps in understanding inequality and how changes at smaller geographic scales impacted the quality of life of individuals. We address this gap in by creating and analyzing the first longitudinal dataset for community human development indicators at the county level in the United States. We systematically characterize the statistics and dynamics of life expectancy, education, and income that have occurred over time at the county scale, illustrating variations in development conditions across the US. We find that while gaps in education and income between counties have not changed much over the last decade, there has been a significant increase in the variance of life expectancy. Gains in life expectancy have largely been concentrated in metropolitan areas while rural counties have lagged, increasing urban-rural disparities over time. Our findings demonstrate that increasing inequalities in life expectancy pose a major challenge for equitable development in the US. Our work also sets a baseline for studying the disruptive effects of the COVID-19 pandemic on human development.

1.1.4 Chapter 5: The Community Human Development Index as a Precision Public Health Vulnerability Metric and Risk Indicator

Accurately assessing and measuring population vulnerability across geographic scales is critical for risk management. Currently, the most widely used metric for vulnerability analysis in the US is the Centers for Disease Control and Prevention's Social Vulnerability Index

(SVI). While the metric has has proven its utility in a number of contexts, its structure has inherent limitations that restricts the way it can be used. Here we propose the Community Human Development Index (CHDI) as a complementary vulnerability metric that can be used to address these gaps. We demonstrate that the CHDI correlates well with the SVI at the neighborhood scale, while also capturing different information about vulnerability of populations. We also show that CHDI predicts community health outcomes as well or better than the SVI, with less variation. Through our analyses, we show that the CHDI can be used as a scalable metric for public health and population vulnerability analysis, allowing for richer, more accurate analyses across a broader range of communities.

CHAPTER 2

THE COMMUNITY HUMAN DEVELOPMENT INDEX (CHDI): LOCALIZING SUSTAINABLE DEVELOPMENT GOALS ACROSS SCALES

2.1 Introduction

Fostering sustainable development requires promoting and assessing a set of diverse objectives across a range of scales [23]. Over the last few years, starting with the Millennium Development Goals, followed by the Sustainable Development Goals [24] and Agenda 2030, policy has acquired a new focus. In general, practical policies and innovations are now asked to be developed over the long term –typically decades – towards transformative efforts such as eradicating extreme poverty (SDG1) or eliminating carbon emissions. Current societal and political approaches to these goals require the creation of composite quantitative indices, involving many different dimensions of sustainability and their evaluation at increasingly smaller scales so that distributional challenges related to inequality can be identified and addressed [23, 25, 4]. This strategy calls for the localization of sustainability goals at the scale of communities (typically neighborhoods) so that minimal standards of living can be promoted everywhere, and not simply based on national or regional averages. Addressing these two issues together: the creation of complex multidimensional indices, and their implementation and measurement across scales with a basis on local communities, are the two main challenges for systematic assessments of progress towards sustainability.

The simplest and most widely used development index with these characteristics is the Human Development Index (HDI). The HDI is the most standard measure of national development that goes beyond economic performance, which is typically measured by real incomes or Gross Domestic Product per capita [4, 26]. It was created to reflect an approach to international development that is philosophically broader and more comprehensive than economic performance, and closer to human capabilities and opportunities [1, 27]. The definition of the HDI has itself changed over time to reflect the concepts underlying this "capabilities" approach to development. It was originally defined as an additive index, where components measuring real incomes, health and educational attainment summed up to a total that could then be ranked across nations [4]. This implied the substitutability between these three different dimensions of development, so that, for example, higher incomes could compensate for poorer health, which was considered non-sensical. Thus, subsequent versions of the HDI built the composite index as a geometric mean of components, implying that all dimensions are necessary for well-being [5].

This will also be the implementation of our disaggregated index at subnational scales, where new challenges arise, for example, in meaningfully accounting for what real incomes are for households in different locations and creating definitions that can be aggregated and disaggregated across scales, from neighborhoods to nations. Only then can we address some of the criticisms of the current use of the HDI, which according to UNDP are entailed by the fact that the "HDI is an average measure and thus masks a series of disparities and inequalities within countries." [28].

We call our novel metric the Community Human Development Index (CHDI). It is designed to assess sustainable development metrics at various subnational levels. It fills a critical need for standardized, internationally recognized metrics that can operate across scales and assess various kinds of distributional effects [23, 29]. The CHDI can be combined with resilience planning frameworks, such as the TripleRM Model (Risk, Resilience, and Resource Management Model), to build in-depth local assessments of communities [30, 31]. The health components of the CHDI can also be a useful entry point in precision medicine, which, according to the National Institutes of Health (NIH), is "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" [32]. The CHDI can be integrated into health dashboards for improved predictive analytics on health outcomes.

Here, we have constructed the CHDI for the US State of Illinois, implemented at the finest scale possible, census tracts. While previous studies have applied HDI methodologies to calculate state averages or to analyze census tracts within a single city, the CHDI dataset reported here is the first to present development metrics for nearly all the census tracts in Illinois [8, 33]. We present an analysis of results and their interpretation in terms of assessing levels of development in a heterogeneous state and specifically issues of equity across neighborhoods in large urban areas here illustrated by Chicago.

2.2 Methodology

2.2.1 Overview of the Community Human Development Index (CHDI)

The Community Human Development Index (CHDI) is a measure designed to be used at the sub-national or community level, such as the State, County, City or neighborhoods proxied, for example, by census tracts in the United States. Like the HDI defined at the national level, the CHDI is the geometric mean of three separate components: 1) the Community Life Expectancy Index (CLEXI) 2) the Community Education Index (CEDI), which is the arithmetic mean of two sub-indices based on Community Mean Years of Schooling (MYSCom) and Community Expected Years of Schooling (EYSCom) and 3) the Community Income Index (CINI).

There are three indices that must be calculated to arrive at the Community Human Development Index: the Community Life Expectancy Index (CLEXI), the Community Education Index (CEDI), and the Community Income Index (CINI). We deal with each component in turn next.

2.2.2 Calculating the Community Life Expectancy Index (CLEXI)

According to the World Bank, life expectancy at birth "is the average number of years a newborn is expected to live" if mortality trends are constant [34]. The UN currently normalizes population level life expectancy at birth between two aggregate extremes: a minimum of 20 years and a maximum of 85 years. This maximum is expected to change over time, as people live longer in leading countries.

Thus, the Community Life Expectancy Index is calculated using life expectancy at birth for people living in each focus community. Therefore, if LECom is the Community Life Expectancy at birth, then the Community Life Expectancy Index (CLEXI):

$$CLEXI = (LECom - 20)/(85 - 20)$$

For this study, census tract level life expectancies were extracted from the US Centers for Disease Control and Prevention (CDC) and the Robert Wood Johnson Foundation's "US Small-area Life Expectancy Estimates Project" (USALEEP), which generated estimates of life expectancy at birth for most US census tracts over the period of 2010-2015.

2.2.3 Calculating the Community Education Index (CEDI)

The CEDI is the mean of 2 different indices: Community Mean Years of Schooling (MYSCom) Index and Community Expected Years of Schooling (EYSCom) Index.

Calculating the Community Mean Years of Schooling Index

Community Means Years of Schooling measures the adult population (25 years or older)'s mean years of education. As most countries report education levels by final attainment, the International Standard Classification of Education (ISCED) is used to assign each degree/level to a certain number of years of schooling.

If Ts is the portion of the adult population who attained education level s, and Ns is the number of years necessary to achieve education level s as designated by the ISCED, then Community Mean Years of Schooling (MYSCom):

$$MYSCom = \sum_{s} T_s \times N_s$$

The Community Mean Years of Schooling Index (MYSICom) can then be calculated by normalizing MYSCom to the UN target goalpost of 15 years:

$$MYSICom = MYSCom/15$$

For this study, education levels for each census tract were extracted from the US Census' American Community Survey (ACS) 5-year estimates for 2012-2017. The ACS reports estimates for total tract population as well as educational attainment for the population 25 years or older, organized by degree.

Calculating the Community Expected Years of Schooling Index

The Community Expected Years of Schooling is the estimated number of years of education a student entering school today will receive based on current enrollment rates. If a is the number of students in a particular age group, and c is the total number of children in that age group, then r_x,, the school enrollment rate of that age group, is:

$$r_x = a/c$$

If b is the lower age limit of all students in the school system, d is the upper age limit of all students in the system, and y is the number of years a particular age group is expected to be in school, then the Community Expected Years of Schooling (EYSCom):

$$EYSCom = \sum_{b}^{d} y \times r_x$$

The Community Expected Years of Schooling Index (EYSICom) can then be calculated by normalizing EYSCom to the UN Goalpost of 18 years:

$$EYSICom = \frac{EYSCom}{18}$$

For this study, student populations and enrollment rates were calculated based on data from the US Census' American Community Survey 5-year estimates for 2012-2017. The ACS reports the total number of residents in each age cohort in a tract, and the number of students enrolled in school by school level and age cohort. Some census tracts do not have residents in every age cohort, which is required to accurately calculate the EYSCom. To account for this, the EYSICom was calculated at the county level and then imputed to all tracts within the county.

Calculating The Community Education Index

The Community Education Index is generated by taking the average of the Community Mean Years of Schooling Index and the Community Expected Years of Schooling Index:

$$CEDI = \frac{MYSICom + EYSICom}{2}$$

2.2.4 Calculating the Community Income Index (CINI)

The UN measures economic standard of living by taking the natural logarithm of the Gross National Income, adjusted for Purchasing Power Parity, per capita (GNIppc). As GNI is a national economic measure, there is no direct equivalent on the local scale. In order to keep values consistent across scales, it is necessary to down-allocate GNI (dGNI) using another economic measure representative of economic performance, such as per capita income, and then divide it by the regional population to calculate the Community Income Index (CINI) for the CHDI:

$$CINI = \frac{(ln(dGNI_{ppc}) - ln(100))}{(ln(75,000) - ln(100))}$$

For this study, tract-level GNIs (GNI_t) were calculated by down allocating the 2017 US GNI PPP (constant 2011 international dollars) reported by the UN Statistics Division using total tract incomes reported by the American Community Survey (ACS) 5-year estimates for 2012-2017:

$$GNI_t = 2015\ US\ GNI \times \frac{Census\ Tract\ Aggregate\ Income}{Total\ US\ Aggregate\ Income}$$

In order to make quality-of-life comparisons across metropolitan areas, tract-level GNIs were then adjusted based on Regional Purchasing Power Parities (PPP) reported by the US Bureau of Economic Analysis [35]. This produced regionally adjusted tract-level GNI (GNIat):

$$GNI_{at} = GNI_t \times \frac{1}{Regional\ PPP}$$

Tracts that were not located in metro-areas were adjusted using the non-metro area PPP reported by the Bureau of Economic Analysis. The adjusted tract-level GNI values were then divided by tract population estimates reported by the ACS 5-year 2012-2017 survey.

2.2.5 Calculating the Community Human Development Index (CHDI)

The Community Human Development Index (CHDI) is calculated by taking the geometric mean of the three indices:

$$CHDI = \sqrt[3]{CLEXI \times CEDI \times CINI}$$

2.3 Results

2.3.1 Description of the Case Study

Selection of Illinois for Case Study

The state of Illinois, where we are based, was selected for a case study in the application of the CHDI. With a combination of urban (city of Chicago) and rural areas (downstate), as well as a diverse set of populations, Illinois is demographically similar to US as a whole [36]. Additionally, Chicago has many distinct communities, which substantially vary in crime rates, health outcomes, personal income, and educational attainment levels [37].

At the Community Level, the CHDI was calculated for census tracts in Illinois, each of which represents 4000 people on average. Census tracts are grouped together to form counties, and counties are grouped to form cities and states, Cities were defined using the United States Office of Management and Budget's definition of Metropolitan Statistical Areas, which represent functional urban areas (integrated labor markets) with strong economic ties and relatively high population density. Metropolitan Statistical Areas are by construction aggregations of counties, and are updated after each decennial census [38].

As noted in the methodology, a variety of sources were used to calculate the CHDI for the census tracts in the state. The US Census Bureau's American Community Survey 5-year estimates for 2012-2017 were used to build both the CEDI and the CINI. The CINI also utilized Gross National Income and Purchasing Price Parity data from the US Bureau of Economic Analysis. The CDC's USALEEP study was used to calculate CLEXI scores. It is important to note that the USALEEP study did not calculate life expectancies for all census tracts, as some do not have a significant population (such as the one that contains O'Hare International Airport) or a complete population structure to calculate life expectancy at birth.

The CHDI reveals the heterogeneities across local communities, illuminating differences that are often obscured by larger scale averages.

2.3.2 Data Analysis and Description of the results

In a global context, the US has a relatively high HDI, mainly due to high income levels. While it has relatively high-income ranking, both education and life expectancy stand to be improved relative to other highly developed nations. Life expectancy in particular lags behind international standards: the US is ranked 37th in the world, behind nations that in other dimensions are much less developed, such as Chile, the Czech Republic, Cyprus, and Costa Rica.

CHDI Rankings of Cities in Illinois

For our case study, we have opted to use the international goalposts (normalizations) set by the UN. This was done to allow for comparisons between the CHDIs calculated for census tracts and the HDIs calculated at the national level. In this way we can compare the same numerical score for the CHDI obtained, for example, for a neighborhood in Chicago to the HDI in various nations, for which levels of development have been characterized at greater length.

Rankings of cities in Illinois reveal the importance of applying human development metrics at a local scale. The rankings of the cities clearly show a shift in the economy and geographic distributions of wealth. Both Champaign-Urbana, IL and Bloomington, IL are small metropolitan areas with large higher education institutions. This has the effect of increasing the area's Community Mean Years of Schooling Index scores, as the adult population around a college tends to be highly educated. However, the presence of a University does not appear to have a large impact on local life expectancy, as CLEXI scores only increase slightly. The city of Chicago is ranked as having the third highest average CHDI, and is analyzed in-depth in the next section.

Conversely, the three Illinois metropolitan areas with the lowest CHDIs are former manufacturing centers. As many jobs in the manufacturing industry do not require higher degrees, the Community Mean Years of Schooling Index tends to be lower, bringing down the overall CEDI score. Additionally, broader shifts in the manufacturing industry and factory closures have caused the local economies of these metropolitan areas to become economically depressed, bringing down incomes measured by the CINI scores. These issues indicate that former manufacturing hubs may be revitalized by investing in educational resources, such as community colleges. Low CLEXI scores can be addressed by increasing access to medical primary care and addressing the local impacts of broader public health issues, such as the opioid crisis.

Overview of Chicago CHDI Results

On average, census tracts in Chicago showed a high CHDI, with a mean of .9421, roughly equivalent to Illinois' average CHDI of .9363. Both CEDI and CINI scores were significantly higher than the US national averages, likely due to the concentration of universities and major businesses in the city.

The application of Purchasing Power Parity at the local level also had the effect of lowering the CINI scores of Chicago and other larger and wealthier metropolitan areas in the state. This reflects the economic reality that the same income often offers less purchasing power in larger metropolitan areas, which tend to be more expensive, for example in terms of housing and local services, than in rural areas.

CHDI values varied widely within the city, with the highest and lowest CHDI tracts in all of Illinois found within Chicago. This reflects both the wealth and education in certain parts of the city, and the poverty concentration in others. In general, high CHDI tracts were located on the North Side and in the Central Business District ("The Loop") of Chicago, while lower average CHDI values were found for the South and West sides, which are poorer and generally plagued by environmental challenges and high crime rates.

The highest CHDI value was 1.1003, reported Tract 8001 in Cook County. All three tracts with the highest CHDI values in the state are located in Glencoe, IL, a northern suburb of the city of Chicago. The lowest CHDI was found in Census Tract 8435, perhaps not surprisingly the location of the Cook County jail.

Neighboring tracts also have extremely different CHDI values. For example, Census Tract 17031836300, located in the Hyde Park neighborhood of Chicago (the location of the University of Chicago, where we work), has a CHDI of .9811, higher than that of any nation in 2018. A neighboring census tract 17031836100, has a CHDI of .8653. The same tract also has a life expectancy at birth of about 70.5 years, equivalent to the life expectancy of birth in Iraq in 2019.

2.4 Conclusion

The Community Human Development Index (CHDI) is a composite index that can be used to measure education, income, and life expectancy outcomes at a community level. The three components of the HDI, income, education, and life expectancy have all been found to be correlated with one another to some extent. Through further research, we can study how strongly these correlations are observed at local scales, and how the correlations differ across communities.

Using the CHDI model to analyze data, it is possible to construct multidimensional indices to measure development across scales. An analysis of census tract data in Illinois showed large disparities in development at multiple scales. The highest CHDI tracts in Illinois were found to have development levels higher than any country in 2019, while only a few miles away, the tracts with lowest CHDI values had values roughly equivalent to Cuba.

In calculating a localized version of a national index, strategic decisions are required to make the index both locally relevant and scalable to higher levels. There are challenges in localizing indices and rendering them meaningful across scales, from local communities to the global scale. For example, researchers have to decide how to down-allocate national measures, such as Gross National Income, and whether or not to account for regional differences in purchasing power parity. Additionally, calculating measures such as expected years of schooling or life expectancy at birth is only possible in communities that have populations in every recorded age cohort. The US has some of the best disaggregated data and small-area measures in the world. Some nations may have similar data standards, but the majority of developing nations do not at present. This makes the application of a localized index difficult in such contexts. Future studies should look into how small-area measures can be estimated in nations with lower levels of data resolution, where a wider range of proxy quantities may also play a role.

The CHDI can serve as a useful tool for Precision Medicine and advancing UNSDG 3 (Good Health and Wellbeing), as well as the World Economic Forum's goal of improving Global Health and Healthcare. The long-term goal of the NIH's Precision Medicine Initiative is to integrate precision medicine approaches into all areas of health and healthcare. Previous studies have shown that the location of a patient's residence within a city can be predictive of their long term health outcomes [39]. As a result, precision medicine approaches often analyze patients' socioeconomic backgrounds as well as their genetic backgrounds and biological baselines. The CHDI can be used as an indicator to predict health outcomes. It can also be used to better target public health efforts within large populations to study, predict and prevent health risks, and reduce the health outcome disparities between neighborhoods. The CHDI can be combined with health datasets, such as those produced by the All of Us Research Program, to develop insights about the interplay of genetic background, socioeconomic background, and chronic disease [40]. The CHDI can also be integrated with patient Health Dashboards to give a broader context to care, allowing physicians to develop treatment plans and timely interventions that are most effective to a patient's specific needs in predictive and preventive medicine.

The CHDI, as an indicator of the well-being of a community, is designed to aid local risk assessment, participatory strategic planning for resilience, and optimal service delivery and resource allocation and management at a community level. The CHDI can utilized in conjunction with global open-source platforms designed to accelerate the delivery of the UNSDGs, such as the World Economic Forum's "UpLink" platform for a multi-stakeholder approach at a grassroots, community level with local leaders and innovators. The CHDI can also be a useful tool in the World Bank's strategy to aid in the advancement of the UNSDGs on a local level.

Overall, this study shows that a detailed examination and understanding of the heterogeneities in human development within cities and across communities is key for the meaningful attainment of international, national, state, city and countywide sustainable development goals and for a robust understanding of the underlying reasons contributing to such outcomes. The CHDI model shows that building indices of development and measuring achievement at the community scale is not only possible, but also necessary to help transform the UN2030 global vision into a reality at every level.

	Metropolitan Statistical Area	CHDI Rank	CHDI	CEDI	CINI	CLEXI
Top 3 Cities in Illinois	Champaign-Urbana	1	.9765	.9881	1.0267	0.9192
	Bloomington	2	.9646	.9668	1.0322	0.9003
	Chicago-Naperville-Elgin	3	.9421	.9236	1.0054	0.9019
National Average	USA	N/A	.920	.899	.956	.905
Bottom 3 Cities in Illinois	Decatur	9	.9175	.8881	1.0037	.8674
	Rockford	10	0.9119	.8765	.9966	.8689
	Danville	11	.8962	.8485	1.0015	.8478

Note: This ranking is focused on cities where the majority of census tracts fell within Illinois state lines. As such, Cape Girardeau, Davenport-Moline-Rock Island, and St. Louis were excluded.

- **CHDI: Community Human Development Index**
- **CLEXI: Community Life Expectancy Index**
- **CEDI: Community Education Index**
- **CINI: Community Income Index**

Table 2.1: CHDI of Cities in Illinois

	Country	2019 HDI Rank	HDI	Education Index	Income Index	Life Expectancy Index
Countries with Highest HDI	Norway	1	0.954	.919	.985	.958
	Switzerland	2	0.946	.896	.965	.979
	Ireland	3	0.942	.918	.955	.955
	USA	15	.920	.899	.956	.905
	Illinois	N/A	.9363	.9141	1.0049	.8949
Countries with Lowest HDI	Chad	187	0.401	.288	.429	.523
	Central African Republic	188	0.381	.353	.310	.505
	Niger	189	0.377	.247	.334	.647

Note: HDI worldwide rankings are based on values from the UNDP 2019 Human Development Report

Table 2.2: Comparisons of US and Illinois CHDI with HDI values of Countries

	Tract Number, County, MSA	CHDI Rank	CHDI	CEDI	CINI	CLEXI
Top 3 Census Tracts in Illinois	Tract 8001, Cook County, Chicago-Naperville-Elgin	1	1.1003	1.0307	1.2355	1.0462
	Tract 8003, Cook County, Chicago-Naperville-Elgin	2	1.0955	1.0270	1.2076	1.0600
	Tract 8006, Cook County, Chicago-Naperville-Elgin	3	1.0916	1.0418	1.1969	1.0431
National Average	USA	N/A	.920	.899	.956	.905
State Average	Illinois	N/A	.9363	.9141	1.0049	.8949
Bottom 3 Census Tracts in Illinois	Tract 3406, Cook County, Chicago-Naperville-Elgin	2917	.8065	.8298	.8579	.7369
	Tract 6903, Cook County, Chicago-Naperville-Elgin	2918	.7794	.8679	.8888	.6138
	Tract 8435, Cook County, Chicago-Naperville-Elgin	2919	.7753	.8294	.6641	.8462
CHDI: Community Human Development Index **CLEXI: Community Life Expectancy Index** **CINI: Community Income Index** **CEDI: Community Education Index**						

Table 2.3: CHDI Values of the Highest and Lowest tracts in the United States

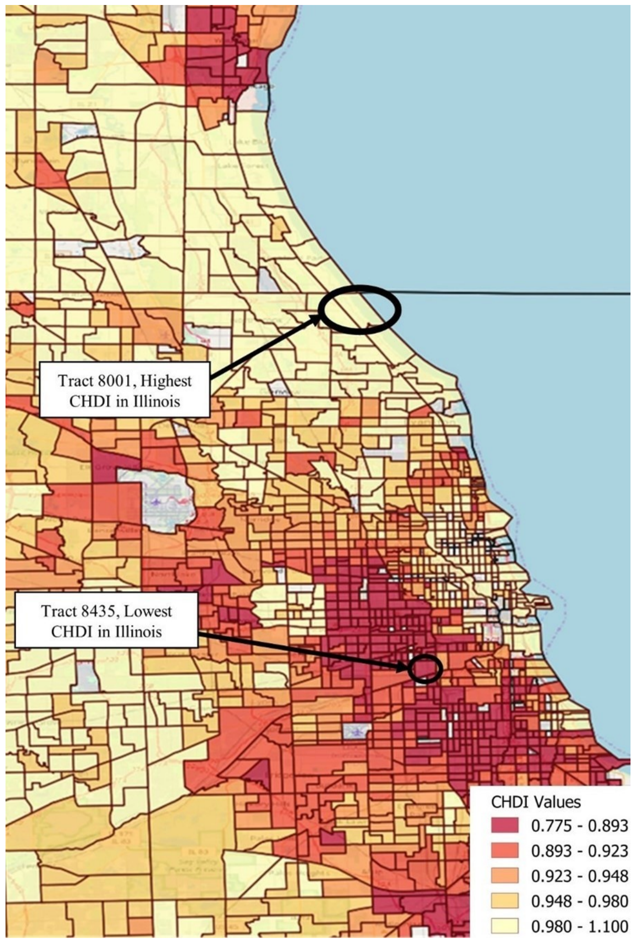


Figure 2.1: Map Showing CHDI values for Census Tracts in Chicago, overlaid a map from OpenStreetMaps. Open spaces are tracts which were not included in the CDC USALEEP Study, such as O'Hare Airport, for which life expectancies could not be calculated. Map shows major disparities in development, with higher CHDI values in the North Side and the central Loop Business District of the city, and relatively low CHDI values on the South and West Sides of the City.

CHAPTER 3

MEASURING AND COMPARING THE COMMUNITY HUMAN DEVELOPMENT OF US CITIES AND NEIGHBORHOODS

3.1 Introduction

Human development remains a major challenge in a world characterized by fast, multifaceted change but also by profound inequalities [41, 23, 42]. Better understanding the underlying processes supporting human development and how they can be accelerated in sustainable ways is a fundamental objective for both science and practice [1, 43], translated into local, national and international policy objectives and commitments.

Over the last few decades two main approaches to human development have emerged at two very different scales. The first was inspired by the work of Amartya Sen [44, 1] and Martha Nussbaum [43], who defined human development in terms of *capabilities*. This approach draws on a deep history of ideas about ethics and moral philosophy [44, 45, 46] as well as on recent data and evidence in many diverse situations worldwide, including a keen awareness of the local conditions of poverty, gender, ethnicity and disability in specific contexts [1, 41, 47]. The capabilities approach to human development inspired holistic measures of progress beyond national accounts (e.g. GDP), culminating in the definition and refinement of the well known Human Development Index (HDI) [46, 4, 5]. The HDI has been measured at the national level since 1990 in annual reports by the United Nations Development Programme (UNDP). The index measures the enabling factors of a long and healthy life, access to knowledge, choice and a decent standard of living [46, 41]. While there have been lively debates regarding the HDI's ability to capture the full complexity of human capabilities [4], and there have been improvements in this metric's construction over time, it has become broadly used by NGOs and national governments alike to measure and compare development levels across very different contexts. Because of its widespread adoption worldwide, the HDI has arguably become the gold standard simplest metric for measuring development at the national level and, as such, it has inspired a race to the top among nations to lead rankings and improve the living conditions of their populations [46, 41]. However, there is clearly a large gap between measuring the HDI at the aggregate scale of nations and the concept of human capabilities, which is tied to local environments where people grow up, live and work. Addressing this mismatch is an active research and policy goal [8] with great promise to connect the rich conceptual framework of capabilities to diverse local outcomes, including inequality, local public health, consequences of urbanization and the distributional effectiveness of social policies [4, 5, 8, 48].

The second approach to studying human development has been more local, with a focus on place-based communities ("neighborhoods") and on the challenge of inequality and segregation, especially in cities. The study of these *neighborhood effects*, as the field became known, has been a major theme in the social sciences for over a century [49, 50], but it gained special importance since the 1980s with the work of William Julius Wilson [51, 52], against the background of deindustrialization in US cities, mass unemployment concentrated among working class Black communities, and the formation of inner city marginalized neighborhoods exhibiting compounding forms of social disadvantage. The literature on neighborhood effects has since grown to give an interdisciplinary account of concentrated local disadvantage that includes sociological, economic, developmental and health considerations [52, 50, 53, 54, 55].It has also identified many different indicators of social disadvantage, with an emphasis on aspects of local human "ecological" effects [52, 49, 50], including concentrated crime [37, 56, 57, 58, 59], school performance [60], trust and collective action [37, 50, 57], and racial and ethnic composition and segregation [51, 52, 50]. Because of its interdisciplinary nature and diverse metrics, the study of neighborhood effects has remained far from unified conceptually, especially in terms of the identification of the web of causal processes that create and maintain cumulative local disadvantage [49, 53, 50, 54]. As a result, policy approaches have remained somewhat narrow and arguably ineffective, including built environment interventions to mitigate disorder [58], crime prevention programs [61, 57, 58], rent vouchers to abandon troubled neighborhoods [62, 49, 54], non-profit social support [57, 59] or cognitive treatment of youth at risk [61, 54]. Recent empirical work using more extensive data sources, such as tax records, has better established the critical importance of temporal exposure to disadvantaged neighborhoods, especially during childhood, with consequences for the life course of individuals including their future income, family structure, and health [63, 64, 59, 54]. Nevertheless, there is more agreement in this literature about data and statistical methods than about causes and solutions, with the scope of human development necessary to address systemic local disadvantage attributed to very different scales, from individuals or communities, to non-profits and public institutions.

To connect these two approaches, we must bring metrics and analyses of human development to the same local scale, reflecting the human experience on a daily basis [1, 43]. Neighborhoods are critical in this sense, because they tie together outcomes to the local environments where people reside, go to school and organize socially to deal with issues of liveability, health and safety [65, 30]. In this way, they represent an ideal scale to measure how local environments enable or inhibit human capabilities. As a consequence, we hypothesize that increases in local human development –measured by the HDI– must be associated with declines in most measures of socioeconomic disadvantage. While we expect these effects to be visible at all scales, we also hypothesize that we should see stronger associations at the neighborhood level and large variations in human development across local communities within large cities, expressing known inequality of neighborhood effects in the more holistic language of human development.

To test these hypotheses, we first show how the HDI can be localized at the neighborhood scale. To this end, we create a novel large dataset characterizing the human development of all cities and neighborhoods in the US, with over 70,000 census tracts. We then develop the methodology to consistently measure the HDI across scales, allowing the direct comparison of human development in neighborhoods, cities, states and nations. We analyze the HDI values across scales in the US to show that human development is typically associated with larger urban areas, but that these indeed present greater inequality between their local communities (stronger neighborhood effects). Finally, we show that high human development in neighborhoods anywhere is associated with the simultaneous and systematic reduction of most forms of social disadvantage both in terms of expected values and risk.

3.2 Results

We now show how human development can be defined and measured consistently down to the scale of local communities, Figure 3.1A. By extension, human development can also be measured at any intermediate scale, including states, metropolitan areas (functionally defined cities) and counties, Figure 3.1B.

3.2.1 Measuring Community Human Development Across Scales

Despite its implementation as an international standard by the UNDP since 1990, there is nothing special about measuring the HDI at the national level: All quantities involved – and their intended significance in terms of human capabilities– apply more meaningfully at the level of households or individuals. Starting from smaller scales will allows us to build up an understanding of collective effects emerging from different local human ecologies and urban network effects [66]. Analyses at local scales also give further insight into how population sorting and filtering processes occurring at different scales impact human development and may lead to place-based inequalities.

It is therefore critical to preserve consistency of the HDI estimation across scales so that we can compare average levels of national development to those of specific small local communities. Figure 3.1A shows how the HDI is defined as a composite measure of i) a long, healthy life, ii) access to knowledge and iii) economic choice and a decent standard of living. These objectives are in turn measured in the latest implementation of the HDI as an international standard via life expectancy at birth, educational attainment and real economic incomes. Although these quantities are positively correlated (Figures 3.4 and 3.5), there remains a large amount of variation unexplained, especially when taken at the scale of local communities (census tracts).

To measure human development as an index (i.e. a number of order 1), these three input quantities are normalized to given international goalposts to form three subindices - I_{LE}, I_E, I_{RI}, respectively. The HDI is then given as the geometric mean of these three normalized components, $HDI = (I_{\mathrm{LE}} \times I_{\mathrm{E}} \times I_{\mathrm{RI}})^{1/3}$, meaning that each of the components is considered essential for high human development and that they are not mutually substitutible [4]. This makes the HDI different from many commonly used indices of vulnerability and socioeconomic status built out of principal component analyses [67] of variables such those in Figure 3.3, which are weighted but additive (substitutible). Therefore, in the context of the HDI, a population that has high income but poor health (or education) will rank low. This is the main reason why the US, despite high mean income, ranks 17 in the world by HDI in 2020, behind many poorer nations.

The life expectancy index, I_{LE} is calculated simply as an indexed value of life expectancy at birth LE, normalized to a maximum of 85 years and a minimum of 20 years, $I_{LE} = (LE - 20)/(85 - 20)$. The education index is made up of two subindices, accounting for mean years of schooling (MYS) and expected years of schooling (EYS). The mean years of schooling index applies to adults (≥ 25 years old) and is computed as the population average, $MYS = \sum_s n(s) \times Y(s)$, where $n(s)$ is the fraction of the adult population who attained education level s, and $Y(s)$ is the number of years necessary to achieve such education level designated by the International Standard Classification of Education (ISCED) [68]. The expected years of schooling index applies to younger populations ($<$ 25 years old), who may still be in school. It is estimated as the expected mean final educational attainment if current school enrollment rates hold, $EYS = \sum_a n_r(a) \times Y_e(a)$, where a is age, $Y_e(a)$ is the expected total years of schooling for age cohort a and $n_r(a)$ is the fraction of the population enrolled in school at age a. The education index is the result of averaging these two subindices with given international goalposts, $I_{\mathrm{E}} = 1/2(EYS/18 + MYS/15)$. Finally, the real Income Index, I_{RI} is usually calculated at the national level using Gross National Income (GNI) per capita (gni), which has no simple subnational equivalent. To create a meaningful definition in small areas, we use average personal income per capita data, I_{pc}, as $\hat{gni} = c_{\mathrm{GNI/I}} I_{\mathrm{pc}}$, with $c_{\mathrm{GNI/I}} = GNI/I$, the ratio of national GNI to total personal income. To create real incomes, we adjust nominal incomes for cost of living at the local level. The US Bureau of of Economic Analysis publishes a local purchasing power parity index (PPP) for metropolitan areas and states, which we use as $\hat{\mathrm{gni}}_{PPP} = c_{\mathrm{GNI/I}}\ I_{\mathrm{pc}}/PPP$, which now applies to each level of geographic aggregation, including states, metropolitan areas and tracts. The income index is then $I_{RI} = (\log \hat{gni}_{PPP} - \log 100)/(\log(75,000) - \log(100))$, normalized to a minimum of \$100 real dollars per person/year and to a maximum of \$75,000. This upper value is commonly surpassed in US census tracts, resulting in a values for the local income index > 1. In their latest report [47], UNDP capped gni at this upper bound for three city states so that it would not dominate the overall HDI, which we do not do here. Moreover, the use of logarithms in the income index follows UNDP construction conventions, recognizing the broad statistics (typically lognormal) of personal income in modern societies and smaller marginal benefits at high incomes [66, 41].

A previous study found that the HDI's even statistical weighting of its three components accurately captured the degree of variation between the components across years at the national level [69]. To assess whether this was true at the census tract level, we performed a Principal Components Analysis of the three sub-indices. We found that the first principal component (PC1) was positively correlated with all three sub-indices, and accounted for 72.5% of the variance. The normalized weights of the Income Index, Education Index, and Life Expectancy Index were .35, 0.36, and .28 respectively, for PC1. This is close to equal weighting (which would be 0.33) of the three components, supporting the use of the HDI equation to aggregate these three quantities at the tract scale.

The consequences of these definitions are transparent and algorithmic. We provide code and data (see Materials and Methods) to allow readers to replicate our definitions and analysis, as well as generalize them to other contexts. Given appropriate data, analogous calculations can be easily developed for other nations and over time, which may improve on a few limitations of present datasets, see Materials and Methods.

3.2.2 Community Human Development in US Cities and States

We now compare the HDI variation cross-sectionally across scales including states, metropolitan areas and neighborhoods. The general HDI statistics in census tracts across the entire nation is very well described by a normal distribution, Figure 3.1C. We also tested a number of alternative statistical descriptions, Table 3.1.

The consistent construction of the HDI across scales allows us to compare human development in each local community to nations or to the temporal trajectory of US development. We see in Figure 3.1C that while Norway (the top nation by HDI in 2020 [47]) has a larger HDI than the US average, about 19.6% of the US population (and 18.5% of census tracts) exceed this level. Similarly, 6.5, 0.7 of the US population and only 0.2% (7.9, 0.9, 0.3% of tracts) lives at standards of development below Russia, Mexico and China, respectively. The discrepancy between percent of population and of tracts at each level reflects the fact that very high HDI communities tend to be part of larger cities and have more populous tracts.

Despite the simplicity of the national tract-level HDI statistics in Figure 3.1C, the same is not observed at smaller scales reflecting regional and local biases to lower or higher human development. Figure 3.1D shows that there is a general statistical tendency for larger US metropolitan areas to display higher levels of human development. An even clearer trend, however, is the rising inequality in human development, measured by the Gini coefficient between tracts, Figure 3.1E. These urban scaling effects of population size [23, 66] are very general and apply also to each HDI subindex and are even larger if income is not adjusted for local purchasing power in larger, more expensive cities, Figures 3.6 to 3.10. Together these statistical trends mean that while larger cities tend to provide better environments for human development overall, they also show a widening gap between their local communities, an issue discussed below under the theme of neighborhood effects.

As illustrations of higher HDI and greater inequality with city size, consider that the ten most developed US local communities, with HDI>1.1, all are part of only three large metropolitan areas. New York City has the greatest number (six), located by New York University and Washington Square and by Central Park, followed by Washington DC (three). The single highest HDI$=1.13$ community in the US was in the Boston-Cambridge-Newton metropolitan area (Coolidge Hill, by Harvard University). By contrast, the tracts with the lowest HDI nationwide are not what we may think of as standard communities. Most are dominated by institutions concentrating disadvantaged populations, particularly jails, but also asylums and rehabilitation centers. For example, in New York City metro, the census tract with lowest HDI in the nation contains the Northern State Prison (and Newark airport) with HDI$=0.443$. The other two are Sing Sing Correctional Facility and Rikers Island. In almost every region, the tracts with the lowest HDI are jails and other institutions that, for various reasons, disproportionately concentrate disadvantage.

These patterns of high and low human development have interesting expressions when used for ranking US states and urban areas, Figure 3.2. US states are significant units of analysis because many important policy decisions occur at this scale, from health care and education to land use and transportation [70]. The top US state for HDI is Massachusetts with $HDI = 0.967$, significantly better than any nation worldwide. By contrast, the bottom HDI state is Mississippi with a value of 0.876, similar to Poland, Figure 3.2. These differences fall along urban-rural differentials in HDI, but are also likely to reflect state level policies concerning education, healthcare and other facets of human capabilities.

For metropolitan areas, Boulder, CO tops the rankings with HDI= 0.982 as the result of high performance in all HDI components, with education and income index values both higher than 1. There is a general "college town effect" driving the highest performing cities irrespective of geography, meaning that cities that concentrate higher education and research tend to do very well. Other examples include Corvallis, OR, and Ames, IA, Silicon Valley (San Jose, CA metro) and Boston-Cambridge. Among the smaller micropolitan areas, Los Alamos, NM (home of the eponymous national laboratory) has the highest HDI of any micro- or metropolitan area in the US with HDI= 1.009.

It is interesting to note that the high HDI college town effect is not purely local as it presents spillover effects into surrounding communities. In fact, college campuses tend to score low on HDI because of a strong concentration of students with little or no income. However, the HDI of surrounding census tracts tends to be elevated, likely because colleges and research institutes create environments that attract other highly educated and entrepreneurial people. This effect is not only observable for micropolitan and small metropolitan areas with large colleges and research facilities (like Boulder and Los Alamos), but also in the local areas around such institutions in larger metropolitan areas, as noted above for Washington Square in New York City and Coolidge Hill in Boston. By contrast, the lowest ranked metropolitan areas, such as Gadsden, AL and Lake Havasu City-Kingman, AZ, tend to be (post-)industrial, agricultural or border cities, mostly in the South and Southwest. These cities perform relatively well in terms of the Income index; Gasden, AL, has an income index value of 0.97. Their relative challenges of human development concentrate on deficits of life expectancy and education.

We see that HDI rankings of US urban areas and states naturally lead to a rich set of comparative analyses revealing consequences of policy, local history and culture across scales, as well as collective socioeconomic dynamics associated with scaling and agglomeration effects in urban areas and their constituent communities to which we now turn.

3.2.3 Community Human Development and Neighborhood Effects

Studies of neighborhood effects start by recognizing and measuring i) inequalities between different neighborhoods, especially in major North American cities [56]; ii) that inequality is place-based, persistent and self-reinforcing [71], creating vicious cycles of cumulative (dis)advantage, and iii) that the effects of neighborhood environments on people are "ecological" and cumulative in terms of temporal exposure, especially during childhood [52, 37, 64, 54, 60]. Moreover, the effects on individuals are complex, involving interlocked social, economic, cognitive and behavioral outcomes, including mental health [56], educational attainment [60, 56], physical health [55, 72, 73], crime [59, 72], lack of trust and other measures associated with social disadvantage [37]. Neighborhood effects are clearly identifiable through spatially resolved maps of socioeconomic outcomes in any city, which often manifest variations within short distances of about a kilometer [23].

We propose that the capabilities approach to human development provides a theoretical frame of reference for this body of knowledge and practice. To do this, we show that greater HDI observed at the scale of neighborhoods is associated with systematic and simultaneous improvements in a large number of diverse socioeconomic indicators, which we characterize statistically. We assembled a set of different metrics for tracts nationwide across a variety of representative neighborhood effects studies, including those by the Opportunity Atlas [63], the US Census American Community Survey and the Centers of Disease Control PLACES dataset, see Materials and Methods. Though some of these metrics relate to HDI inputs (i.e. children in poverty), most reflect outcomes expressing distinct facets of disadvantage at different life stages, as well as segregation and poverty, Figure 3.3. Examples are teen pregnancy, poor mental health, or incarceration.

As an example, Figure 3.3A-B illustrates the typical relationship between HDI and teenage pregnancy rates at the tract level across all places in the USA (71,513 tracts). Figure 3.3A highlights tracts from three different metropolitan areas - Memphis TN, Phoenix AZ, and San Francisco CA - with low, medium and high average human development, respectively. The average relationship between teenage pregnancy rates and HDI shows a clear negative correlation reflecting both the decrease in the mean expected rate and its variance with larger community HDI. We characterized this behavior in three ways. First, by a simple linear regression fit (blue line) that captures most of the average variation and, second, by a statistical model of rates as a Beta distribution, with parameters dependent on the HDI (red and green lines), see Supplementary Text for motivation and explanation. In addition, we also performed a factor analysis of this set of variables.

The linear fit gives us the essence of this variation: average teenage pregnancy rates, $p(\text{HDI}) = p_0 - a \times \text{HDI}$ at the tract level display a negative slope of $a = 1.671$ versus HDI. This linear fit predicts that at HDI$= 1.02$, teenage pregnancy rates should vanish. This is not quite correct because this decrease slows down at very high HDI. The Beta distribution model, $p(\text{HDI}) \sim \text{Beta}(\alpha, \beta|\text{HDI})$, gives a much better description of data both in terms of the HDI dependence of the mean (solid red line) and its variance (dashed lines) or 99% confidence interval (green). To estimate this distribution, we binned census tracts according to their HDI values and estimated the Beta parameters (α, β) for each bin, Figure 3.30. We then fitted the variation of these parameters with HDI to produce a continuous description of the conditional statistics (red and green lines in Figure 3.3B). In most cases, the variation of the distribution average and standard deviation is well fit by a logistic curve (solid red line), which approximates a linear relationship at mid-HDI (blue line), while saturating to zero at high HDI and slowing down their increase while also increasing variation at low

HDI. Tables 3.2 and 3.3 summarize linear fits for various quantities at the metropolitan and micropolitan level, illustrated in Figures 3.11 to 3.19. Tables 3.4 and 3.5 summarize results for the same quantities at the tract level, illustrated in Figures 3.21 to 3.29. In addition, we show (Figures 3.20 and 3.32 and Supplementary Text) that race, ethnicity, or being foreign do not show strong correlations with HDI, mainly because low HDI communities have many different compositions at the national level [74]. Greater local shares of Black and Hispanic populations become more associated with lower HDI in more specific urban environments –such as Baltimore or Chicago– where neighborhood effects have been studied in association to racial segregation [51, 37, 59, 75].

In addition to linear regressions and Beta density modeling, we have also performed a principal component analysis (PCA) of the joint 18 variables in Figure 3.3 to show that HDI increases correlate to simultaneous decreases in all social disadvantage rates expressed as the first PCA component, which accounts for 54.5% of the relative variance, Table 3.6.

Figure 3.3C shows the slope of the average variation a in each of these rates with HDI, see also Figure 3.3D inset. Normalizing these rates by their values at low HDI obtains a simple general picture where all rates decrease as human development increases and vanish around or slightly above HDI= 1, Figure 3.3D. This expresses a universal trend -regardless of geography or local culture- towards the systemic mitigation and even eradication of many apparently different societal challenges at high levels of human development, measured by the HDI.

3.3 Discussion

We have shown how human development can be measured consistently across scales from nations down to local neighborhood communities. In this way, the HDI's historical role as the major indicator of socioeconomic progress among nations and its meaning signaling expanding human capabilities can be brought to bear on challenges of local development at the neighborhood and urban scales. This approach provides an integrative framework for the concept of human capabilities and their formation through a person's life course with the challenge of neighborhood effects and associated design and evaluation of sustainable development policies in cities.

Consistent with this integrated picture, we have found that larger cities tend to be sources of higher human development but are also associated with starker inequalities between their neighborhoods. The typical places with the highest human development are not necessarily the richest, but rather those with more intensive activity in higher education and research regardless of geography, a finding we called the "college town effect." These places allow us to visualize what a future of more widespread human development may look like. Conversely, a very different kind of place concentrates disadvantage, especially jails and other institutions, with public housing projects also often in this category. These extremes illustrate the environments that create and destroy human capabilities, as sources and sinks of human development.

Progress in theory and practice of local human development will necessarily require measurement and analysis across time, which has been already transformative at the national level [46, 41, 47]. Such longitudinal analyses remain a challenge given present data collections and, especially, the manner in which they are dispersed across different organizations. The creation of a local community HDI in the US and its relation to other tract characteristics over time hinges on an enormous statistical effort to coordinate and integrate a number of different data sources, including the US Census, the Bureau of Economic Analysis, the Internal Revenue Service, the Robert Wood Johnson Foundation, the National Association of Public Health Statistics and Information Systems and the Centers for Disease Control, among others. This kind of data collection, organization and analysis, is very recent and, as far as neighborhood level life expectancy is concerned, has been produced only once (USALEEP). We hope that the present analysis along with other recent influential studies of neighborhood effects and improvements in data sciences and technology motivate such integrated data collections on an ongoing basis towards producing a reliable benchmark for promoting human development across the US over time and systematically mitigating inequalities of human capabilities.

As this type of evidence becomes more available, we expect it to drive a number of breakthroughs in policy and in social theory associated with the integrated understanding of complex social processes across scales [76, 77, 78]. First, creating more complex localized indices of sustainable development at the community scale remains the main challenge underlying the United Nations Sustainable Development Goals, adopted by 193 different nations around the world [20]. Demonstrating that such localization is possible and useful at identifying context appropriate and human-centric processes of development is critical for a fast convergence to sustainability worldwide. The localization of sustainable development indices at smaller scales has a number of other critical implications, for example allowing us to gauge inequities linked to gender or race and ethnicity much more directly and minimizing adverse distributional effects in policy, as now mandated by the US federal government.

Second, and most important, the capabilities approach to human development along with recent findings of life course effects from varying exposure to neighborhood environments during childhood and adolescence [64, 54, 79] has the potential to help unify emerging findings in sociology, public health, life history theory, behavioral economics, and social psychology into a common synthetic framework capable of generating practical and systemic solutions to urgent problems of community development in the US and around the world [77, 78]. The scope of this convergence could be profound both for interdisciplinary social theory and for development policy. National evidence on the importance of tracking human development via the HDI and promoting context appropriate public policies - in different political, economic and cultural contexts - shows that this can be done. The urgency of crises of equity and sustainability, together with advances in data sciences make it also easier and more necessary.

3.4 Methods

3.4.1 Data Sources

Tract-level average real incomes were calculated by down-allocating the 2015 US GNI PPP (constant 2011 international dollars) reported by the UN Statistics Division using total tract incomes from the American Community Survey (ACS) for 2010-2015. These were then adjusted to create real incomes based on regional purchasing power parities reported by the US Bureau of Economic Analysis. Non-metro tracts were adjusted for real incomes using state level, non-metro PPP. Population estimates, school enrollment, and educational attainment for the population 25 years and older were taken from the ACS 5-year estimates for 2010-2015. Life expectancy data was taken from the US Small-area Life Expectancy Estimates Project (USALEEP). Tracts where life expectancy data was not reported received county-level life expectancy values from 2014. Tracts that do not have residents in every age cohort also were filled in using county-level values for education. These make up 3.05% of the cases. For the five census tracts located in counties that also had incomplete age cohorts, only Mean Years of Schooling was used to calculate the Education Index. Social disadvantage metrics in Figure 3.3 were obtained from the Opportunity Atlas [63], the American Community Survey (ACS) 5-year estimates for 2015-2019 and the CDC PLACES database, see supplementary text for details. Data and python code is provided online at `https://github.com/mansueto-institute/local-hdi`.

3.4.2 Interactive Map of CHDI

A detailed interactive map of HDI at the census tract level, showing statistics and comparison to international standards, is available online at `https://communityhdi.org`.

3.4.3 Best fits and Beta density estimation

Linear and sigmoid best fits were estimated using the python package *curvefit* from *scypy.optimize*. Beta parameter estimation and density selection were performed using the python package *reliability* . Principal component analysis was performed on the set of 17 social disadvantage metrics plus the HDI at the census tract level using python package *sklearn.decomposition*, see Supplementary Text for details.

3.5 Code Availability

The code used to aggregate and generate the data underlying this work is available at `https://github.com/mansueto-institute/local-hdi`. Data sources can be found in the Materials and Methods section.

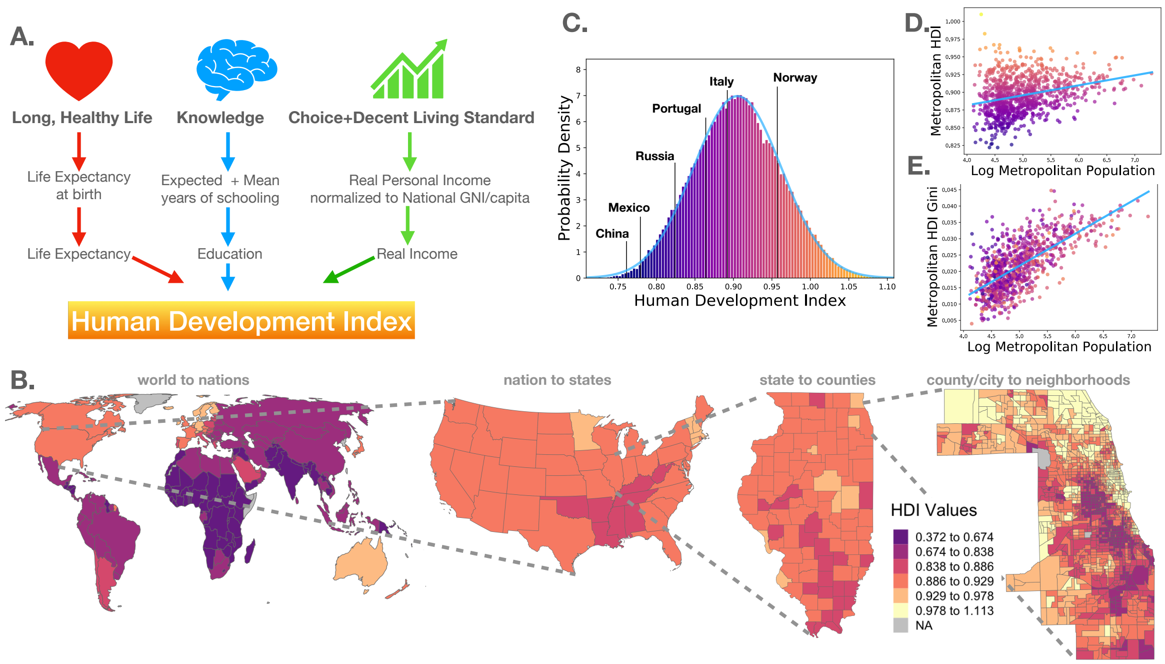
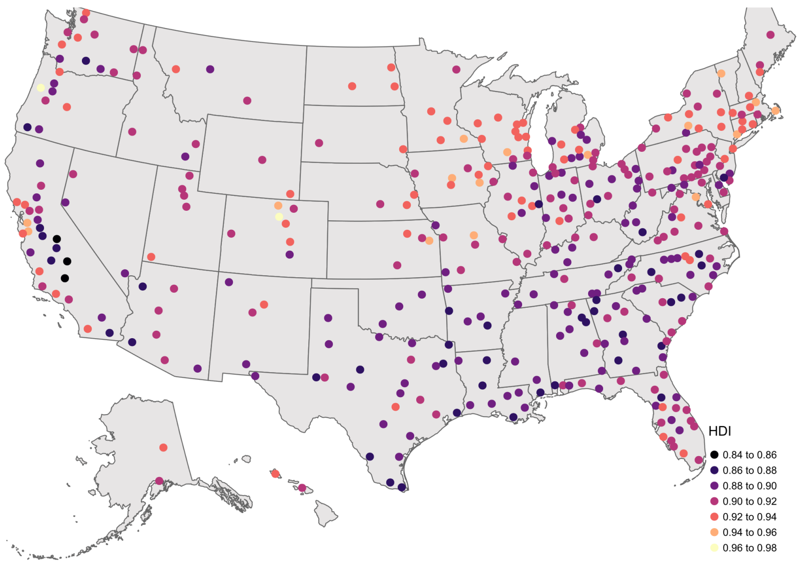


Figure 3.1: Human Development Index (HDI) definition and its expression across scales. A. The HDI is defined as the geometric mean of 3 components, indexing educational attainment, life expectancy and real incomes to given international standards. B. The HDI (dis)aggregation across spatial scales from nations to states, counties and cities to neighborhoods (census tracts). We typically observe the greatest HDI variation at the local level of neighborhoods within larger urban areas such as Chicago. C. The statistical distribution of neighborhood level HDI is very well described by a simple Normal distribution, $N(\mu, \sigma)$ with mean $\mu = 0.905$ and standard deviation $\sigma = 0.057$ (see Table 3.1 for statistical tests). D. Larger metropolitan areas show on average larger HDI but the effect is noisy, characterized by a linear fit on the logarithm of population size (blue line) with slope $0.011(0.005, 0.017)$ and intercept $0.825(0.810, 0.841)$, where brackets indicate 95% confidence intervals in parameter estimates. E. The inequality in human development (Gini coefficient) among local communities within metropolitan areas increases on average linearly with the logarithm of population size (blue line) with slope $0.0098(0.0092, 0.0105)$. The increase in human development with city size and associated rise in local inequality motivates the connection between the HDI as a general indicator of human well being and analyses of neighborhood effects.



State/Federal District	HDI	Metropolitan Area	HDI
Massachusetts	0.9407	Boulder, CO	0.9624
Connecticut	0.9378	Corvallis, OR	0.9622
Vermont	0.9348	Ann Arbor, MI	0.9570
New Hampshire	0.9345	Iowa City, IA	0.9567
Minnesota	0.9330	San Jose-Sunnyvale-Santa Clara, CA	0.9557
New Jersey	0.9281	Bridgeport-Stamford-Norwalk, CT	0.9553
District of Columbia	0.9273	Ames, IA	0.9532
Maryland	0.9271	Lawrence, KS	0.9496
North Dakota	0.9244	Boston-Cambridge-Newton, MA-NH	0.9491
Colorado	0.9240	San Francisco-Oakland-Hayward, CA	0.9486
Texas	0.8964	Gadsden, AL	0.8647
South Carolina	0.8935	McAllen-Edinburg-Mission, TX	0.8641
Nevada	0.8919	Lake Havasu City-Kingman, AZ	0.8637
Oklahoma	0.8879	Laredo, TX	0.8614
Tennessee	0.8874	Dalton, GA	0.8614
Kentucky	0.8843	Brownsville-Harlingen, TX	0.8605
Alabama	0.8839	Yakima, WA	0.8604
Louisiana	0.8838	Pine Bluff, AR	0.8601
West Virginia	0.8832	Bakersfield, CA	0.8595
Arkansas	0.8798	Visalia-Porterville, CA	0.8579
Mississippi	0.8762	Madera, CA	0.8572

Figure 3.2: Human Development in US States and Metropolitan Areas. Mapping the HDI of different metropolitan areas (left), allows us to observe which places lead solutions that promote greater human capabilities, and which lag. The differences can be ranked, with the top and bottom States and Metropolitan Areas shown on the right. States and cities with larger public efforts in education, healthcare and innovation tend to be leaders. This pattern is clearer at the city level where "college towns" and urban areas with higher concentrations of education and research top HDI rankings, regardless of geography even in lower performing states. The top HDI micropolitan area is Los Alamos, NM.

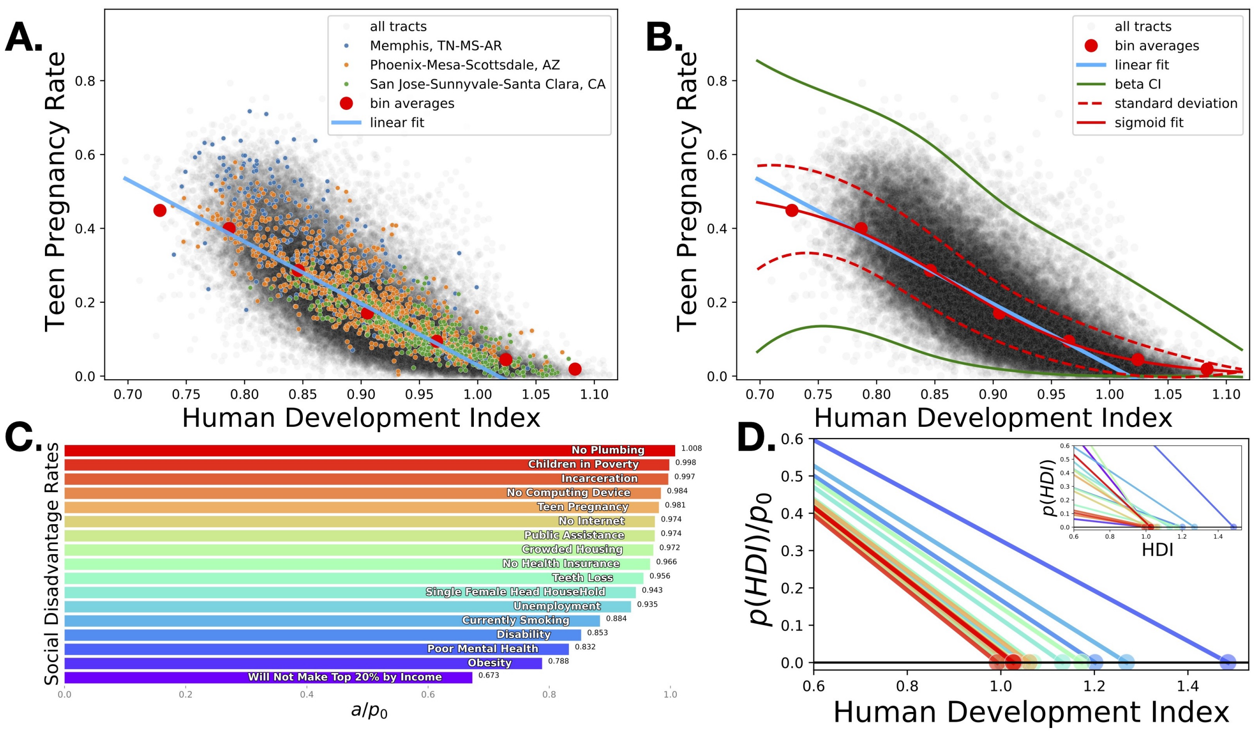


Figure 3.3: Higher human development index values are associated with many different lower social disadvantage rates. A. Highlights for three different metropolitan areas with low, medium and high HDI for teen pregnancy rates. The average rate (red circles) decreases approximately linearly (blue line) with HDI. B. A Beta distribution with parameters varying with HDI gives a better description of the rate statistics, including slowing down at very high HDI and associated variance (risk) reduction. C. The slopes of the negative linear relation between 17 rates of social disadvantage and HDI, shown in Figure 3.3D inset. D. When adjusted for rate initial magnitudes at low HDI, all social disadvantage rates display similar slopes and vanish for HDI$\rightarrow$ 1 or slightly above, see Supplement for additional Figures 3.21 to 3.29 and parameter estimation Tables 3.4 and 3.5.

3.6 Supplementary Materials

3.6.1 Variable Definitions and Meaning

Measures of neighborhood effects were taken from a diverse range of datasets. The sources and definitions of these variables are detailed below.

Ten variables were taken from the 2015-2019 American Community Survey (ACS) 5-year estimate data [80]. *No Plumbing* is the "The percentage of all ACS housing units that do not have complete plumbing facilities." *Children in Poverty* is the "The percentage of children under age 18 living in households that are below the poverty line in the ACS." *No Computing Device* is the "Percentage of people that live in households that do not have computing device of any kind in the ACS." *No Internet* is the "Percentage of ACS households that have no Internet access."*Public Assistance* is the "The percentage of all ACS occupied housing units that receive public assistance income (general assistance and Temporary Assistance to Needy Families)." *Crowded Housing* is the "The percentage of ACS occupied housing units that have more than 1.01 persons per room." *No Health Insurance* refers to "The percentage of the ACS population that have no health insurance, public or private." *Single Female Head of Household* is the "The percentage of all ACS occupied housing units with a female householder and no spouse of householder present." *Unemployment* refers to "The percentage of ACS civilians ages 16 years and over in the labor force that are unemployed." *Disability* is the "The percentage of the ACS population who have one or more disabilities."

A number of variables were also taken from the Opportunity Atlas [63]. *Teenage Birth Rate* represents "Fraction of women who grew up in this area who claimed ever a child who was born when the women were between the ages of 13 and 19 as a dependent when filing taxes." *Incarceration Rate* represents "Fraction of children who grew up in this area who were in prison or jail on April 1, 2010." *Will Not make top 20 Percent of Earners* is 1 minus "Fraction of children who grew up in this area who have average individual income in 2014-15 (in their mid-30s) in the top 20 percent of the national income distribution for children born in the same year. "

Health outcomes data was taken from the US Centers for Disease Control PLACES Project [81], which provides small area estimates for 27 chronic disease rates, health behaviors, and chronic conditions. The estimates are calculated using a multilevel regression and poststratification framework. These estimates have been independently validated [82].

Tooth Loss represents biennial prevalence of "Respondents aged $\geq$ 65 years who report having lost all of their natural teeth because of tooth decay or gum disease." divided by total "Respondents aged $\geq$ 65 years." *Currently Smoking* represents annual prevalence of "Respondents aged $\geq$ 18 years who report having smoked $\geq$ 100 cigarettes in their lifetime and currently smoke every day or some days" divided by "Respondents aged $\geq$ 18 years who reported information about cigarette smoking." *Poor Mental Health* represents annual prevalence of "Respondents aged $\geq$ 18 years who report 14 or more days during the past 30 days during which their mental health was not good" divided by "Respondents aged $\geq$ 18 years who report or do not report the number of days during the past 30 days during which their mental health was not good." *Obesity* represents the annual prevalence of "Respondents aged $\geq$ 18 years who have a body mass index (BMI) $\geq$ 30.0 kg/m^2 calculated from self-reported weight and height." divided by "Respondents aged $\geq$ 18 years for whom BMI can be calculated from their self-reported weight and height." This excluded the following: "data from respondents with heights less than 3 ft or $\geq$ 8 ft, data from respondents weighing less than 50 lbs or $\geq$ 650 lbs, data from respondents with BMI less than 12 kg/m2 $\geq$ 100 kg/m2, and pregnant women."

3.6.2 Linear Fits, Urban Definitions and Visualization

Figures 3.4 to 3.20 show linear fits characterizing the variation of (components) of HDI with population and various population rates commonly used to express social disadvantage.

Linear fits were performed in python using package *curvefit* from *scypy.optimize*.

Metropolitan and micropolitan population is the sum total of the population reported by the American Community Survey for each tract within a US Census defined Metropolitan or Micropolitan Statistical Area. These are functional definitions that define cities in terms of unified labor markets, meaning that they include in the same spatial unit both places of work and places of residence, determined based on measured commuting flows. In practice, this includes in the same urban definition central urban cores and suburbs and typically several political units such as "cities," townships and counties.

Though the term "metropolitan" is often used to refer to all these urban areas, in Figures 3.6 to 3.10 we use it in its more specific meaning, to refer to urban areas so defined with populations above 50,000. Micropolitan Statistical Areas, or simply Micropolitan Areas refer to smaller functional cities with populations between about 10,000 and 50,000 people. The largest metropolitan area in the US is New York City with just over 20 million people. There are 381 metropolitan areas in our data set and 536 micropolitan areas, making up a total of 917 in total in the continental US and Alaska and Hawaii.

In statistical studies at the tract level, non-urban (micro or metropolitan) census tracts are also included, as shown in the visualization map `https://communityhdi.org`. Linear fits are done in a similar way using the value of social indicators as the dependent variable and of the HDI as the independent variable, with data as census tracts. This procedure can be repeated at any scale, such as individual metropolitan areas, states, or political cities. Tables 3.2 and 3.3 summarize the linear best fits for the dependence of social disadvantage rates, computed as the population weighted community HDI in each tract in the urban area, versus total population.

3.6.3 Beta Distribution Estimation

To characterize the statistical non-linear dependence of various social disadvantage rate on the community HDI, we modeled these rates as stochastic variables following a Beta distribution. The Beta distribution is the most standard statistical model for variables that describe a continuous probability for an event, such as the probability of teenage pregnancy or incarceration for someone in a given community. In this sense, the Beta distribution for social disadvantage rates conditional of HDI also provides the Bayesian conjugate to many generative processes that can be used to generate event data, including binomial, negative binomial and geometric distributions.

To model the dependence of the two parameters of the Beta distribution on HDI, we estimated the parameters of the distribution for census tracts divided into a number of bins, shown in Figure 3.3A-B and Figures 3.21 to 3.29 and 3.32 by their centers as red dots. For each of these intervals (bins) of HDI, we then collect all corresponding tracts and estimate parameters of the Beta distribution, see Figure 3.30. We verified that the Beta distribution is a general good description of the frequency distribution of tracts for many social disadvantage rates and other variables using (Fit_Beta_2P, Fit_Everything) from the python library *reliability* and a Bayesian Information Criterion test, see Figure 3.30 for an illustration. Tables 3.4 and 3.5 give the values for Beta Average and Beta standard deviation, which are equivalent to the two parameters of the distribution but easier to interpret. We then fit both average and distribution by linear and sigmoid functions to obtain an estimate of the continuous variation of the distribution parameters on HDI. These estimates are shown as both the blue solid lines (linear fit) and red and green sigmoid curves in Figure 3.3A-B and Figures 3.21 to 3.29, as well as Figure 3.32. We see in all cases that the sigmoid curves (which have an extra parameter) give a better fit than the linear model, see Tables 3.4 and 3.5.

3.6.4 Principal Component Analysis

In addition to giving a statistical characterization of pairwise relationships between HDI and social disadvantage rates, we also performed a principal component analysis (PCA) of these associations taking the set of variables in Figure 3.3 together with HDI at the tract level as separate dimensions. This makes for 18 separate dimensions, including HDI. A total of 71,436 census tracts with all variables given across the nation was then used as points, with variance to be accounted for via PCA analysis.

Figure 3.31 shows the relative variance explained by the first principal components, as is usually shown in PCA analysis. We see that the first and second components are by far the most influential, accounting for 54.5% and 8.7% of the total variance at the census tract level. Other dimensions explore various covariations of social disadvantage rates in the 16-dimensional orthogonal directions to the first component.

The first PCA dimension accounts for the linear inverse correlation between the HDI (first element of the vector) and all other components, see Table 3.6, showing the near proportionality of decreases in HDI and of *all* other rates of social disadvantage, supporting the pairwise results shown in Figure 3.3. The second component is “orthogonal” to the first in a number of dimensions, with coefficients of both signs, see Table 3.6.

This analysis therefore supports, in our view, the arguments of the main paper that communities with HDI approaching unity should see predictably small rates of social disadvantage along many distinct dimensions.

3.6.5 Effects of Race and Ethnicity

Many studies of neighborhood effects have provided evidence of the association of social disadvantage rates and racial and ethnic segregation at the community level, especially of Black populations in specific cities [83, 84]. We explored the general influence of race and ethnicity rates (population fraction) with values of the HDI, which we show in the paper is systematically negatively correlated with many dimensions of social disadvantage, Figure 3.3.

Specifically, Figure 3.20 shows the correlation of the fraction of Black, and Hispanic population versus the value of the HDI at the metropolitan level. The correlation is very small (very low R^2), in effect because at low human development there are cities with many different racial and ethnic makeups, some with very small Black or Hispanic populations and others with strong majorities, for example in the South and the border with Mexico, respectively.

The estimation of the correlation between the HDI and fraction Black or Hispanic at the community level (census tracts) reveals a similar pattern, with weak correlations ($R^2 = 0.143$ and 0.083 for Black and Hispanic respectively), see Figure 3.32. Though there is some negative association at this level of minority concentration to HDI, especially for fraction of Black residents, the features of the statistics are as for Metropolitan areas, where low HDI tracts come in many different rates of race and ethnicity.

Within specific metropolitan areas (not shown) the correlation between these race and ethnicity variables at the tract level and the HDI is very context dependent. It is weak or insignificant in many urban areas such as New York City, but it is stronger in cities where racial segregation and associated poverty has been known to be stronger. Examples are Baltimore, Chicago or Memphis for Black rates at the neighborhood level, and Los Angeles for Hispanic. Other variables accounting for the rates of foreign residents or English not spoken at home are even less correlated with HDI, and in some cases show in fact weak positive correlations.

We conclude from this analysis that race and ethnicity variables are not significantly associated with changes in human development at the national level because low human development situations are not *specific* to high rates of Black or Hispanic population in particular. Many places in the nation with very low human development (for the US) have small or vanishing minority populations, such as Appalachia. At the same time, race and ethnicity can show strong associations with HDI in specific cities, many of which have been the focus of past studies and where race and poverty may be strongly segregated in small areas, the original motivation for the modern literature on neighborhood effects. A detailed analysis of these effects in the context of specific cities is therefore necessary and will be developed elsewhere.

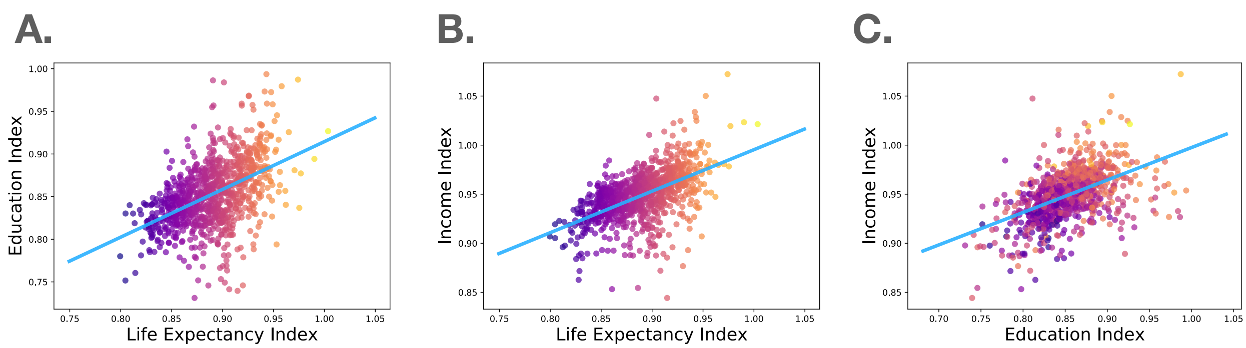


Figure 3.4: Correlation of components of HDI for Metropolitan and Micropolitan Areas. Though the components of the Human Development Index (Education, Income, and Life Expectancy) are positively correlated with one another, there is a large degree of variance observed. A. Dependence of the Education Index (EI) on the Life Expectancy Index (LEI). The best fit line (blue) has intercept $0.354[0.293, 0.414]$ and gradient $0.560[0.492, 0.628]$ B. Dependence of the Income Index (II) on the Life Expectancy Index (LEI). The best fit line (blue) has intercept $0.572[0.533, 0.611]$ and gradient $0.422[0.379, 0.466]$ C. Dependence of the Income Index (II) on the Education Index (EI). The best fit line (blue) has intercept $0.666[0.633, 0.699]$ and gradient $0.330[0.292, 0.369]$. Numbers in parentheses indicate the 95% confidence internal in the parameter estimates.

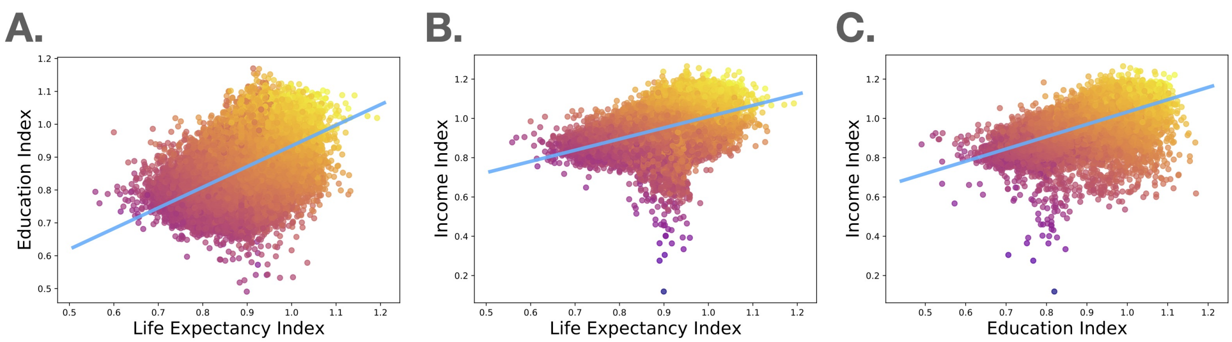


Figure 3.5: Correlation of components of HDI at the census tract level. A higher degree of variance is observed between the quantities at the tract level than at larger geographies, such as metropolitan and micropolitan areas. This demonstrates the importance of working at the tract level, where measures are less correlated. A. Dependence of the Education Index (EI) on the Life Expectancy Index (LEI) for tracts. The best fit line (blue) has intercept $0.305[0.298, 0.312]$ and gradient 0.628 $[0.620, 0.636]$ B. Dependence of the Income Index (II) on the Life Expectancy Index (LEI) for tracts. The best fit line (blue) has intercept 0.438 $[0.431, 0.445]$ and gradient 0.570 $[0.562, 0.577]$ C. Dependence of the Income Index (II) on the Education Index (EI) for tracts. The best fit line (blue) has intercept 0.404 $[0.400, 0.409]$ and gradient 0.627 $[0.622, 0.632]$. Numbers in parentheses indicate the 95% confidence internal in parameter estimates.

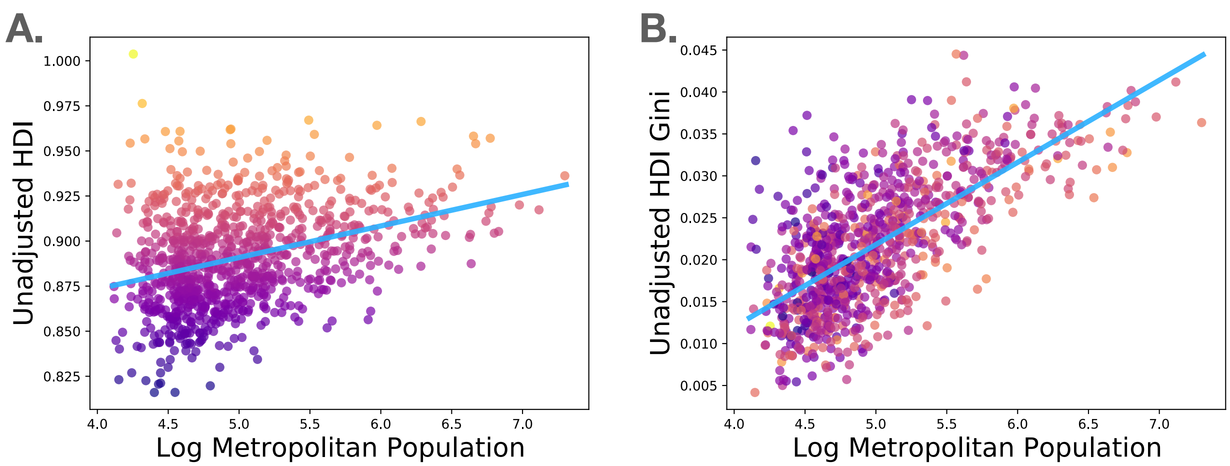


Figure 3.6: The unadjusted HDI -computed using nominal income rather than with local purchasing power parity correction- for all metropolitan areas in the US. This includes so-called micropolitan areas and metropolitan areas. A. Average unadjusted HDI on city population size. The best fit line (blue) has intercept $0.803[0.787, 0.819]$ and gradient 0.017, $[0.014, 0.021]$, $R^2 = 0.12$. B. The dependence of the unadjusted HDI Gini coefficient (inequality) on population size. The best fit line (blue) has intercept -0.027, $[-0.030, -0.024]$ and gradient 0.0097, $[0.0091, 0.0105]$, $R^2 = 0.48$. Parenthesis indicate the 95% confidence internal in the estimates of both parameters.

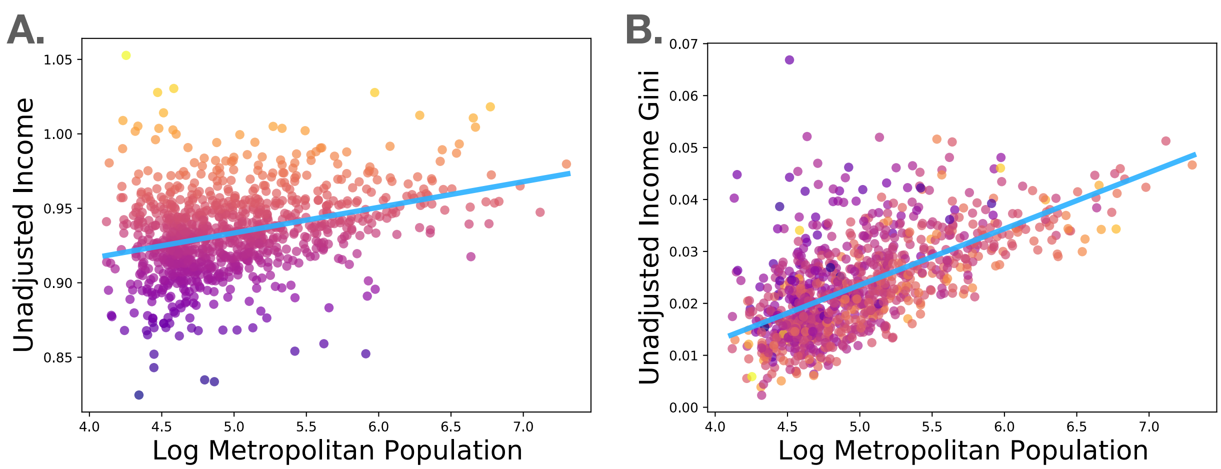


Figure 3.7: The unadjusted income index -using nominal income rather than with local purchasing power parity correction- versus population for all metropolitan and micropolitan areas in the US. A. The average unadjusted income index on city population size. The best fit line (blue) has intercept 0.847, $[0.831, 0.864]$ and gradient 0.017, $[0.014, 0.020]$. B. Scaling of unadjusted income index Gini coefficient with population size. The best fit line (blue) has intercept -0.031, $[-0.035, -0.026]$ and gradient 0.0108, $[0.0099, 0.0117]$. Numbers in parentheses indicate the 95% confidence internal in the estimates of both parameters.

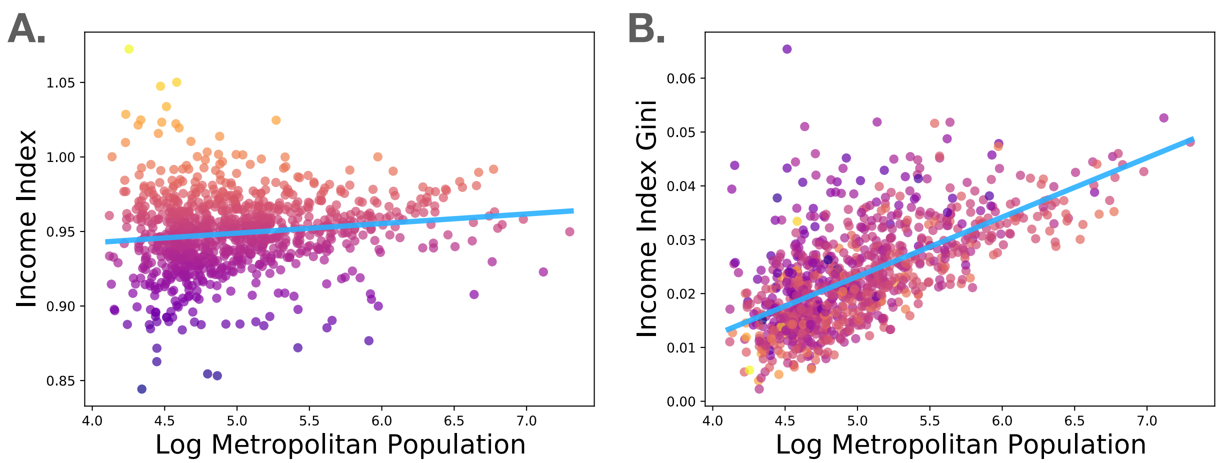


Figure 3.8: The adjusted income index and its Gini coefficient versus population for metropolitan and micropolitan areas. A. The best fit line for the income index versus logarithmic of population has gradient 0.0064 $[0.0033, 0.0095]$ and intercept 0.9168 $[0.9012, 0.9322]$. B. The dependence of the Gini coefficient of the income index on logarithm of populations has gradient 0.0110 $[0.010, 0.01199173]$ and intercept -0.0319 $[-0.0363, -0.0275]$.

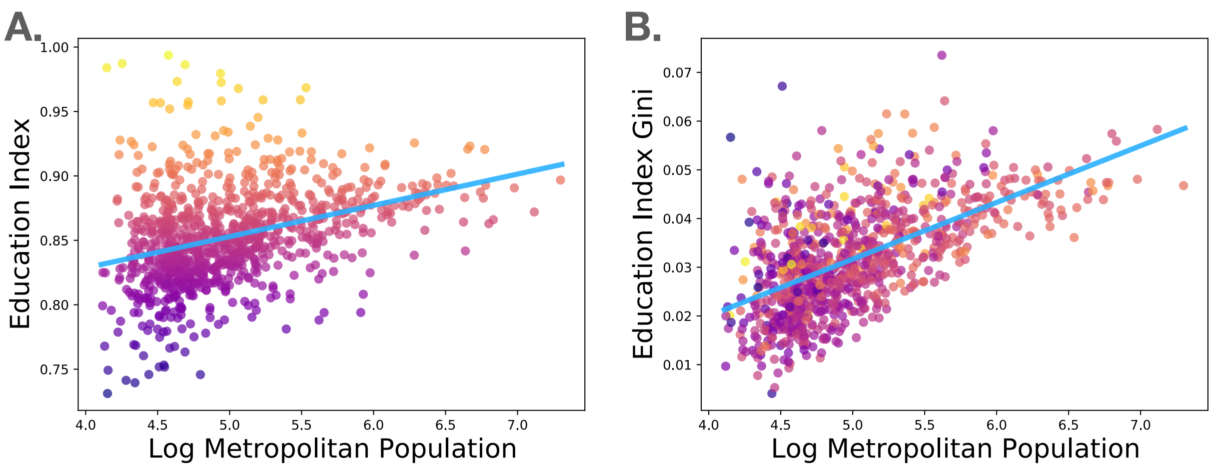


Figure 3.9: The dependence of the education index and its Gini coefficient on population size for both metropolitan and micropolitan areas in the US. A. The dependence of the education index with logarithm of the population is characterized by a linear fit (blue line) with gradient 0.0243 $[0.020, 0.0286]$ and intercept 0.7316 $[0.7097, 0.7535]$. B. The dependence of the Gini coefficient of the income index on the logarithm of population is characterized by a linear best fit (blue line) with gradient 0.0116 $[0.0106, 0.0127]$ and intercept -0.0265 $[-0.0318, -0.0213]$

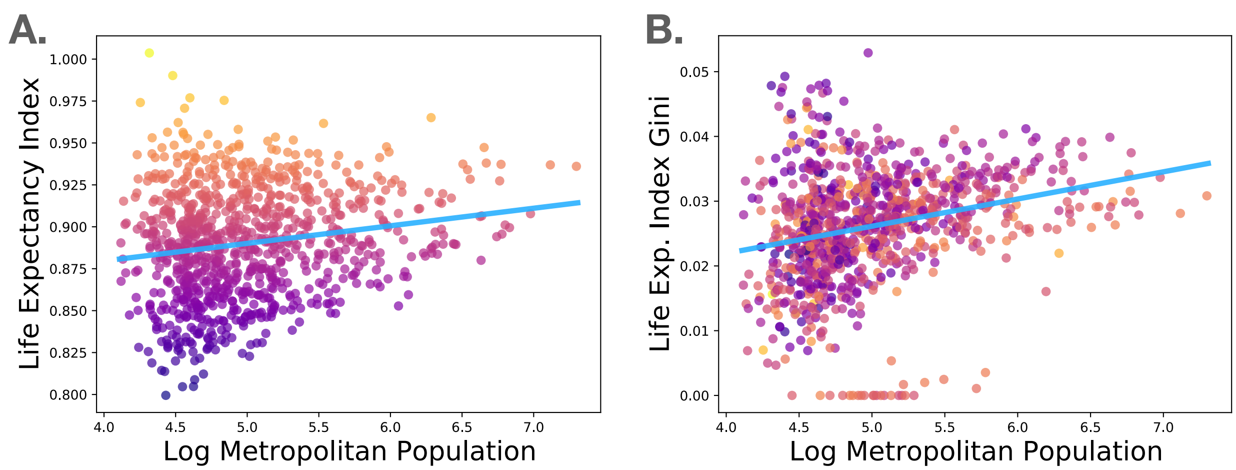


Figure 3.10: The dependence of the life expectancy index and its Gini coefficient on the population size of metropolitan and micropolitan areas in the US. A. The best fit line (blue) for the life expectancy index on the logarithm of population has gradient 0.0105 $[0.0066, 0.0144]$ and intercept 0.8377 $[0.8181, 0.8574]$. B. the Gini coefficient dependence on the logarithm of population (blue line) has gradient 0.0042 $[0.0032, 0.0052]$ and intercept 0.0051 $[0.0005, 0.0103]$

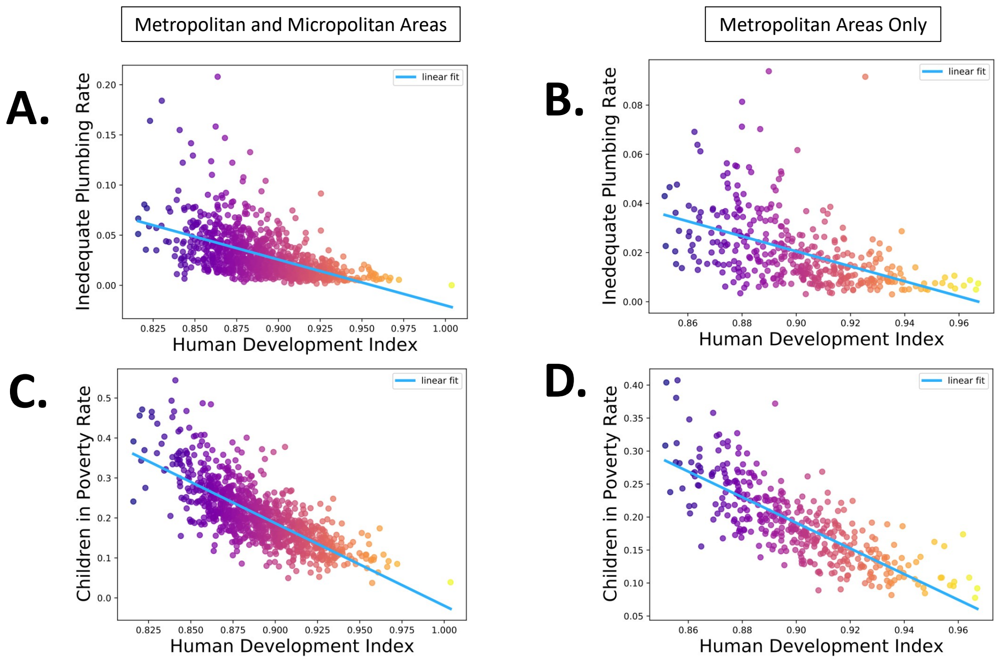


Figure 3.11: Inadequate plumbing and children in poverty by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Table 3.3.

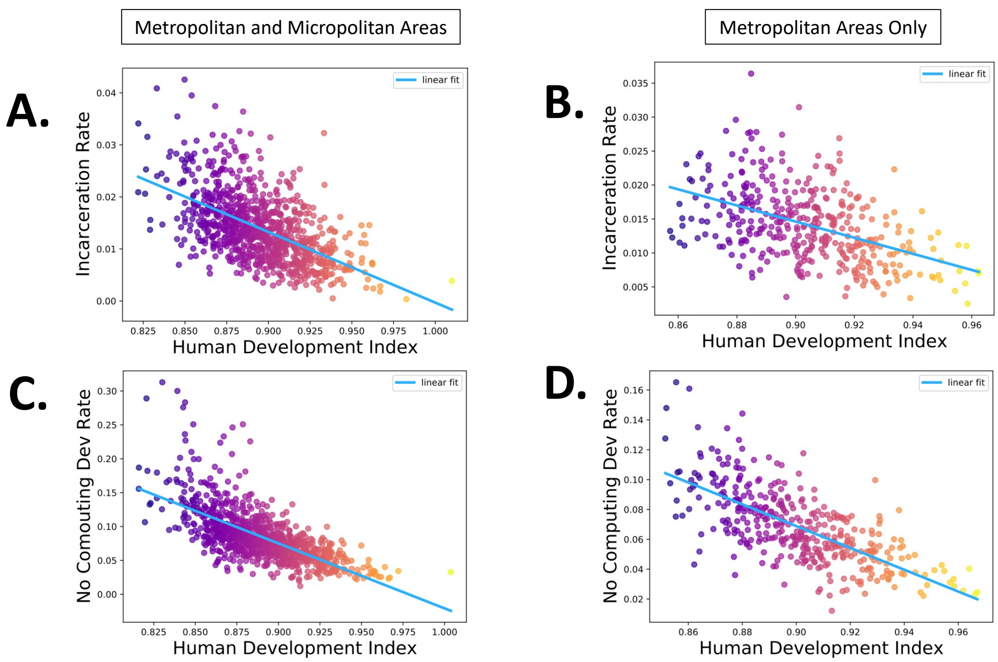


Figure 3.12: Incarceration rate and fraction of individuals living in households with no computing device by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Tables 3.2 and 3.3.

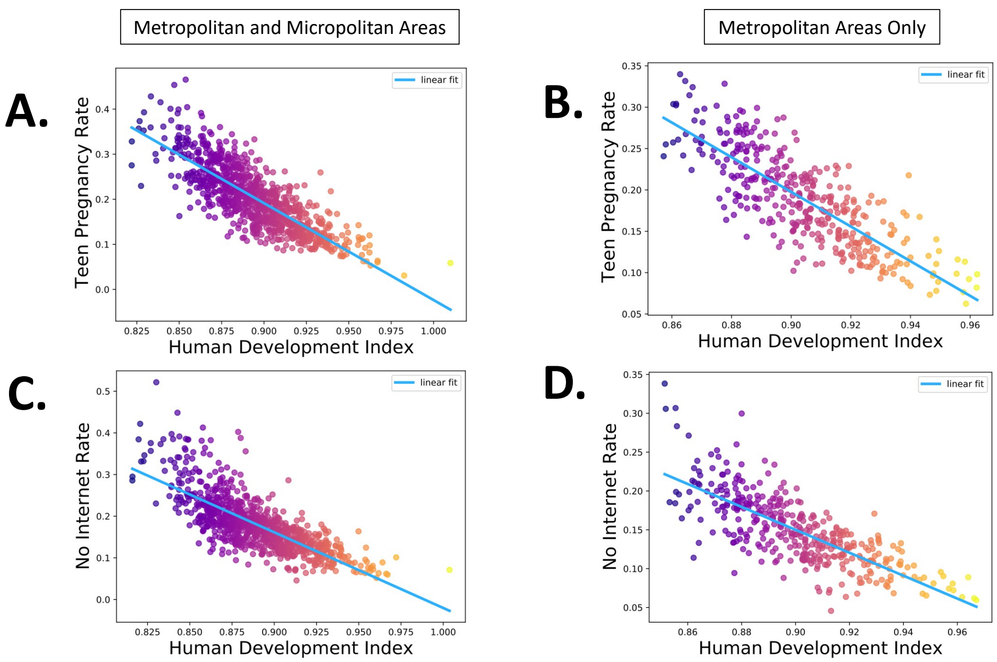


Figure 3.13: Teenage pregnancy rate and fraction of households that have no internet access by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Tables 3.2 and 3.3.

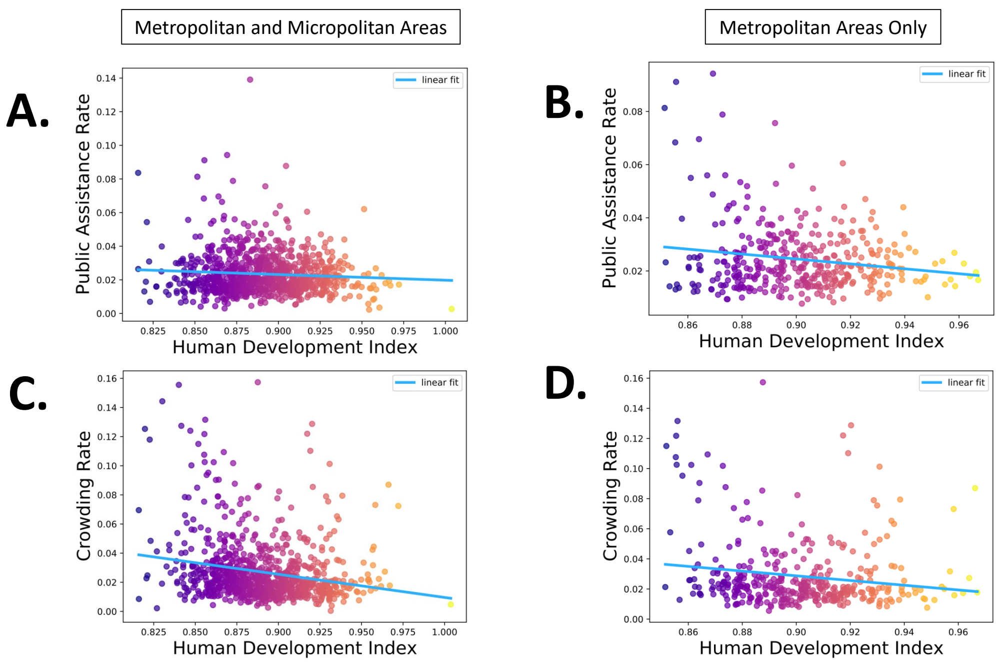


Figure 3.14: Fraction of housing units that receive public assistance and fraction of households with crowding (> 1 person/room) by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Table 3.3.

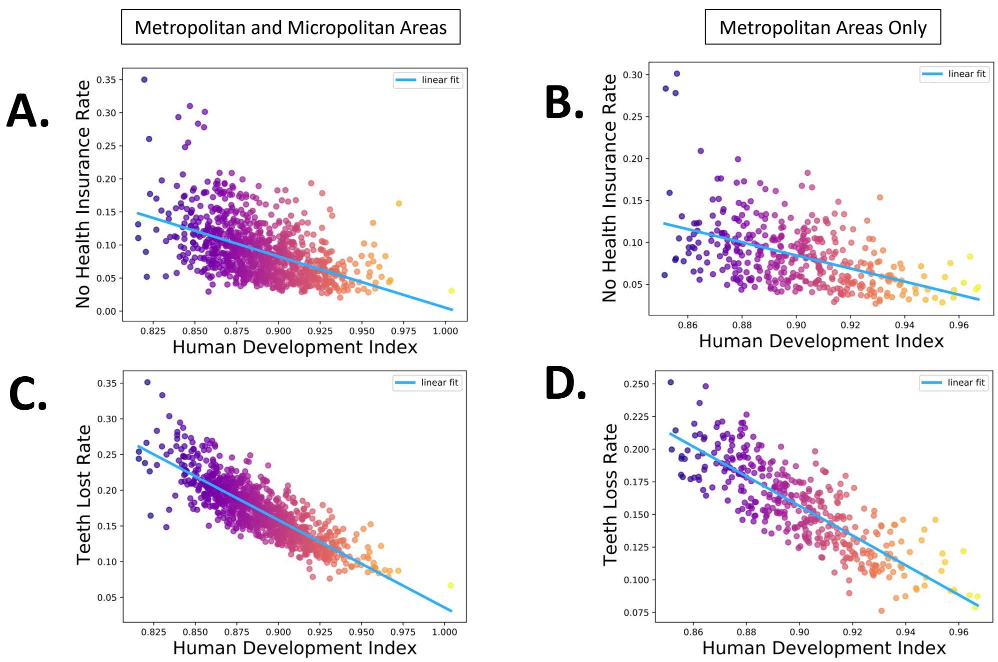


Figure 3.15: Fraction of people with no health insurance coverage and teeth loss (fraction aged ≥ 65 years who report having lost all of their teeth) by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Tables 3.2 and 3.3.

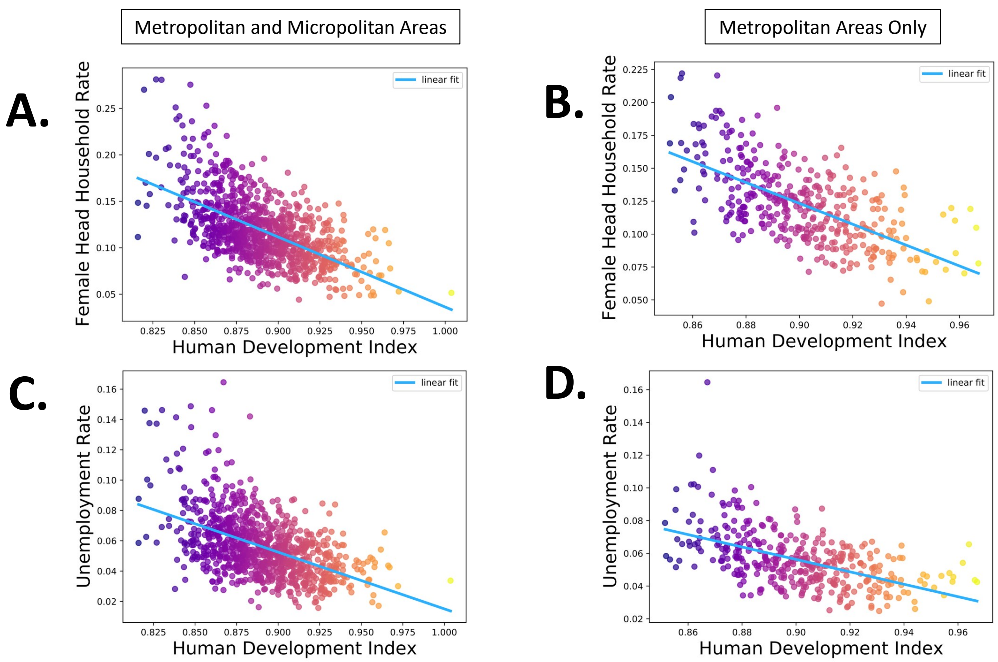


Figure 3.16: Fraction of households with a single female householder and population aged ≥ 16 years in the labor force that are unemployed by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Tables 3.2 and 3.3.

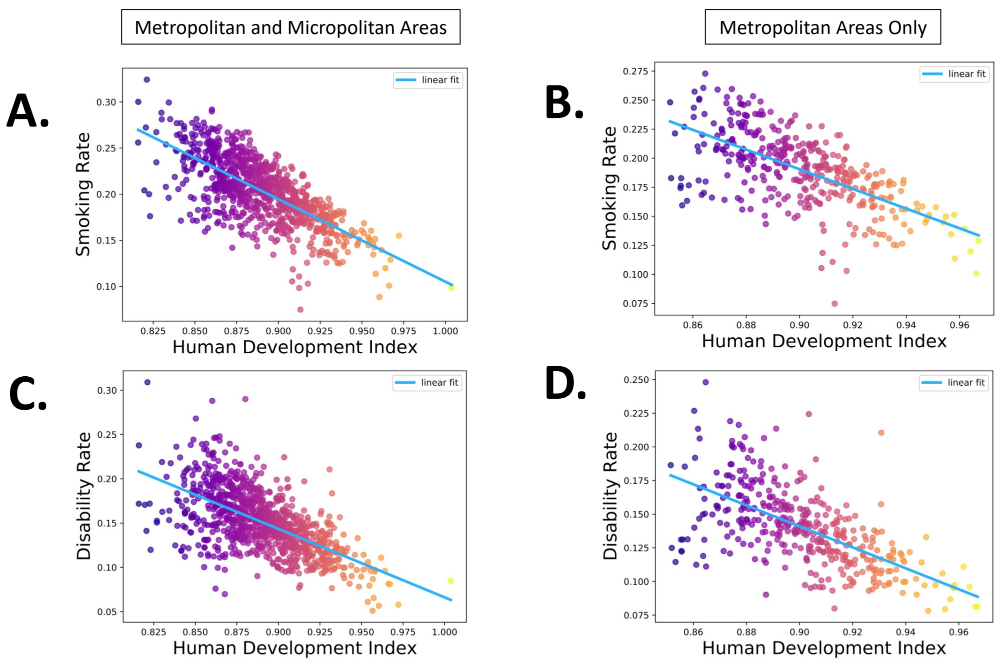


Figure 3.17: Fraction of individuals who currently smoke and fraction of individuals who report one or more disabilities by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Table 3.2.

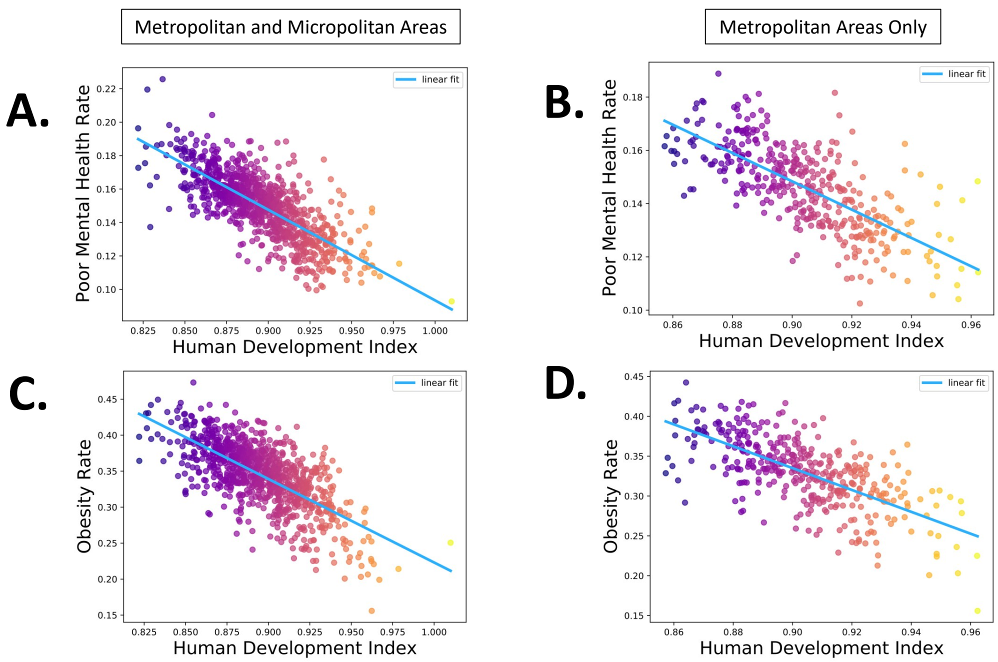


Figure 3.18: Fraction of individuals with poor mental health (fraction who report 14 or more days during the past 30 days during which their mental health was not good) and fraction of individuals who are obese by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Table 3.2.

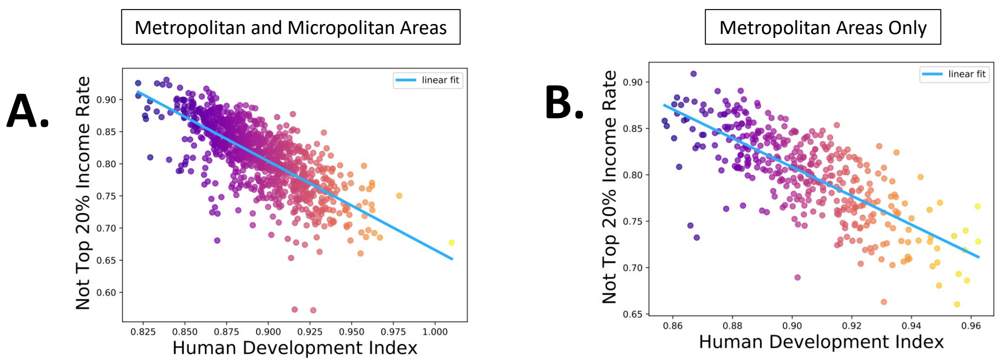


Figure 3.19: Fraction of children who grew up in this area who do not reach individual income in 2014-15 in the top 20 percent of the national income distribution by Metropolitan and Micropolitan Area vs HDI. Best line fits (in blue) are given in Table 3.2.

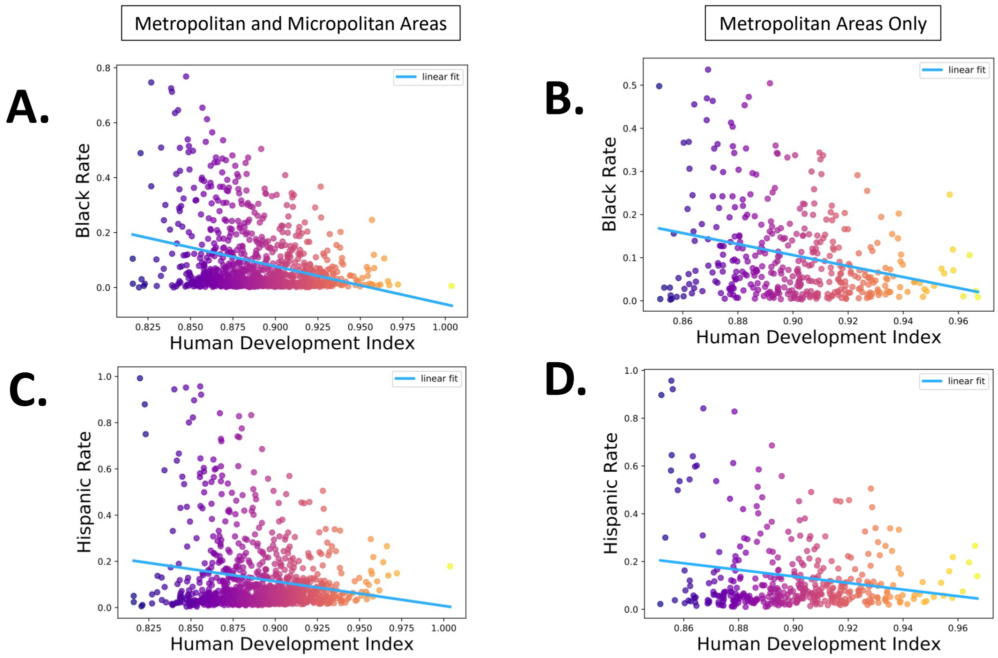


Figure 3.20: The relationship between the Community Human Development Index (CHDI) and the fraction of Black (top) or Hispanic (bottom) population in 917 Metropolitan and Micropolitan areas (left) or limited to 381 Metropolitan areas (right). A. The linear relation (blue line) has gradient $-1.381(0.288)$ and intercept $1.319(0.257)$, with $R^2 = 0.091$; B. The linear relation (blue line) has gradient $-1.274(0.455)$ and intercept $1.253(0.410)$, with $R^2 = 0.076$; C. The linear relation (blue line) has gradient $-1.062(0.387)$ and intercept $1.069(0.345)$, with $R^2 = 0.032$; D. the linear relation (blue line) has gradient $-1.379(0.686)$ and intercept $1.379(0.618)$, with $R^2 = 0.0409$. In all cases the correlation is barely significant and largely the result of greater variation of population composition at lower HDI values, meaning that low human development cities have many different racial and ethnic makeups, including some with relatively large Black or Hispanic populations. Note however that low HDI is not associated at this scale to mostly Black or mostly Hispanic urban areas.

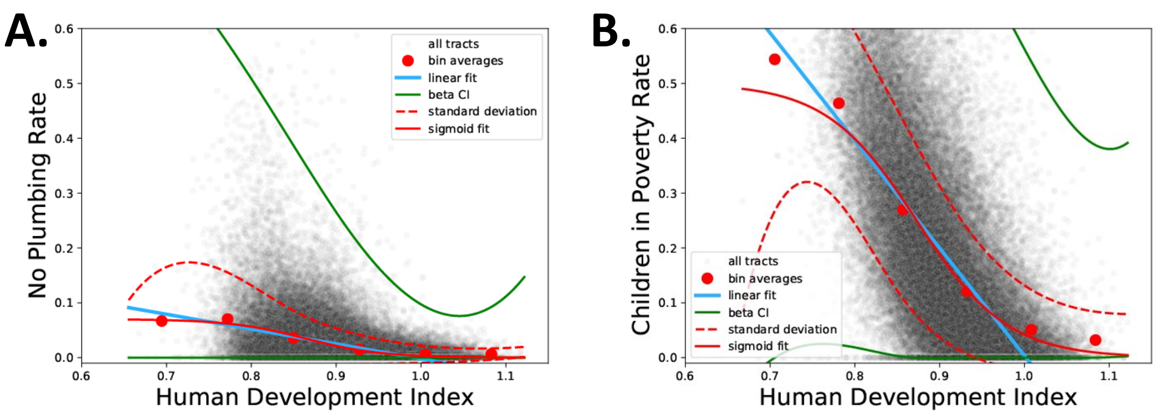


Figure 3.21: Rates of A. households without plumbing and B. children in poverty versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Table 3.5.

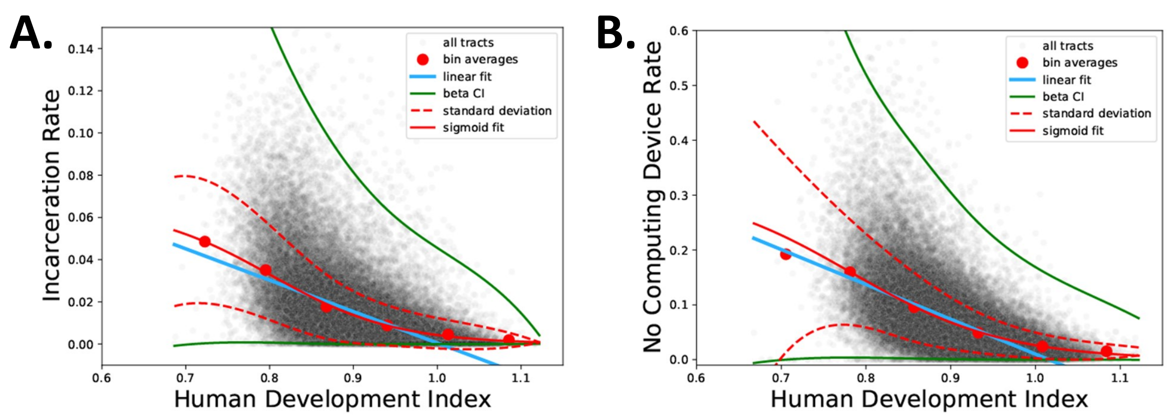


Figure 3.22: Rates of A. incarceration per person and B. households without a computing device for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Tables 3.4 and 3.5.

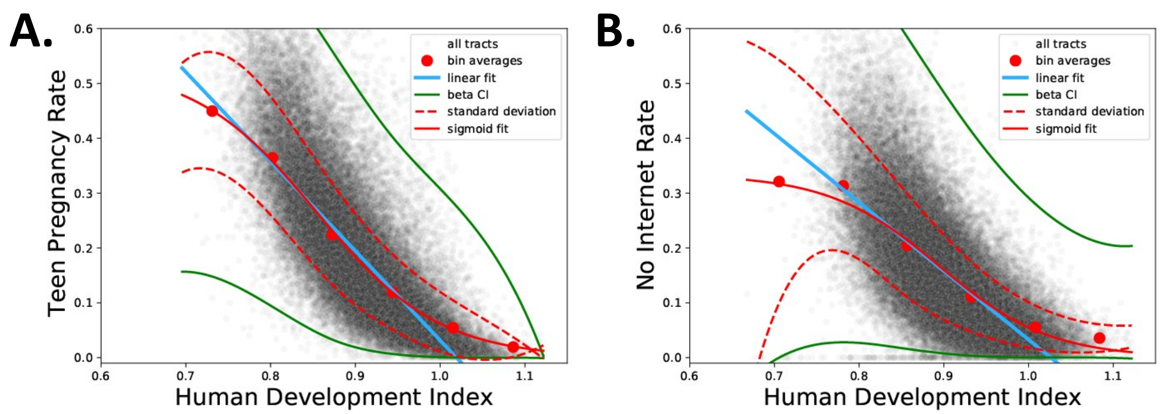


Figure 3.23: Rates of A. teen pregnancy and B. households without internet access versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Tables 3.4 and 3.5.

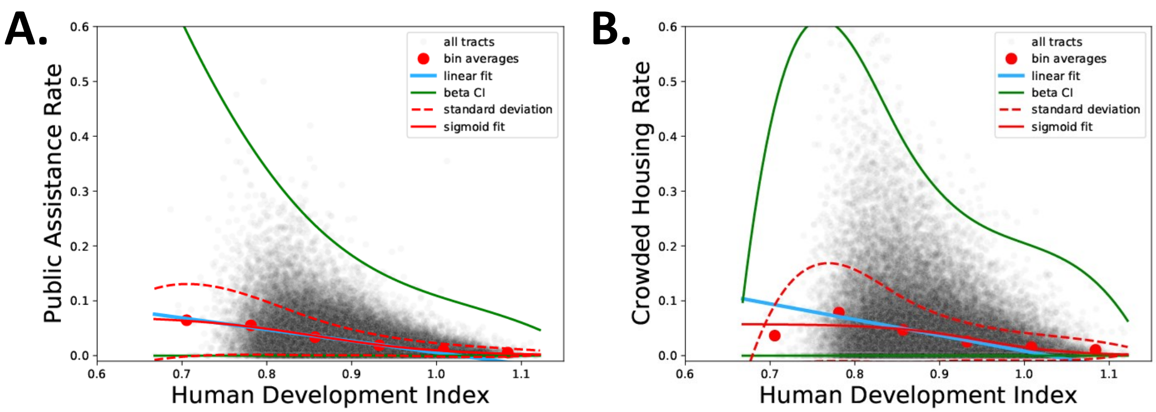


Figure 3.24: Rates of A. people on public assistance and B. crowded households versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Table 3.5.

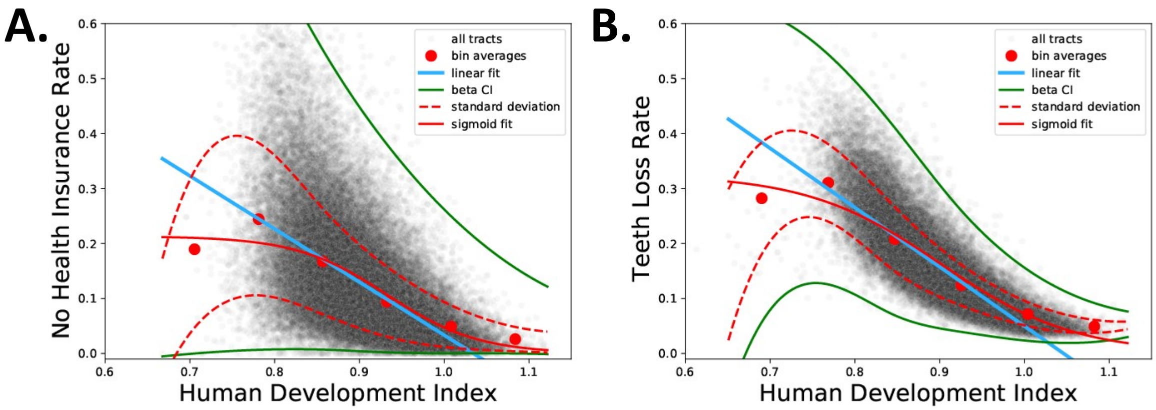


Figure 3.25: Rates of A. people without heath insurance and B. people with teeth loss versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Tables 3.4 and 3.5.

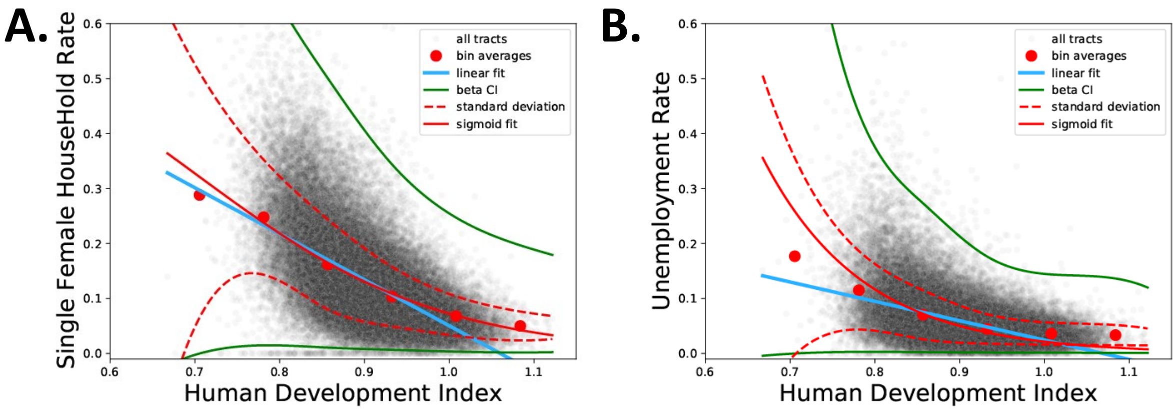


Figure 3.26: Rates of A. single female headed households and B. unemployment versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Tables 3.4 and 3.5.

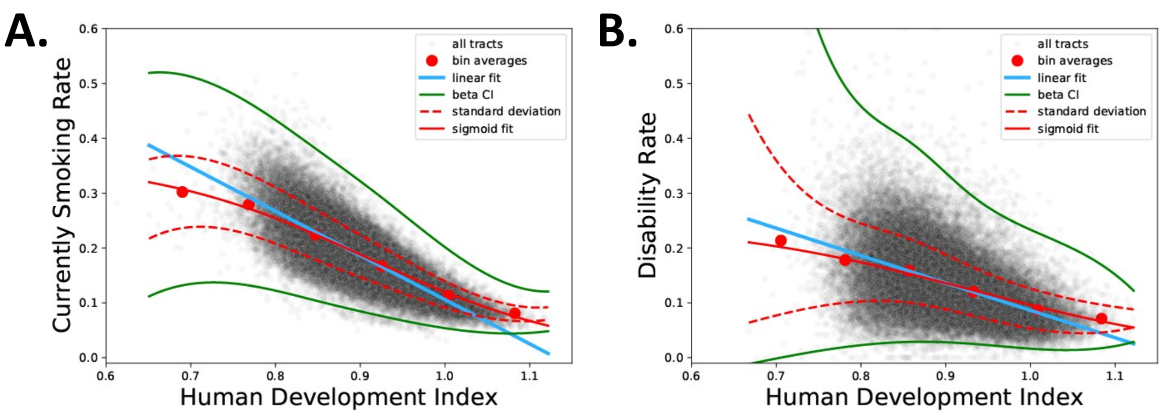


Figure 3.27: Rates of A. people currently smoking and B. people with disability versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Table 3.4.

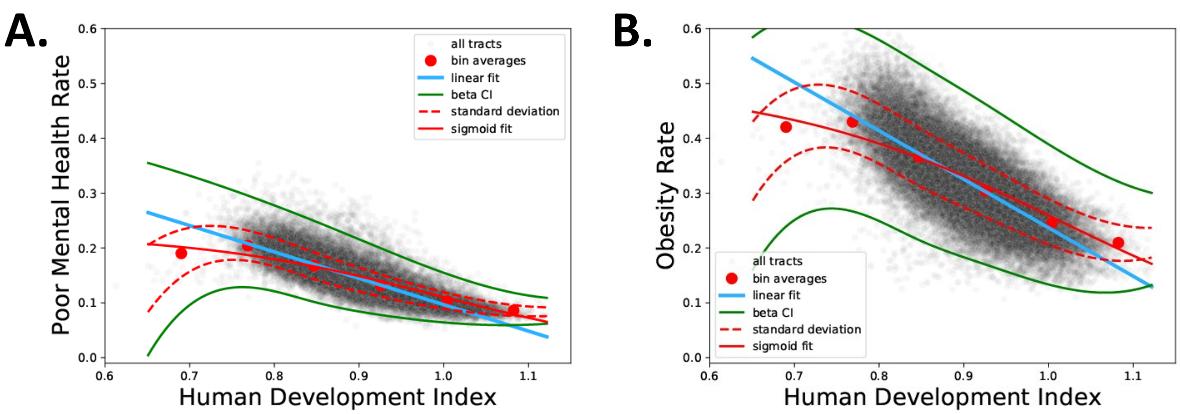


Figure 3.28: Rates of A. poor mental health and B. obesity versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Table 3.4.

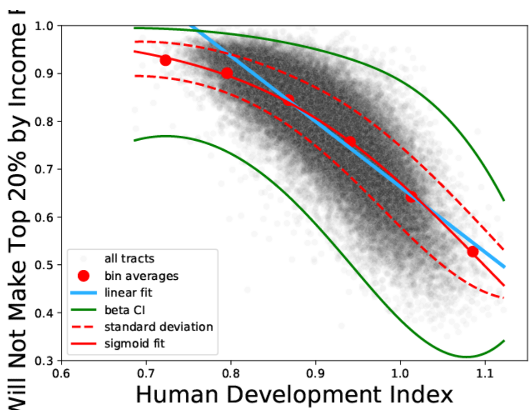


Figure 3.29: Rates of people not making it to the top 20% of income versus HDI for census tracts nationwide. The blue line shows a linear fit, while the red and green lines show the estimation for Beta distribution mean (solid red), standard deviation (dashed red) and 95% confidence interval (green). The red dots show the bin averages used for the sigmoidal fits. Parameter estimations and goodness of fit are given in Table 3.4.

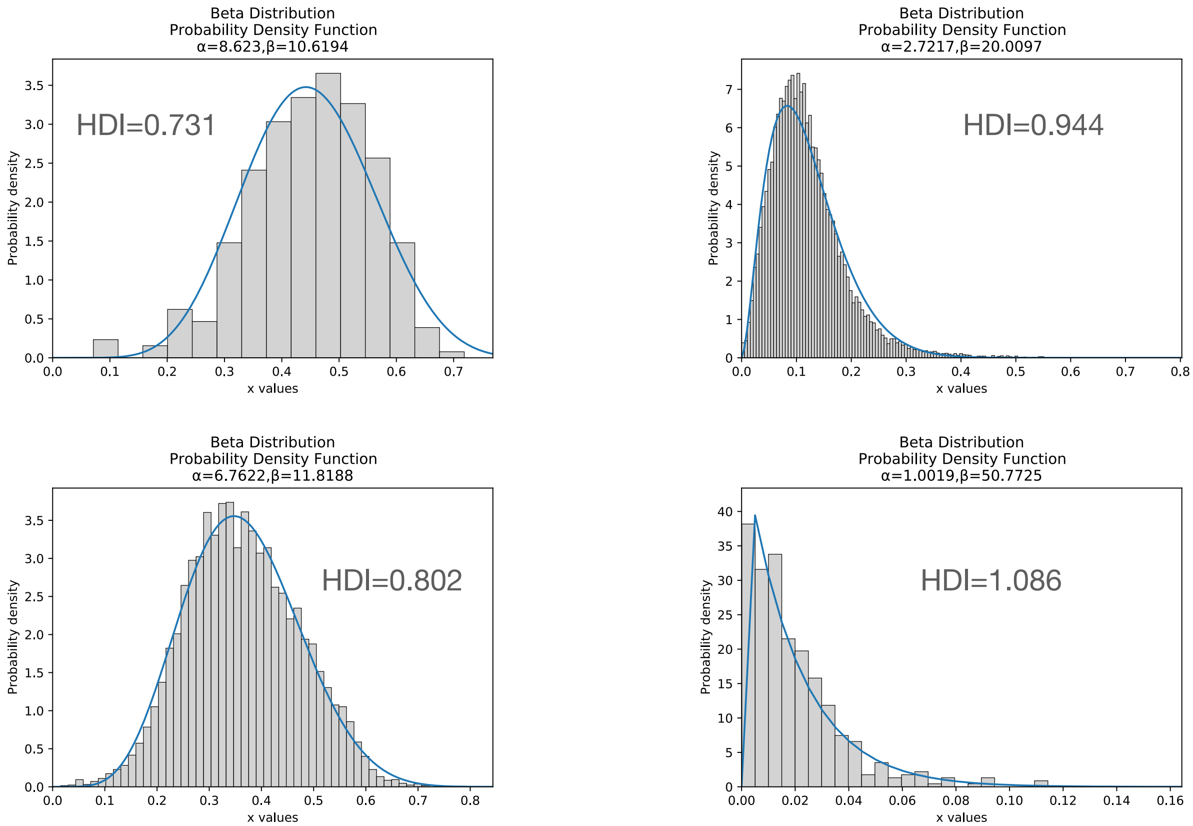


Figure 3.30: Examples of Beta distribution fits for the variation in Teen Pregnancy rates for sets of tracts with different HDI. Note how the histograms changes shape and how the Beta distribution is able to estimate all cases reasonably well. See Table 3.4 for parameter estimates.

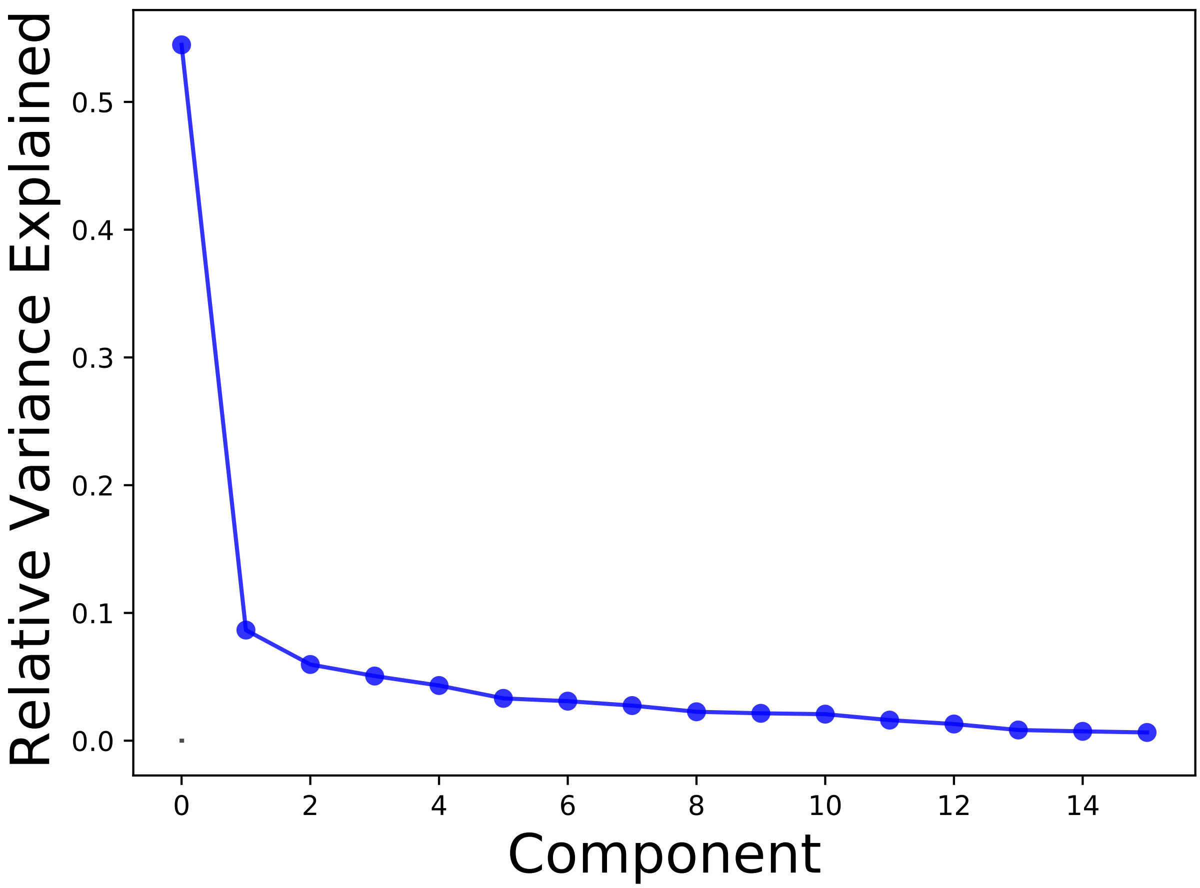


Figure 3.31: The relative variance explained by rank ordered PCA components for the variables of Figure 3.3. The first component corresponds to the decrease in all social disadvantage rates with increases in HDI, see also Table 3.6.

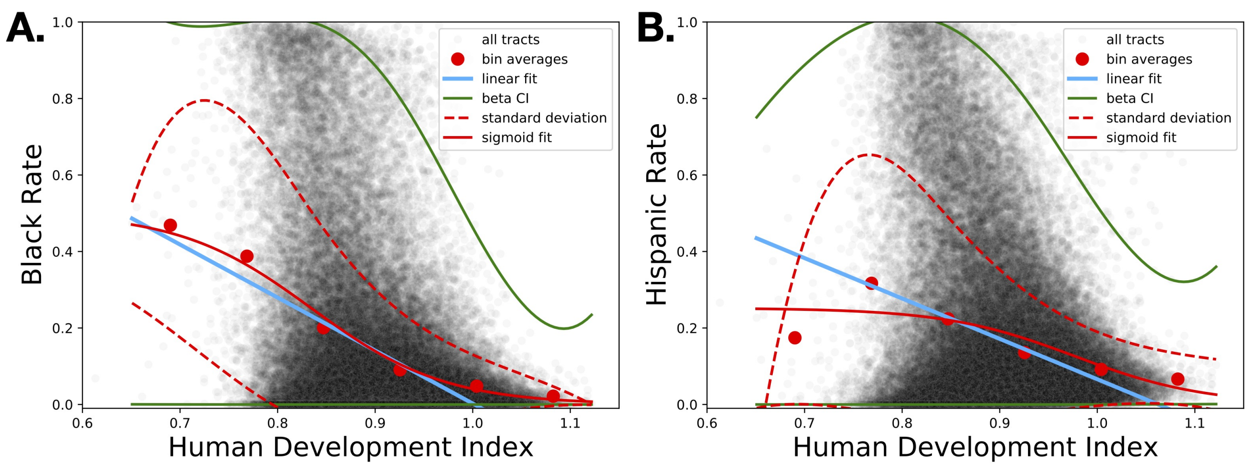


Figure 3.32: The correlation between the fraction of **A.** Black and **B.** Hispanic population fractions in 72349 census tracts nationwide versus the community HDI. While there is is a small negative association in both cases (gradient=$-1.056(0.026)$, $R^2 = 0.142$ for Black and gradient=$-1.385(0.025)$, $R^2 = 0.083$ for Hispanic), the data clearly show that a linear fit is inadequate and that there are many different situations at low human development, while at high HDI most tracts are mixed but fractions of minority population are small, especially for Black.

Table 3.1: Table of Density Distribution best fits to HDI census tract data in the US. Parameter names are as conventionally listed for the distribution in the first column. Log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (AD) are different goodness of fit criteria for preferring a specific distribution given data. The 2 parameter Normal distribution is gives the best fit among distributions listed.

Distribution	α	β	γ	μ	Sigma	Log-likelihood	AICc	BIC	AD
Normal 2P				0.905097	0.0571608	104426	-208849	-208831	9.6342
Lognormal 2P				-0.101714	0.0633808	104312	-208620	-208601	16.5568
Lognormal 3P			0	-0.101714	0.0633808	104312	-208618	-208590	16.5568
Loglogistic 2P	0.903917	27.4721				103444	-206884	-206865	79.4194
Loglogistic 3P	0.903917	27.4721	0			103444	-206882	-206854	79.4194
Weibull 3P	0.487969	8.66643	0.442444			101928	-203849	-203822	372.56
Weibull 2P	0.932037	16.5051				100465	-200926	-200908	590.255
Gumbel 2P				0.933802	0.056969	98521.3	-197039	-197020	879.804

Table 3.2: CHDI and Social Indicator Fits for Metropolitan and Micropolitan Areas Part 1

Dependent Variable vs. HDI	Linear Model for Metros and Micros $y = a - b \times hdi$	Linear Model for Metros Only $y = a - b \times hdi$
Teen Pregnancy	$a = 2.019$ (0.045)	$a = 1.911$ (0.065)
	$b = 2.041$ (0.050)	$b = 1.911$ (0.072)
	$R^2 = 0.643$	$R^2 = 0.648$
Incarceration	$a = 0.135$ (0.012)	$a = 0.121$ (0.010)
	$b = 0.135$ (0.014)	$b = 0.118$ (0.011)
	$R^2 = 0.308$	$R^2 = 0.215$
Will Not Make Top 20% by Income	$a = 2.047$ (0.040)	$a = 2.207$ (0.0.66)
	$b = 1.381$ (0.044)	$b = 1.554$ (0.073)
	$R^2 = 0.508$	$R^2 = 0.540$
Poor Mental Health	$a = 0.635$ (0.014)	$a = 0.626$ (0.024)
	$b = 0.541$ (0.016)	$b = 0.531$ (0.027)
	$R^2 = 0.555$	$R^2 = 0.500$
Obesity	$a = 1.383$ (0.036)	$a = 1.569$ (0.069)
	$b = 1.160$ (0.040)	$b = 1.371$ (0.077)
	$R^2 = 0.472$	$R^2 = 0.453$
Teeth Loss	$a = 1.261$ (0.023)	$a = 1.178$ (0.036)
	$b = 1.225$ (0.026)	$b = 1.135$ (0.040)
	$R^2 = 0.694$	$R^2 = 0.672$
Currently Smoking	$a = 1.000$ (0.026)	$a = 0.953$ (0.048)
	$b = 0.895$ (0.029)	$b = 0.847$ (0.053)
	$R^2 = 0.498$	$R^2 = 0.395$
Unemployment	$a = 0.389$ (0.018)	$a = 0.397$ (0.028)
	$b = 0.374$ (0.021)	$b = 0.379$ (0.031)
	$R^2 = 0.254$	$R^2 = 0.280$
Disability	$a = 0.839$ (0.029)	$a = 0.842$ (0.045)
	$b = 0.772$ (0.033)	$b = 0.779$ (0.050)
	$R^2 = 0.370$	$R^2 = 0.388$

Table 3.3: CHDI and Social Indicator Fits for Metropolitan and Micropolitan Areas Part 2

Dependent Variable vs. HDI	Linear Model for Metros and Micros $y = a - b \times hdi$	Linear Model for Metros Only $y = a - b \times hdi$
Children in Poverty	$a = 2.044$ (0.055)	$a = 1.939$ (0.072)
	$b = 2.063$ (0.061)	$b = 1.943$ (0.080)
	$R^2 = 0.547$	$R^2 = 0.606$
No Health Insurance	$a = 1.327$ (0.058)	$a = 1.320$ (0.096)
	$b = 1.340$ (0.065)	$b = 1.329$ (0.106)
	$R^2 = 0.312$	$R^2 = 0.289$
No Computing Device	$a = 0.938$ (0.031)	$a = 0.726$ (0.035)
	$b = 0.959$ (0.034)	$b = 0.730$ (0.039)
	$R^2 = 0.452$	$R^2 = 0.475$
Single Female Household	$a = 0.790$ (0.030)	$a = 0.0836$ (0.045)
	$b = 0.753$ (0.034)	$b = 0.792$ (0.050)
	$R^2 = 0.345$	$R^2 = 0.393$
Public Assistance	$a = 0.053$ (0.013)	$a = 0.107$ (0.024)
	$b = 0.033$ (0.014)	$b = 0.092$ (0.026)
	$R^2 = 0.005$	$R^2 = 0.030$
Crowded Housing	$a = 0.168$ (0.022)	$a = 0.169$ (0.044)
	$b = 0.159$ (0.024)	$b = 0.156$ (0.049)
	$R^2 = 0.042$	$R^2 = 0.025$
No Plumbing	$a = 0.436$ (0.022)	$a = 0.295$ (0.025)
	$b = 0.456$ (0.024)	$b = 0.305$ (0.027)
	$R^2 =0.270$	$R^2 = 0.241$
No Internet	$a = 1.796$ (0.045)	$a = 1.477$ (0.058)
	$b = 1.816$ (0.051)	$b = 1.475$ (0.064)
	$R^2 = 0.579$	$R^2 = 0.579$

Table 3.4: CHDI and Social Indicator Fits at Tract Level Part 1

Dependent Variable vs. HDI	Linear Model $y = a - b \times hdi$	Beta Average $y = \frac{y_0}{(1+e^{k(hdi-hdi_0)})}$	Beta Standard Deviation
Teen Pregnancy	$a = 1.662$ (0.004) $b = 1.630$ (0.005) $R^2 = 0.596$	$y_0 = 0.532$ (0.017) $k = 13.872$ (0.673) $hdi_0 = 0.853$ (0.006) $R^2 = 0.999$	$y_0^{std} = 0.116$ (0.004) $k_0^{std} = 13.317$ (1.375) $hdi_0^{std} = 0.971$ (0.008) $R^2 = 0.995$
Incarceration	$a = 0.149$ (0.0007) $b = 0.148$ (0.0008) $R^2 = 0.285$	$y_0 = 0.065$ (0.006) $k = 13.400$ (1.371) $hdi_0 = 0.801$ (0.018) $R^2 = 0.997$	$y_0^{std} = 0.159$ (0.050) $k_0^{std} = 7.610$ (0.405) $hdi_0^{std} = 0.572$ (0.062) $R^2 = 0.999$
Will Not Make Top 20% by Income	$a = 2.028$ (0.003) $b = 1.365$ (0.003) $R^2 = 0.642$	$y_0 = 0.991$ (0.015) $k = 7.339$ (0.470) $hdi_0 = 1.100$ (0.004) $R^2 = 0.998$	$y_0^{std} = 0.086$ (0.018) $k_0^{std} = -8.753$ (5.403) $hdi_0^{std} = 0.785$ (0.059) $R^2 = 0.870$
Poor Mental Health	$a = 0.577$ (0.001) $b = 0.480$ (0.001) $R^2 = 0.683$	$y_0 = 0.225$ (0.039) $k = 7.009$ (2.854) $hdi_0 = 0.993$ (0.058) $R^2 = 0.949$	$y_0^{std} = 2.613e+03$ (1.693e+08) $k_0^{std} = 3.410$ (2.181) $hdi_0^{std} = -2.559$ (1.902e+04) $R^2 = 0.969$
Obesity	$a = 1.120$ (0.002) $b = 0.883$ (0.002) $R^2 = 0.567$	$y_0 = 0.500$ (0.073) $k = 5.972$ (2.010) $hdi_0 = 1.012$ (0.053) $R^2 = 0.971$	$y_0^{std} = 0.069$ (0.006) $k_0^{std} = 5.265$ (1.427) $hdi_0^{std} = 1.070$ (0.034) $R^2 = 0.984$
Teeth Loss	$a = 1.127$ (0.002) $b = 1.078$ (0.002) $R^2 = 0.699$	$y_0 = 0.325$ (0.047) $k = 12.660$ (4.140) $hdi_0 = 0.901$ (0.035) $R^2 = 0.959$	$y_0^{std} = 0.155$ (0.042) $k_0^{std} = 7.366$ (1.395) $hdi_0^{std} = 0.758$ (0.078) $R^2 = 0.992$
Currently Smoking	$a = 0.911$ (0.002) $b = 0.805$ (0.002) $R^2 = 0.647$	$y_0 = 0.362$ (0.024) $k = 7.787$ (0.838) $hdi_0 = 0.909$ (0.021) $R^2 = 0.996$	$y_0^{std} = 0.074$ (0.003) $k_0^{std} = 7.988$ (0.548) $hdi_0^{std} = 0.904$ (0.013) $R^2 = 0.998$
Unemployment	$a = 0.376$ (0.002) $b = 0.352$ (0.002) $R^2 = 0.222$	$y_0 = 6.746e+04$ (1.736e+10) $k = 8.426$ (4.684) $hdi_0 = -7.744e-01$ (3.061e+04) $R^2 = 0.966$	$y_0^{std} = 5.723e+04$ (1.933e+10) $k_0^{std} = 1.450e+01$ (1.243e+01) $hdi_0^{std} = -1.3847e-01$ (2.334e+04) $R^2 = 0.967$
Disability	$a = 0.585$ (0.002) $b = 0.499$ (0.003) $R^2 = 0.247$	$y_0 = 0.248$ (0.028) $k = 6.538$ (1.124) $hdi_0 = 0.929$ (0.040) $R^2 = 0.992$	$y_0^{std} = 4.161e+04$ (1.184e+10) $k_0^{std} = 6.429$ (4.237) $hdi_0^{std} = -1.226$ (4.436e+04) $R^2 = 0.949$

Table 3.5: CHDI and Social Indicator Fits at Tract Level Part 2

Dependent Variable vs. HDI	Linear Model $y = a - b \times hdi$	Beta Average $y = \frac{y_0}{(1+e^{k(hdi-hdi_0)})}$	Beta Standard Deviation
Children in Poverty	$a = 1.965$ (0.007)	$y_0 = 0.500$ (0.041)	$y_0^{std} = 1.982e+04$ (1.117e+09)
	$b = 1.961$ (0.008)	$k = 18.757$ (4.159)	$k_0^{std} = 4.580$ (2.625)
	$R^2 = 0.427$	$hdi_0 = 0.872$ (0.015)	$hdi_0^{std} = -1.695$ (1.233e+04)
		$R^2 = 0.986$	$R^2 = 0.968$
No Health Insurance	$a = 0.998$ (0.004)	$y_0 = 0.213$ (0.034)	$y_0^{std} = 0.299$ (0.026)
	$b = 0.965$ (0.005)	$k = 18.602$ (10.268)	$k_0^{std} = 7.812$ (0.468)
	$R^2 = 0.332$	$hdi_0 = 0.932$ (0.035)	$hdi_0^{std} = 0.764$ (0.024)
		$R^2 = 0.896$	$R^2 = 0.999$
No Computing Device	$a = 0.645$ (0.002)	$y_0 = 0.318$ (0.020)	$y_0^{std} = 7.556e+04$ (1.883e+10)
	$b = 0.635$ (0.003)	$k = 11.037$ (0.624)	$k_0^{std} = 9.672$ (4.514)
	$R^2 = 0.340$	$hdi_0 = 0.781$ (0.013)	$hdi_0^{std} = -5.973e-01$ (2.582e+04)
		$R^2 = 0.999$	$R^2 = 0.980$
Single Female Household	$a = 0.887$ (0.003)	$y_0 = 0.650$ (0.168)	$y_0^{std} = 7.868e+04$ (2.182e+10)
	$b = 0.836$ (0.004)	$k = 6.959$ (0.822)	$k_0^{std} = 8.734$ (5.239)
	$R^2 = 0.346$	$hdi_0 = 0.701$ (0.070)	$hdi_0^{std} = -7.218e-01$ (3.184e+04)
		$R^2 = 0.997$	$R^2 = 0.961$
Public Assistance	$a = 0.216$ (0.001)	$y_0 = 0.072$ (0.007)	$y_0^{std} = 0.379$ (0.159)
	$b = 0.210$ (0.001)	$k = 13.045$ (2.290)	$k_0^{std} = 7.018$ (0.483)
	$R^2 = 0.167$	$hdi_0 = 0.856$ (0.022)	$hdi_0^{std} = 0.538$ (0.088)
		$R^2 = 0.990$	$R^2 = 0.999$
Crowded Housing	$a = 0.295$ (0.002)	$y_0 = 0.057$ (1.813e-02)	$y_0^{std} = 0.080$ (0.023)
	$b = 0.287$ (0.003)	$k = 18.572$ (2.114e+01)	$k_0^{std} = 13.413$ (10.606)
	$R^2 = 0.105$	$hdi_0 = 0.945$ (7.318e-02)	$hdi_0^{std} = 0.944$ (0.072)
		$R^2 = 0.639$	$R^2 = 0.775$
No Plumbing	$a = 0.268$ (0.002)	$y_0 = 0.070$ (8.362e-03)	$y_0^{std} = 0.131$ (0.014)
	$b = 0.270$ (0.002)	$k = 22.253$ (8.558)	$k_0^{std} = 12.826$ (2.020)
	$R^2 = 0.136$	$hdi_0 = 0.863$ (2.149e-02)	$hdi_0^{std} = 0.824$ (0.023)
		$R^2 = 0.962$	$R^2 = 0.993$
No Internet	$a = 1.283$ (0.004)	$y_0 = 0.334$ (0.035)	$y_0^{std} = 7.865e+04$ (1.875e+10)
	$b = 1.25101972$ (0.004)	$k = 15.32755914$ (3.960)	$k_0^{std} = 7.166$ (3.627)
	$R^2 = 0.480$	$hdi_0 = 0.894$ (0.023)	$hdi_0^{std} = -1.039$ (3.334e+04)
		$R^2 = 0.977$	$R^2 = 0.969$

Table 3.6: Principal Components

Variables	First Component	Second Component
Explained Relative Variance	54.5%	8.7%
HDI	1.00	1.00
Teen Pregnancy	-0.956	-2.034
Incarceration	-0.749	-1.928
Will Not Make Top 20% by Income	-0.896	0.032
Poor Mental Health	-0.999	-0.425
Obesity	-0.929	3.113
Teeth Lost	-1.047	0.478
Currently Smoking	-0.968	3.739
Unemployment	-0.683	-2.041
Disability	-0.652	8.628
Children in Poverty	-0.880	-2.639
No Health Insurance	-0.667	-5.061
No Computing Device	-0.802	4.791
Single Female Head Household	-0.759	-6.677
Public Assistance	-0.539	-4.978
Crowded Housing	-0.318	-12.448
No Plumbing	-0.556	6.678
No Internet	-0.903	3.769

CHAPTER 4

COMMUNITY HUMAN DEVELOPMENT INDICATORS IN US COUNTIES: AN ANALYSIS OF THE DIVERGENCE OF LIFE EXPECTANCIES ACROSS THE UNITED STATES OVER TIME

4.1 Introduction

Understanding changes in human development over time is critical for developing scientific frameworks that support this complex phenomenon and for formulating policy interventions that improve people's quality of life [85, 86]. One general approach for conceptualizing human development is the *capabilities approach*. The capabilities approach goes beyond political concepts of liberalism tied to (often unenforceable) general freedoms by asking how physical, social and institutional living environments actually promote individual, context appropriate capabilities (i.e. agency and success) [1, 27, 87]. Formulated primarily by Amartya Sen and Martha Nussbaum, the approach has been widely adopted by economists, development researchers, and NGOs as a way of measuring development across a wide range of communities and contexts [85, 88, 43]. The development economist Mahbub ul Haq adapted the capabilities approach at the United Nations Development Programme and sought to quantify its concepts by creating the Human Development Index (HDI). The HDI is a metric designed to go beyond traditional measures of development based on Gross Domestic Product (GDP), by measuring a non-substitutible composite of education attainment, real income, and life expectancy [26]. Since its introduction in late 1980s, the HDI has been improved and used to measure changes in patterns of development primarily at the national level. These applications, and their relation to other quantities relevant for development, have led to a rich set of analyses about how countries change and develop holistically over time [89, 90].

A major area of study in human development deals with whether development across different units (nations, regions) is convergent or divergent [91, 92]. Convergent development means that different units manifest over time more similar levels of development, usually with the implication that initially less developed units experience faster progress and catch up with more developed ones [93]. Divergent development implies, by contrast, that over time metrics of development in different regions become more different, expressing increasing inequality of human experiences and capabilities. Previous studies have examined in what circumstances nations move toward similar levels of development, with the expectation that initially underdeveloped nations would change rapidly at first and more slowly at later times, shaping a general process of convergence [94].

For the components of the HDI- Education, Income, and Life Expectancy- multiple studies have documented that the majority of developing countries have broadly converged toward development benchmarks, with inequality decreasing across the world [91, 95]. A recent United Nations report, in particular, found that between 1970 and 2005 low HDI countries achieved faster growth in development than high HDI countries, and that this growth was primarily achieved through improvements in life expectancy and education [96], emphasizing the importance of not focusing solely on national accounts. Notably, rates of convergence have also been uneven, with faster convergence in parts of East Asia and Eastern Europe, and slower rates of progress in Sub-Saharan Africa [97] over this period. Some studies have also noted this convergence at subnational and regional scales, though others have identified growing within-nation inequality, particularly in low- and middle-income countries [9, 98].

Fewer studies have examined these questions at the metropolitan (city) scale, where intra-national growth patterns can also diverge (or converge) over time. This question is of great interest everywhere, but assumes particular importance within large nations with distributed (federal) governments such as the USA, where subnational units such as states, cities and counties exert important powers over critical components of human development, such as education, healthcare, services and housing policy.

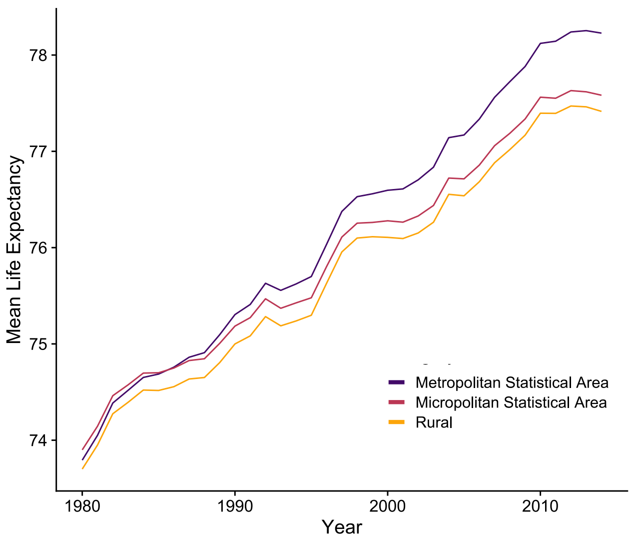
Countries typically show high degrees of internal heterogeneity, with regional differences in economic outputs, access to services, and levels of urbanization [8, 99, 100]. Significant policies are enacted at the subnational level. To understand the relationship between policy, intervention, and development, growth and change must also be measured at multiple geographic scales. New datasets have enabled measurements of growth at the local level, enabling newer, richer analyses of change [21, 30]. Here, by combining data from a variety of emerging sources, we have created a comprehensive dataset showing the evolution of development index values at the county level in the United States over time. We find that education, income, and life expectancy indices all display growth over time in the United States, with a Weibull distribution in life expectancy and a Log-logistic distribution for education and income indices across counties. We also find that income and education neither significantly converged nor diverged across counties from 2010 to 2019, while life expectancy diverged rapidly over the period of 1980 to 2014.

4.2 Results

4.2.1 Divergent Trends in Life Expectancy

Life expectancy at birth, on average, increased in the United States from 1980 to 2014, from 73.8 to 79.1. The greatest single gain in life expectancy in a one year period was an increase of 1.35, in Bronx County, NY, in 1997, while the greatest decrease was a decline of 1.05 years in Baltimore City in 2013. Over the 34 year period study, the Aleutians East Borough saw the greatest increase in life expectancy, of 12.958 years. A number of other Alaskan counties also saw major increases in life expectancy over this time period, including the Aleutians West Census Area, the North Borough, and the Kodiak Island Borough. These increases are likely attributable to systematic improvements in local health care access, improved

A



B

Metropolitan Area	Difference From 1980
Cape Coral-Fort Myers-Naples, FL	7.8815
Edwards-Glenwood Springs, CO	7.2362
New York-Newark, NY-NJ-CT-PA	6.7669
San Jose-San Francisco-Oakland, CA	6.6612
North Port-Sarasota, FL	6.3582
Los Angeles-Long Beach, CA	6.2799
Washington-Baltimore-Arlington, DC-MD-VA-WV-PA	6.2362
Miami-Fort Lauderdale-Port St. Lucie, FL	6.1295
Raleigh-Durham-Chapel Hill, NC	6.1020
Rapid City-Spearfish, SD	5.9463
Amarillo-Borger, TX	2.4806
Wichita-Arkansas City-Winfield, KS	2.4705
Richmond-Connersville, IN	2.4218
Martin-Union City, TN-KY	2.3853
Oklahoma City-Shawnee, OK	2.3389
Joplin-Miami, MO-OK	2.3231
Hot Springs-Malvern, AR	2.2315
Johnson City-Kingsport-Bristol, TN-VA	2.1755
Tulsa-Muskogee-Bartlesville, OK	1.9935
Cleveland-Indianola, MS	1.7551
Jonesboro-Paragould, AR	1.7167

C

County Name	Difference From 1980
Aleutians East Borough, AK	12.9582
Aleutians West Census Area, AK	12.9582
North Slope Borough, AK	11.7270
New York County, NY	10.8175
Sumter County, FL	9.8818
San Francisco County, CA	9.7692
Kings County, NY	9.3830
Kodiak Island Borough, AK	9.2654
Loudoun County, VA	9.2078
Collier County, FL	8.9509
Grundy County, TN	-0.2611
Perry County, AL	-0.4039
Kiowa County, OK	-0.5370
Perry County, KY	-0.5704
Estill County, KY	-0.7068
Powell County, KY	-0.8128
Clay County, KY	-0.9643
Breathitt County, KY	-1.0166
Leslie County, KY	-1.3832
Lee County, KY	-1.4451
Owsley County, KY	-2.2043

Figure 4.1: Change in Life Expectancy By Level of Urbanization. From 1980 to 2014, all categories saw average increases in life expectancy. However, metropolitan areas saw larger gains after the mid-1990s. (B) Top 10 and Bottom 10 Metropolitan Areas based on gain in life expectancy at birth from 1980. (C) Top 10 and Bottom 10 Counties based on change in life expectancy at birth from 1980.

sanitation and water systems, and preventative measures to reduce unintentional deaths, such as drownings [101, 102].

Despite average national increases in life expectancy, several counties saw life expectancy decreases over this period. Owsley County, Kentucky experienced the largest decrease over this period, with a drop of 2.204 years. Of the 13 counties that saw life expectancy declines in this period, 11 are located in Appalachia, as defined by the Appalachian Regional Commission [103]. The exceptions are Kiowa County, OK, and Perry County, AL. The root causes of this concerning trend are diverse- the region has higher than average rates of chronic disease and unintentional deaths (including drug overdoses), and both infant and mid-life mortality are higher in Appalachia than other parts of the United States [104, 105].

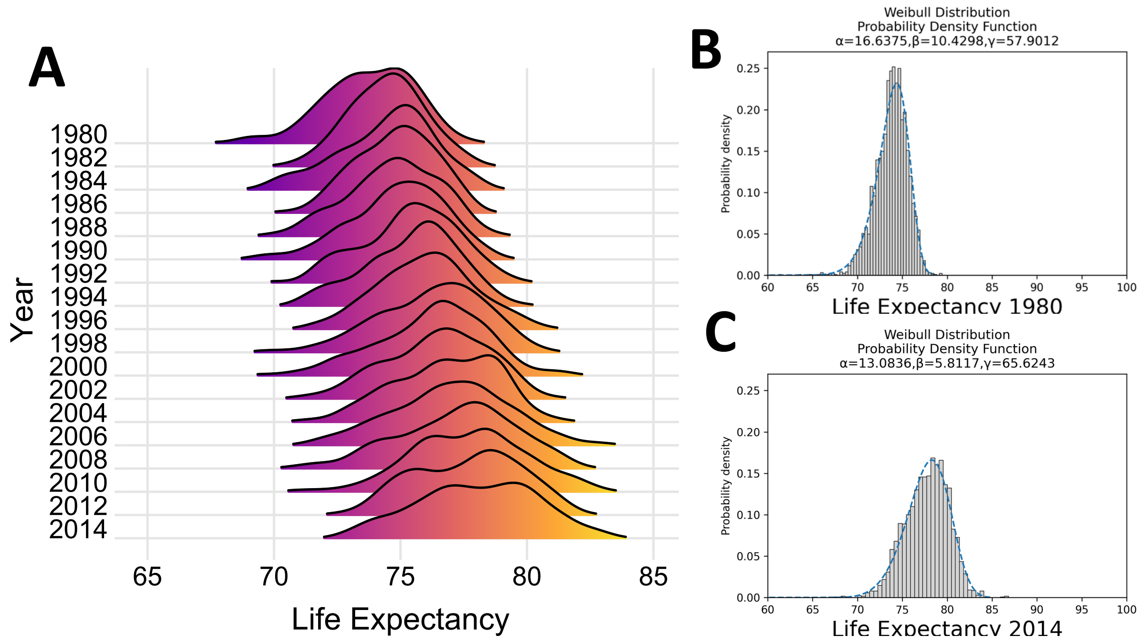


Figure 4.2: Changes in Life Expectancy Distributions across US Counties over time. (A) Ridgeline plot showing life expectancy distributions for even years. From 1980 to 2014, the distribution shifted rightward and widened. Life expectancy is well fit by a Weibull 3-parameter distribution, as shown in (B) and (C) and the Supplement.

Previous studies have observed a divergence between urban and rural life expectancies over time [106, 107]. We observe this phenomenon in our analysis as well. In 1980, the average life expectancy for residents of metropolitan areas (larger urban areas with populations greater than 50,000) was 73.79 years, and for rural areas it was 73.70, a negligible difference. However, the urban-rural difference has grown over time, with larger gaps emerging in the mid-1990s, see Figure 4.1 when urban life expectancy increased faster. In 2014, the gap between the two categories was 0.8 years. Micropolitan areas, which are urbanized environments with populations between 10,000 and 50, 000, had the highest average life expectancy of the three categories in 1980, with a value of 73.89, but since then have only been marginally higher than the rural life expectancy average. The growing urban advantage in life expectancy has persisted in more recent years, and is believed to be a product of multiple factors,including reductions in deaths from infectious diseases such as HIV, the end of the crack-cocaine crisis, and improved access to heath care over time [106]. Conversely, health care and living conditions have not improved to the same extent in US rural areas, and have in many cases degraded from previous levels with the epidemic of opioids and large suicide rates (most frequently by gunshot) primarily a rural phenomenon [108, 109, 110].

Geographic differences also emerge when comparing the gains made in life expectancy between different metropolitan areas. While all metropolitan areas in the dataset increased in life expectancy since 1980, Naples, FL, saw the largest gain, of 7.88 years, see Figure 4.1. This may be due to the rapid population and economic growth the city and surrounding counties experienced from 1995 to 2007 [111, 112]. Other major cities, including New York, Los Angeles, San Francisco, and Miami, also saw major life expectancy gains. Meanwhile, most cities with the smallest gains, including Jonesboro, AR and Cleveland, MS, were concentrated primarily in the South. Midwestern metropolitan areas, such as Chicago-Naperville, IL-IN-WI, and Detroit-Warren-Ann Arbor, MI, on average had middling gains (4.76 years and 4.55 years, respectively).

Statistically, the distribution of life expectancies across US counties is well modeled by a Weibull distribution with three parameters, see Figure 4.2. A Weibull distribution has the form $\mathrm{f(t)} = \frac{\beta t^{\beta-1}}{\alpha^{\beta}} \mathrm{e}^{-(\frac{t}{\alpha})^{\beta}}$, where $\alpha > 0$ is a scale parameter, β is a shape parameter, and the distribution can be shifted to the left or right by a γ parameter. This distribution is used in a wide variety of applications, including weather forecasting, reliability analysis, and survival analysis [113], which make sense in the context of life expectancy. For the county life expectancies from 1980 to 2014, both the alpha and beta parameter decreased over time, while the gamma parameter increased, reflecting the increasing average life expectancy, see Figure 4.1. Plotting the difference in county life expectancies from the national average over time, and setting a common starting point at the beginning of the data set in 1980, shows the rapid divergence in life expectancy at the county level over time, with the highest

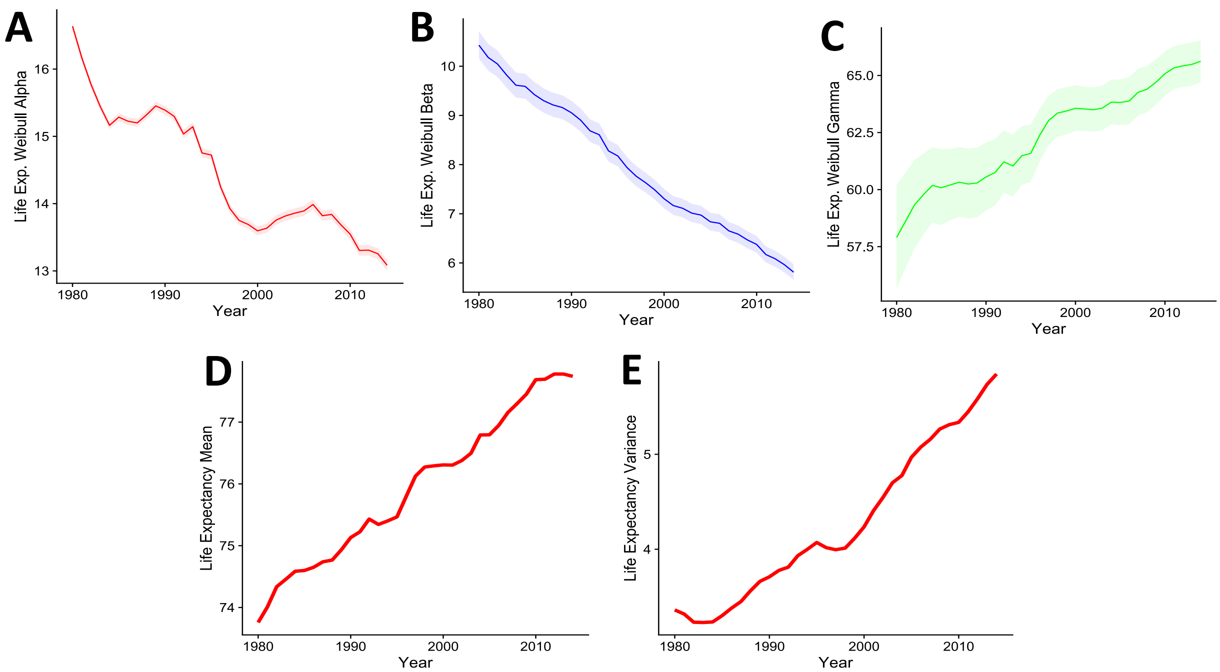


Figure 4.3: Change in Life Expectancy Distribution Parameters Over Time. Life Expectancy in the US is fit well by a three-parameter Weibull Distribution. (A) Alpha and (B) Beta parameters decrease over time, while the right shift in the distribution (C) gamma increases over time, showing the overall increase in life expectancy (D) mean. The concurrent decreases in the alpha and beta parameters reflect the rapidly increasing (E) variance over time. 95% confidence interval is shown as shaded region surrounding lines.

counties showing rapidly increasing gains over the national average, and the converse for counties with lower than average life expectancies, see Figure 4.4. County life expectancies have an increasing coefficient of variation as the mean life expectancy increases, indicating that the standard deviation increased at a faster rate than the mean over this time period, see Figure 4.8. These divergences have likely increased since 2014, as improvements in life expectancy have largely stalled across the US, and even declined in some places [107].

4.2.2 Trends in Income and Education

Both income and education indices were calculated at the county level based on data released by the American Community Survey. As such, they are available from 2010 to 2019, a smaller but more recent time frame than the life expectancy data.

The income index, adjusted regionally for real incomes by the US Census consumer price index (CPI), rose from 0.935 in 2010 to 0.979 in 2019. Glenwood Springs, CO had the highest population weighted average Income Index (II) value in 2019, with a value of 1.028. Cleveland-Indianola, MS had the lowest average II value in 2019, at 0.901. The greatest increases in the Income Index over this time period were in a mix of smaller metropolitan areas, such as Fargo-Wahpeton, ND-MN (increase of 0.059) and mid-sized cities, such as Pittsburgh-New Castle-Weirton, PA-OH-WV (increase of 0.053). Metropolitan areas that saw the smallest increases tended to be cities that already had very high income index values in 2010, such as Miami-Fort Lauderdale-Port St. Lucie, FL (increase of 0.029). In keeping with this trend, the largest one-year increase in value was for Cleveland-Indianola, MS in 2014, when it increased by a value of 0.01, to an adjusted II value of 0.885.

Education indices tend to be relatively stable over time, as the education levels of the adult population (measured by the mean education index) and school enrollment rates (as measured by the expected education index) do not fluctuate significantly at the county level over this period. The population weighted average expected education index rose from 0.893 to 0.899, while the mean education index rose 0.860 to .886. Most of the gains in mean education from 2010 to 2019 were in smaller, southern metropolitan areas, including Cape Girardeau-Sikeston, MO-IL, which saw a gain of 0.05, and the Columbus-Auburn-Opelika, GA-AL MSA, with a gain of 0.04. Several smaller towns also saw losses in the educated adult population over this time, including Kokomo-Peru, IN and Moses Lake-Othello, WA. Most of these towns experienced sustained decreases over time, dropping from year to year. While the overall drops were small (approximately 0.01 or less), they represent a surprising sustained decrease in an important development metric over time.

In order to gain a deeper understanding of the underlying dynamics of these quantities, and to build predictive models moving forward, we fitted multiple probability distributions

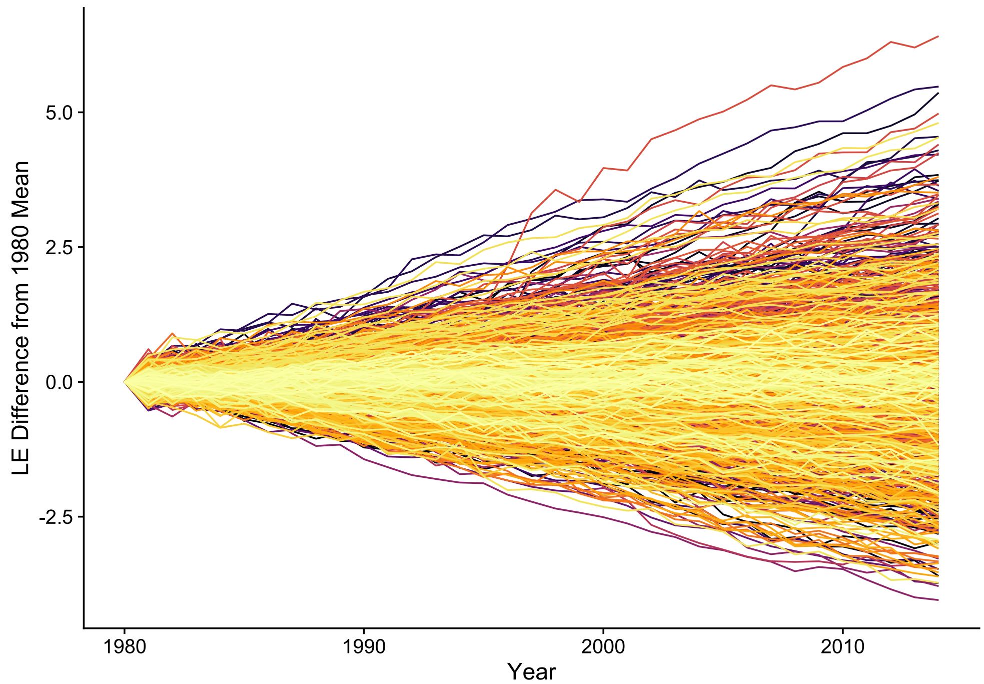


Figure 4.4: Life Expectancy Difference From Mean Over Time. To demonstrate divergence in values over time, 1980 life expectancy values have been set as the origin. On average, counties that had higher than average life expectancy in 1980 only increased in life expectancy gains over time, deviating farther from the average. The outlier at the top of the graph is Bronx County, New York, which saw major life expectancy gains over this period.

to the county-level income index, mean education index, and expected education values for multiple years. Distribution fits were evaluated based on log-likelihood values, calculated using the python package *reliability*. Both education indices and the income index are best characterized as Log-logistic distributions, which is a logistic function of the logarithm of the index see Tables 4.2 to 4.4 and Figures 4.5 to 4.7 and 4.9. This distribution form has a probability density function of $f(t) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{t}{\alpha}\right)^{\beta}\right)^2}$ where $\alpha > 0$ is once again a scale parameter, as well as the median of the distribution, and β is a shape parameter. The distribution often captures the dynamics of a system that has rates which increase initially and then slow down or saturate (like a logistic function), and has been widely used in economics to model incomes of populations[114].

4.3 Discussion

Understanding human development at the subnational level is essential for achieving more equitable human development and meeting local and state sustainability plans and international agreements, such as the United Nations Sustainable Development Goals [21]. Longitudinal assessments of local development also allow for more granular analyses of how local infrastructure, policy, and economic shifts impact the human capabilities and well being of populations. Our study shows that from 2010 to 2014 the United States has seen steady overall growth in development in almost all counties, but also persistent disparities. US metropolitan areas (larger cities) show at present the highest overall levels of human development and have improved the most over the last few decades, especially in terms of life expectancy.

After separating the subindices of the Human Development Index, we find that almost all counties have experienced growth in life expectancy, income and education. However, this growth has been distributed unequally, with large metropolitan areas seeing larger gains in life expectancy after the late-1990s, and rural counties, largely in Appalachia and Southeastern United States, stalling and even experiencing drops in life expectancy up to 2014. Recent studies have shown that an overall stalling of life expectancy has occurred across the United States, even prior to the COVID-19 pandemic, largely due to the opioid epidemic and other "deaths of despair" and also some increases in neurovascular deaths [108]. Further studies will have to examine whether these emerging challenges are leading to even more extreme divergences in life expectancies across US counties and regions. In terms of education and income, some small and mid-sized metropolitan areas have seen growth, but many have also had declines from 2010 to 2019.

We observe, consequently, that levels of human development across US counties have not tended to converge over time. Income and education have shown a consistent level of variation and dispersion over the last decade, while life expectancy, measured consistently between 1980 to 2014, showed increasing dispersion. Life expectancy has alternated between periods of major increases and periods of slow growth across the last 34 years. The root causes for each stalling period can vary- for example, the stalling period observed in the 1980's was largely due to the AIDS epidemic, while the period of improvement afterwards was partly due to the antiretroviral therapies becoming more widely available [115, 116]. While some health trends are universal across the US, including rising obesity and drug overdoses, the relative impact of each of these trends on mortality vary significantly by county [108, 117, 118].

The counties that saw the greatest increases in life expectancy, such as Aleutians East Borough and other Alaskan counties, made gains by taking an approach that was tailored to local needs. For example, to deliver medical care to rural Native American communities throughout the state, the Alaskan Tribal Health System has a Community Health Aide Program (CHA/P), where trained aides and practitioners embedded in communities provide front-line medical care, with support from physicians at larger regional centers. This system, along with steady improvements in infrastructure and sanitation systems, have likely helped reduce deaths from unintentional injuries and infectious disease [119, 101, 120, 102].

Our study is limited by the lack of a consistent life expectancy dataset from 2014 on, and the lack of longitudinal county level income and education data prior to 2009. The lack of consistent data, specifically life expectancy data, at smaller geographic scales poses a significant challenge to tracking and improving human development at the local level but can be tackled by a systematic analysis of death recoprds and the adoption of transparent life history modeling. This remains a practical challenge at present. As we have demonstrated, significant variation exists at the county level, and promoting equitable health, education, and economic outcomes will required a more sustained, organized effort to collect and report relevant data, including health histories and causes of death along with life expectancy estimates.

Nevertheless, the present study contributes to setting a baseline understanding for trends in development metrics in the US prior to the COVID-19 pandemic. The pandemic has been highly disruptive to social, political, educational and economic systems across the globe, representing one of the largest challenges to human development in recent history [121, 122, 123]. As data emerges quantifying the impacts of the pandemic on populations, studies should examine how it disrupted the local development trends described in this paper. Future studies should also examine changes in subnational development in a wider range of countries, to gain a broader understanding of how economic, education, and health related change occurs in a number of different contexts.

4.4 Materials and Methods

4.4.1 Data Sources

Life expectancy data was obtained from the Institute for Health Metrics and Evaluation's *US Health Map*. Data on school enrollment rates, average level of education, and income levels were taken from the American Community Survey 1-year and 5-year estimates. Percent Purchasing Power, used to calculate real incomes across counties, was taken from the Bureau of Economic Analysis.

4.4.2 Calculating the Community Human Development Index

The Human Development Index is made up of three subindices: the life expectancy index (I_{LE}), the education index (I_E), and the income index (I_{RI}). The HDI is calculated as the geometric mean of these subindices, $HDI = (I_{\mathrm{LE}} \times I_{\mathrm{E}} \times I_{\mathrm{RI}})^{1/3}$. The HDI's construction as a geometric mean prevents any one of the subindices to dominate the performance of the overall index. For a county to have a high HDI value, it must have simultaneously high scores in all three subindices.

The life expectancy index, I_{LE} is calculated simply as an indexed value of life expectancy at birth LE, normalized to a maximum of 85 years and a minimum of 20 years, $I_{LE} = (LE - 20)/(85 - 20)$. The education index is made up of two subindices, accounting for mean years of schooling (MYS) and expected years of schooling (EYS). The mean years of schooling index applies to adults (≥ 25 years old) and is computed as the population average, $MYS = \sum_s n(s) \times Y(s)$, where $n(s)$ is the fraction of the adult population who attained education level s, and $Y(s)$ is the number of years necessary to achieve such education level designated by the International Standard Classification of Education (ISCED). This measure is available for the American Community Survey 1-year estimates from 2010 on, and for the 5-year estimates from 2012 on. As the ACS 1 year estimates are only calculated for counties with 65,000 or more people, the HDI trends presented here is calculated for counties with 65,000 individuals of more from 2010 to 2014.

The expected years of schooling index applies to younger populations (< 25 years old), who may still be in school. The Index is measured as the probable mean final educational attainment if current school enrollment rates hold, $EYS = \sum_a n_r(a) \times Y_e(a)$, where a is age, $Y_e(a)$ is the expected total years of schooling for age cohort a and $n_r(a)$ is the fraction of the population enrolled in school at age a. The final education index is calculated as the average of the two subindices with their respective goal values, $I_{\mathrm{E}} = 1/2(EYS/18 + MYS/15)$.

Finally, the real Income Index, I_{RI} is usually calculated at the national level using Gross National Income (GNI) per capita (gni), which has no simple subnational equivalent. To create a meaningful definition in small areas, we use average personal income per capita data, I_{pc}, as $\hat{gni} = c_{\mathrm{GNI/I}} I_{\mathrm{pc}}$, with $c_{\mathrm{GNI/I}} = GNI/I$, the ratio of national GNI to total personal income. To create real incomes, we adjust nominal incomes for cost of living at the local level. The US Bureau of of Economic Analysis publishes a local purchasing power parity index (PPP) for metropolitan areas and states, which we use as $\hat{\mathrm{gni}}_{PPP} = c_{\mathrm{GNI/I}}\ I_{\mathrm{pc}}/PPP$, which now applies to each level of geographic aggregation, including states, metropolitan areas and tracts. Three MSAs,(Enid, OK, Poughkeepsie, NY, and Twin Falls, ID) did not have Purchasing Power Parities reported for some years. For these three years, the last reported value was substituted for the missing values. The income index is then $I_{RI} = (\log \hat{gni}_{PPP} - \log 100)/(\log(75,000) - \log(100))$, normalized to a minimum of \$100 real dollars per person/year and to a maximum of \$75,000.

4.5 Supplementary Materials

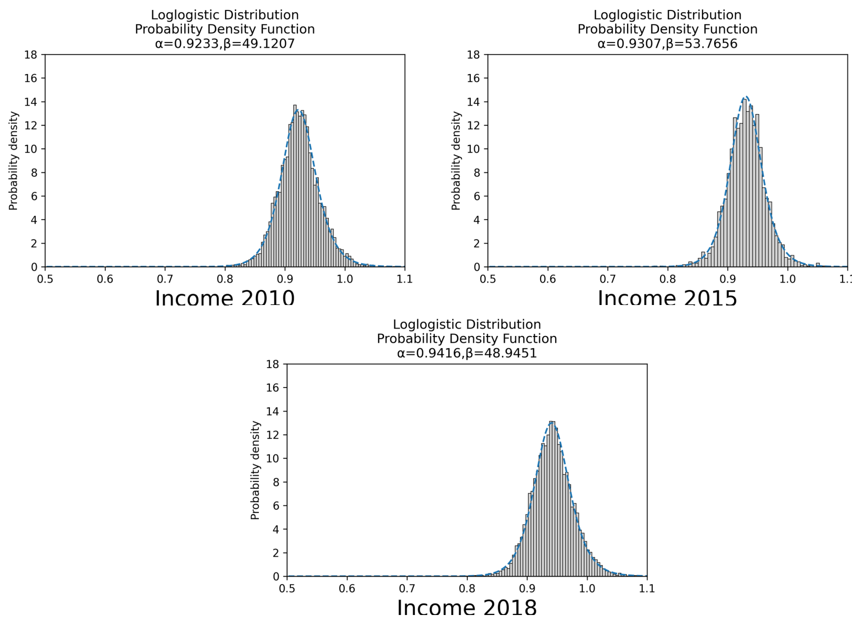


Figure 4.5: Distribution of County-level Income Index values at three time points, fit by a log-logistic curve. Income index values gradually increased over time, but distribution shape parameters did not change significantly.Comparison of fits can be found in Table 4.2

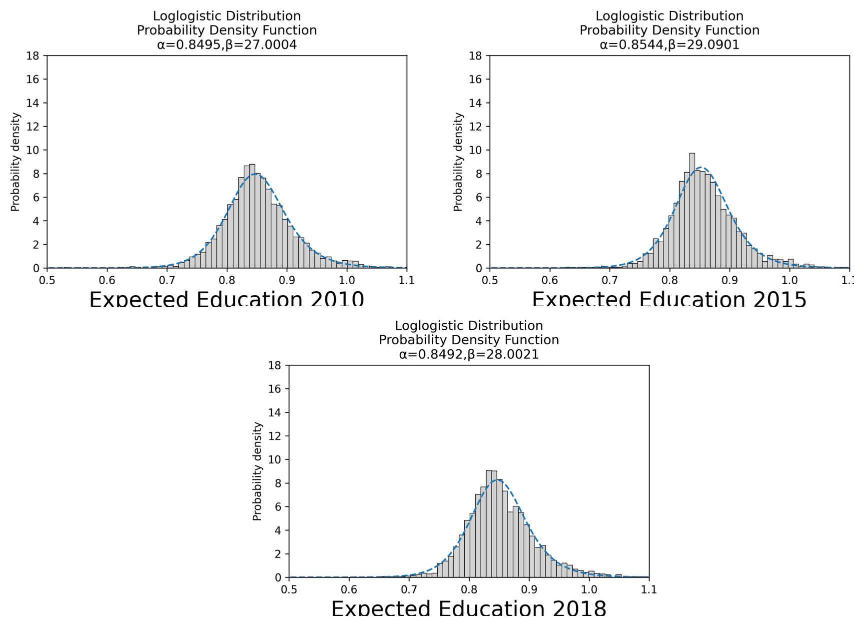


Figure 4.6: Distribution of County-level Expected Education Index values at three time points, fit by a log-logistic curve. Distribution shape parameters did not significantly change in the time period under study. Comparison of fits can be found in Table 4.3

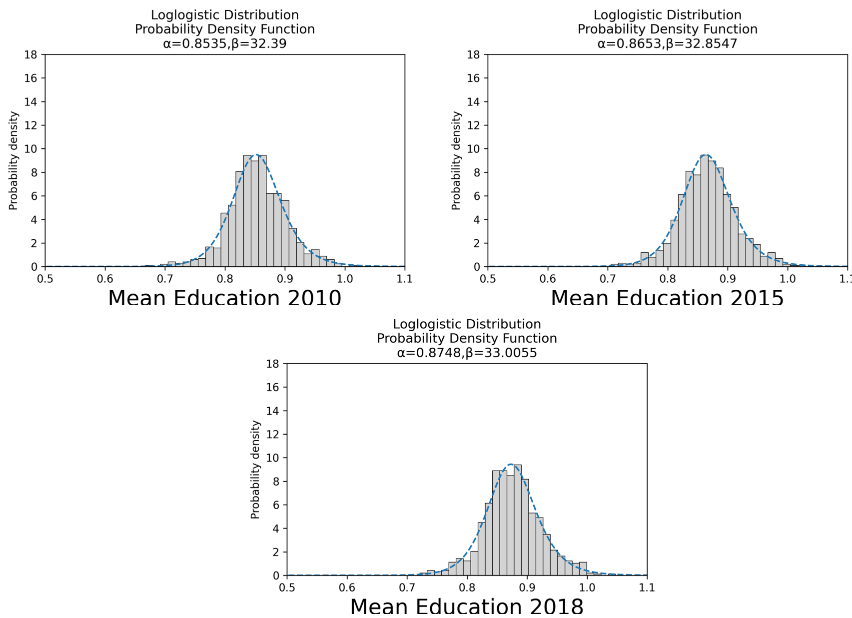


Figure 4.7: Distribution of County-level Mean Education Index values at three time points, fit by a log-logistic curve. Distribution shape parameters did not significantly change in the time period under study.Comparison of fits can be found in Table 4.4

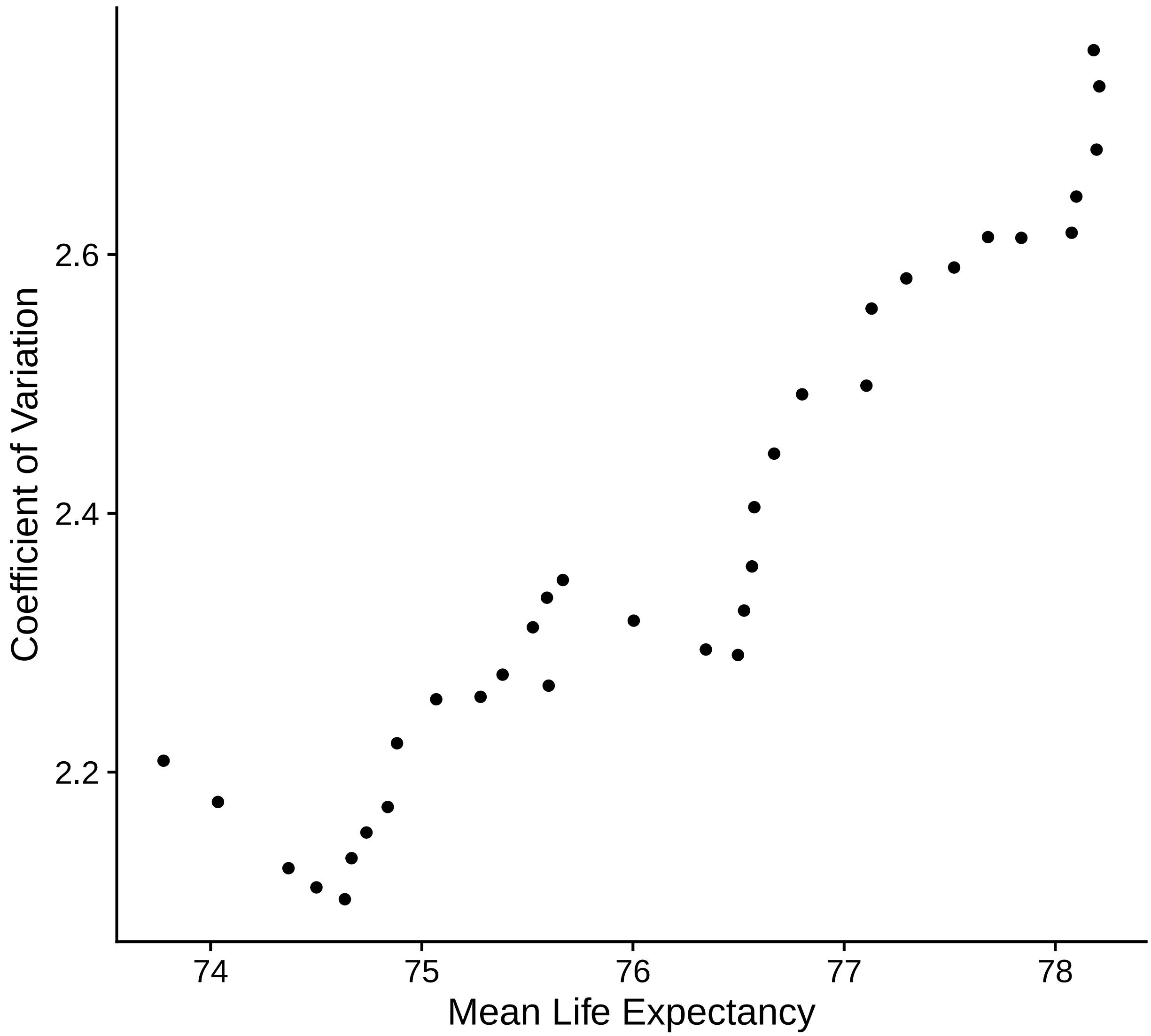


Figure 4.8: Comparison of the Coefficient of Variation County-level Life Expectancy values vs Mean Life Expectancy. As the mean is correlated with year in this dataset, the points moving from left to right also represent subsequent years. Note that in a null model, where standard deviation is increasing at the same rate as the mean, the coefficient of variation plotted against the mean would be a flat, horizontal line. We observe a general trend of the coefficient of variation increasing as the mean increases, indicating that the standard deviation is increasing at a faster rate than the mean.

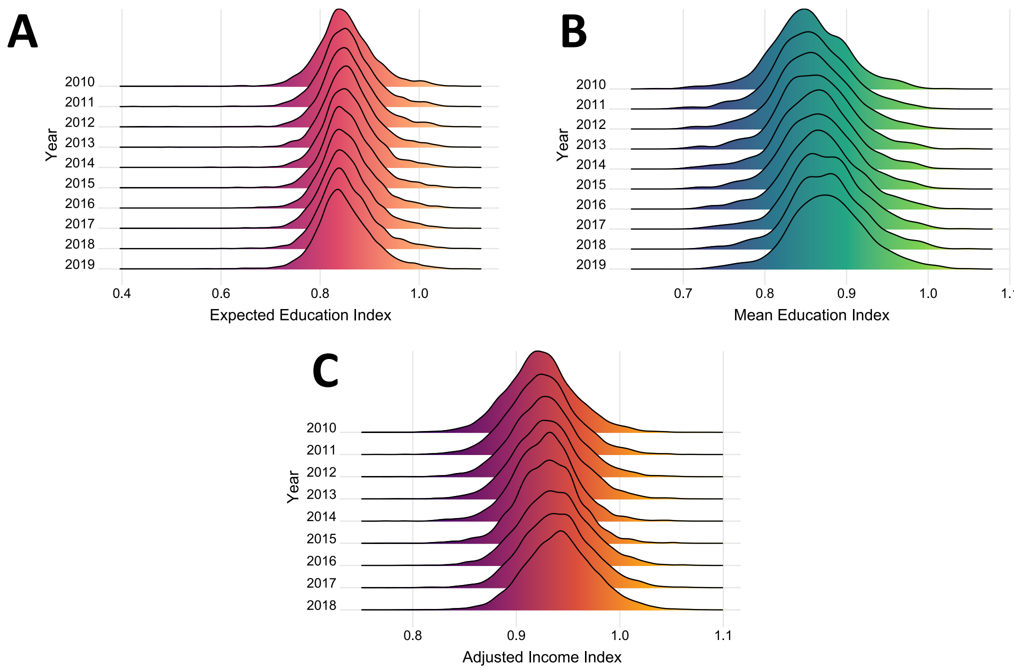


Figure 4.9: Distributions for (A) Expected Education Index, (B) Mean Education Index, and (C) Income Index adjusted for real incomes, for counties in the United States. Note that for all three quantities, there were no significant changes in distribution structure. All three distributions have long tails. This reflects the diversity of community types in the dataset, which ranges from college towns that overperform on the two education metrics, resort cities, which overperform on the income index, and communities which contain large prisons, which have incarcerated populations with concentrated disadvantage.

Table 4.1: Table of Density Distribution best fits to 1980 County Life expectancy data in the US. Distributions are sorted by goodness-of-fit based on Log-likelihood. Other goodness-of-fit criteria include Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (AD). The 3-parameter Weibull distribution is the best fit among density distributions.

Distribution	Alpha	Beta	Gamma	Mu	Sigma	Lambda	Log-likelihood	AICc	BIC	AD
Weibull_3P	16.6375	10.4298	57.9012				-6209.97	12426	12444.1	2.38509
Weibull_2P	74.6025	47.1671					-6243.25	12490.5	12502.6	8.15692
Gumbel_2P				74.6198	1.58217		-6260.71	12525.4	12537.5	10.6786
Loglogistic_2P	73.8564	74.0114					-6264.32	12532.6	12544.7	8.67578
Loglogistic_3P	73.8564	74.0114	0				-6264.32	12534.6	12552.8	8.67578
Normal_2P				73.7753	1.79198		-6291.1	12586.2	12598.3	12.0053
Lognormal_2P				4.30072	0.024538		-6322.14	12648.3	12660.4	15.4179
Lognormal_3P			0	4.30072	0.024538		-6322.14	12650.3	12668.4	15.4179
Exponential_2P			61.247			0.079819	-11084.9	22173.9	22186	1071.76
Exponential_1P						0.013555	-16655.8	33313.6	33319.7	1374.84

Table 4.2: Table of Density Distribution best fits to 2010 Income Index in the US. Distributions are sorted by goodness-of-fit based on Log-likelihood. Other goodness-of-fit criteria include Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (AD). The 2-parameter Log-logistic distribution is the best fit among density distributions.

Distribution	Alpha	Beta	Gamma	Mu	Sigma	Lambda	Log-likelihood	AICc	BIC	AD
Loglogistic_2P	0.923255	49.1207					15414.8	-30825.6	-30811.7	1.14091
Loglogistic_3P	0.710785	37.8154	0.212395				15415.3	-30824.5	-30803.6	0.97962
Lognormal_2P				-0.0797	0.036442		15376	-30748.1	-30734.2	6.9698
Lognormal_3P			0	-0.0797	0.036442		15376	-30746.1	-30725.2	6.9698
Normal_2P				0.924008	0.033697		15365.7	-30727.4	-30713.5	9.35583
Weibull_3P	0.173691	5.04981	0.763826				15153.5	-30301	-30280.1	46.8611
Weibull_2P	0.940428	26.443					14585	-29166.1	-29152.2	NaN
Gumbel_2P				0.941094	0.036154		14398.5	-28792.9	-28779	NaN
Exponential_2P			0.765196			6.29682	6547.31	-13090.6	-13076.7	2258.14
Exponential_1P						1.08224	-7178	14358	14365	3329.11

Table 4.3: Table of Density Distribution best fits to 2010 Expected Eudcation Index data in the US. Distributions are sorted by goodness-of-fit based on Log-likelihood. Other goodness-of-fit criteria include Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (AD). The 2-parameter Log-Logistic distribution is the best fit among density distributions.

Distribution	Alpha	Beta	Gamma	Mu	Sigma	Lambda	Log-likelihood	AICc	BIC	AD
Loglogistic_2P	0.849463	27.0004					4557.17	-9110.33	-9098.24	1.90437
Loglogistic_3P	0.740786	23.5464	0.10857				4557.39	-9108.77	-9090.63	1.65449
Lognormal_2P				-0.16231	0.068028		4485.38	-8966.76	-8954.67	10.0555
Lognormal_3P			0	-0.16231	0.068028		4485.38	-8964.76	-8946.62	10.0555
Normal_2P				0.852144	0.05794		4479.78	-8955.55	-8943.46	14.3338
Gamma_3P	0.009449	41.0831	0.463962				4448.47	-8890.93	-8872.79	14.9533
Weibull_3P	0.355248	5.83576	0.521133				4313.03	-8620.05	-8601.9	47.7756
Weibull_2P	0.879707	14.1077					4159.97	-8315.93	-8303.83	77.6761
Gumbel_2P				0.881991	0.064148		4015.52	-8027.03	-8014.93	103.264
Exponential_2P			0.528678			3.09151	403.22	-802.437	-790.34	993.437
Exponential_1P						1.17351	-2632.56	5267.12	5273.17	1259.43

Table 4.4: Table of Density Distribution best fits to 2010 Mean Education Index data in the US. Distributions are sorted by goodness-of-fit based on Log-likelihood. Other goodness-of-fit criteria include Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Anderson-Darling statistic (AD). The 2-parameter Log-Logistic distribution is the best fit among density distributions.

Distribution	Alpha	Beta	Gamma	Mu	Sigma	Lambda	Log-likelihood	AICc	BIC	AD
Loglogistic_2P	0.853502	32.39					1318.66	-2633.3	-2623.94	0.432838
Loglogistic_3P	0.853502	32.39	0				1318.66	-2631.29	-2617.24	0.432838
Normal_2P				0.854454	0.04759		1310.71	-2617.4	-2608.03	1.87619
Lognormal_2P				-0.15886	0.056025		1307.23	-2610.45	-2601.08	1.85783
Lognormal_3P			0	-0.15886	0.056025		1307.23	-2608.43	-2594.39	1.85783
Weibull_3P	0.224074	4.6894	0.64881				1296.93	-2587.82	-2573.78	4.82901
Gamma_3P	0.008418	35.1402	0.558629				1295.47	-2584.91	-2570.86	3.476
Weibull_2P	0.876973	18.1995					1256.07	-2508.12	-2498.75	12.7885
Gumbel_2P				0.878336	0.048892		1232.92	-2461.82	-2452.46	16.7536
Exponential_2P			0.667082			5.33697	543.774	-1083.53	-1074.16	210.086
Exponential_1P						1.17034	-679.222	1360.45	1365.14	331.061

CHAPTER 5

THE COMMUNITY HUMAN DEVELOPMENT INDEX AS A PRECISION PUBLIC HEALTH VULNERABILITY METRIC AND RISK INDICATOR

5.1 Introduction

Communities are facing a multitude of economic, social, and environmental challenges. Due to the COVID-19 pandemic, there have been unprecedented disruptions of interconnected systems [124, 125, 126]. This has occurred against a backdrop of growing economic inequality and increased natural disruptions due to climate change [127, 128, 129, 130] .The degree to which each of these challenges affects each community varies significantly, requiring a local approach to addressing multifaceted challenges. One aspect of taking a local approach to addressing these challenges is understanding how populations differ from community to community, and how this makes them more susceptible to or resilient to disruption from a range of hazards [131, 132, 133, 30, 31]. In the disaster risk management literature, this is often referred to as assessing population or social vulnerability [134]. Population vulnerability is inherently multidimensional, as it aims to characterize the many different ways a community could be impacted by an event [14, 135]. This concept has long been part of the fabric of risk management, but it has taken on a renewed focus since the late 2000's, after disasters such as Hurricane Katrina, where the negative impacts of the disaster were disproportionately felt in poorer communities [136, 137].

In the United States, one of the most common metrics for measuring population vulnerability is the US Centers for Disease Control's "Social Vulnerability Index" [138, 139]. First released in 2011, the SVI was designed to help researchers and officials, particularly public health officials, to identify vulnerable populations during major disruptions [139]. It is calculated using fifteen population variables, such as "Below Poverty" and "Multi-Unit structures", organized into four major topic areas, using data from the US Census Bureau's American Community Survey. Communities are assigned a percentile rank based on the raw value of the score for each variable, with 0 being the least vulnerable and 1 the most vulnerable. The rank scores are then summed and ranked again for each topic area, and the topic area scores are summed and ranked to create the final SVI score for a community [138, 140]. An updated version of the Social vulnerability Index is released every two years, at both the census tract scale and the county scale. The index is widely used in a number of different contexts to determine population vulnerability across the United States [141].

As with any index or model, there are gaps and challenges inherent in the design and calculation of the Social Vulnerability Index. First, because the SVI is designed to be comprehensive and used for almost any disaster, the variables it includes are not always relevant to assessing population vulnerability in a particular context [142, 143]. For example, lack of access to a vehicle, which is found in both relatively poor populations as well as relatively wealthy populations in urban centers, may be critical in some but not all disaster contexts.

Second, as the SVI is calculated using percentile ranks, it is by its very nature a comparative metric. This means that SVI scores for a particular county can change year to year purely based on how all other counties are performing, rather than any on-the-ground increases or decreases in the factors that increase vulnerability. Thus, improvements in the underlying indicators cannot be tracked from year to year- the SVI cannot provide a metric to measure longitudinal improvement or reduction in risk.

Finally, because it is calculated using multiple variables from the US Census Bureau, the SVI is a US specific metric that is difficult to apply in international contexts. Many countries do not measure the metrics used in the SVI, and when they are measured, they may not be calculated in the same way. This prevents any meaningful cross-country studies of vulnerability at the local level and limits the scope of use for the metric. In order to address these gaps, it is necessary to use additional metrics alongside the SVI in order to gain a more holistic understanding of population vulnerability at different geographic scales.

Any metrics that are used to measure vulnerability need a strong evidence-base to justify their relevance. One such metric is the Human Development Index, developed in the 1990's at the United Nations [4]. The HDI is based off the capabilities approach, a philosophical and economic framework developed by Amartya Sen and Martha Nussbaum which emphasized the importance of expanding people's freedoms so that they can have agency over decisions in their lives [1, 27, 43, 144]. A team at the United Nations operationalized this broader framework into a metric by creating the HDI, an index based on populations' health, education, and standard of living [46, 144, 26]. In the last 30 years, the HDI has become the gold standard for measuring quality of life and development, creating a race between countries to improve their scores over time [47, 26].

The HDI and the capabilities framework has a natural connection to population vulnerability - populations that have lower HDI scores have limited opportunities, freedom, and resources to make decisions, making them more vulnerable to disruptions from disasters [126, 145]. Additionally, the HDI addresses many of the gaps of other social vulnerability metrics, such as the SVI. The HDI uses a smaller set of variables, all of which have been demonstrated to be critical to human capabilities; it is designed to be measured longitudinally, allowing populations to observe growth over time independently of the performance of other populations; and it is designed to be used in international contexts, allowing for comparisons across countries [26].

Despite this, until recently, there was limited recognition of the overlap of the two concepts, with population vulnerability being used largely in disaster risk management studies and the capabilities approach used in economics and development. This is largely due to the fact that, due to data limitations, the majority of studies calculated HDI at the national scale, generating a single value for each country each year. This constrained its use in disaster risk management, which is often done at smaller geographic scales, such as counties or metropolitan areas (cities).

We have recently leveraged novel datasets to create a scalable, more localized implementation of the HDI, that can be applied to census tracts, counties, cities, and states in the US [21, 22]. Here, we show that the HDI and SVI correlate well with one another across US census tracts, while also capturing different information about the populations that reside there. We also demonstrate the importance of the HDI in public health population vulnerability, showing that it performs as well or better than the SVI in predicting health indicators and outcomes in US Census Tracts. Through this work, we show that HDI can serve as an effective complement to the SVI in measuring the population vulnerability of communities, serving as a metric that can be used by urban scientists and planners to reduce population vulnerability and risk in the future.

5.2 Results

5.2.1 Comparison of SVI and CHDI

We begin with a overview of the SVI and HDI at the census tract level in the United States.

US HDI at the census tract level had a mean of .905, a range from 0.443 to 1.113, and is normally distributed. SVI scores are allocated based on percentile rank, and therefore have a uniform probability distribution with a range from 0 to 1. There are four tracts with an SVI of 1.0000, including a tract in Reading, PA, a city that has experienced challenges due to the loss of heavy industries, and a tract in Belle Glade, FL, a city in the interior of the state which has a largely agriculture based economy and a relatively high poverty rate [146, 147]. Conversely, the three of the four tracts with a score of 0.0000 were located in suburban communities, including Mt. Pleasant, SC and Ypsilanti, MI.

HDI correlates relatively well with SVI at the census tract level, see Figure 5.1B. A simple

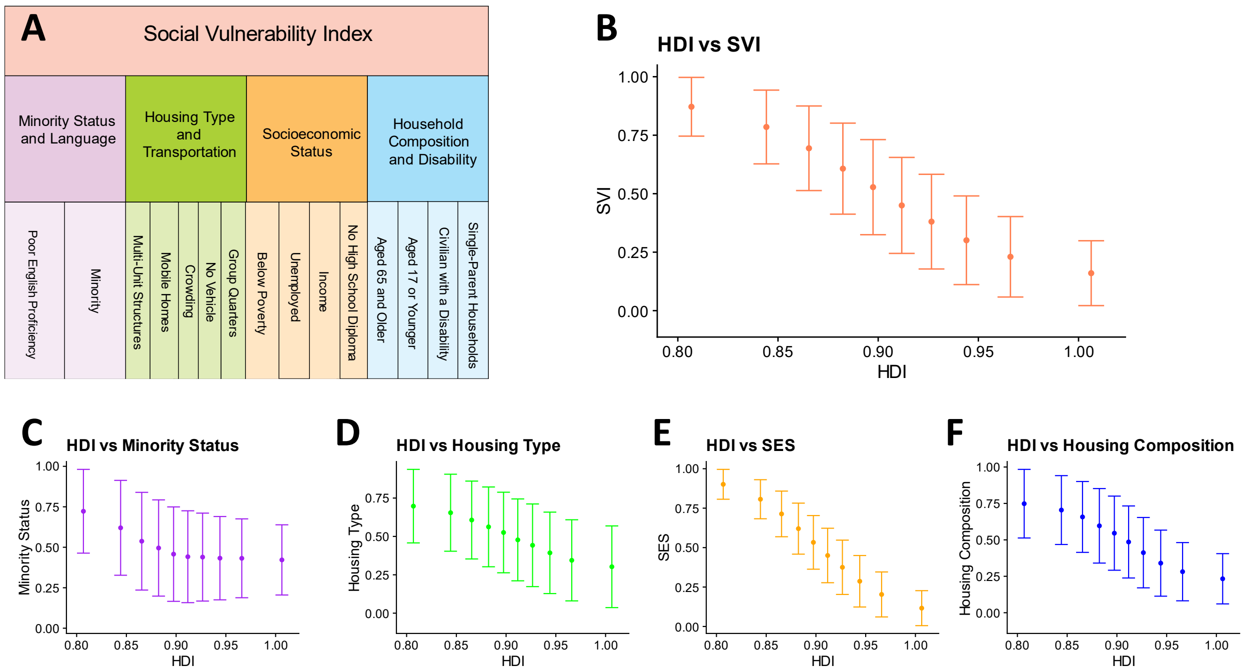


Figure 5.1: Comparison of HDI with SVI and SVI components.(A) A visual breakdown of the CDC's SVI and its components, adapted from the SVI website. (B) SVI value by HDI value. While the relationship between the two can be described by a linear fit, a sigmoid curve is slightly better at capturing the change between SVI and HDI. HDI correlates poorly with some of the components of the SVI, such as (C) Minority Status and (D) Housing Type, and has relatively good fits with the others (E) Socioeconomic Status and (F) Housing Composition.

linear fit of the data, $SVI = -3.929(HDI) + 4.059$ provides a fairly accurate description of the trend ($R^2 = 0.6048$). A logistic curve fit provides a slightly more close fit to the trend, with an $R^2 = 0.6182$, see Table 5.2. This indicates that at low development levels, increases in development lead to relatively large decreases in SVI, with the rate of reduction slowing down at higher development levels. However, there is large degree of variation throughout, with an average standard deviation of .181 across all bins. The greatest SVI standard deviation, 0.22, is found in the highest HDI bracket, tracts with HDI values from 0.98 to 1.11.

The outliers from the trend lines are illustrative of the differences between the two indicies.

For example, a tract in Temple City, CA with a high HDI of 0.9961, a higher value than the HDI of Norway, also had the high Social Vulnerability Index score of .8023, primarily because of high vulnerability scores in the "Minority" and "Housing Composition" subtopics. In Figure 5.2, we use HDI to predict SVI, to see how well a measure of development predicts a measure of vulnerability, and map the residuals to see where they diverge the most. Overall, communities in major cities with relatively high education, income and life expectancy perform better on the HDI than the SVI, while more suburban communities with the same characteristics tend to perform slightly better on the SVI than the HDI, see Figure 5.2. However, the counties where the indices diverge the most in their assessment of vulnerability are in certain rural counties, with SVI ranking certain counties in the South as more vulnerable than HDI, see Figure 5.9.

A breakdown of the SVI topic areas against HDI further reveals the similarities and differences between the two metrics, see Tables 5.1 and 5.2 and Figures 5.4 to 5.8. Out of the four topic areas, Socioeconomic Status (SES) best correlated with HDI, with an $R^2 = .7443$ for the Logistic fit. This is unsurprising, given that SES directly includes two of the three components of HDI- income and education. Results for other topic areas are more mixed. Housing Composition and Disability Status, a topic area that focuses on the age distributions in households, correlates somewhat poorly overall ($R^2 = 0.3456$), though with a narrowing standard deviation at higher HDI. Housing Type and Transportation, a metric that focuses on density and access to motor vehicles, fares even more poorly, with an $R^2 = 0.1870$. A number of high-HDI tracts are located in dense cities, while the Housing Type metric prioritizes single family homes.

The greatest divergence is between HDI and Minority status. Minority status is made up of 2 components: Percentage minority, and percent population aged 5 or older who speak English "Less than well." Initially, minority status decreases with increasing HDI. However, for tracts with an HDI value of .89 of greater, the average Minority Status value remains at

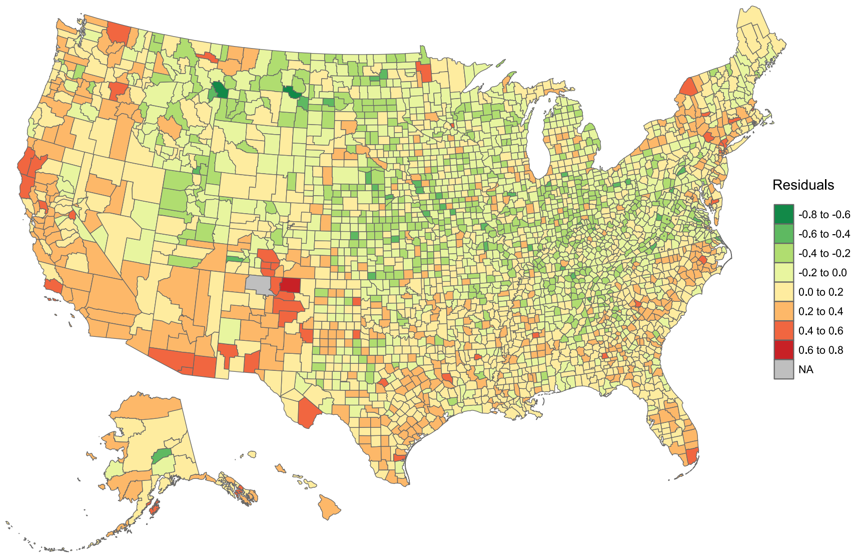


Figure 5.2: Map of Residuals of Logistic Fit to US County Human Development Index and Social Vulnerability Index Values. Areas in red represent counties where SVI is higher than that predicted by the fit, and green counties represent areas where SVI is lower than that predicted by the fit to HDI. On average, SVI reports higher vulnerability than HDI in urban areas. The CDC did not report a SVI value for Rio Arriba County, New Mexico.

0.45. Our previous work also showed a similar finding, where population fraction of minority poorly correlated with HDI across the US [22]. This reflects the wide range of communities in the US, which include largely minority communities that have high education, income, and life expectancy values. A tract in Newton Centre, in Newton, MA had an HDI value of 1.098, putting it in the 99th percentile of HDI values in the United States, and yet received a relatively high vulnerability in score in Minority Status of 0.6213.

5.2.2 CHDI for Assessing Long Term Health Outcomes and Vulnerability

Vulnerability to negative long term outcomes remains a distinctive process separate from that of disaster vulnerability. Using the CDC PLACES and USALEEP Dataset, which utilizes

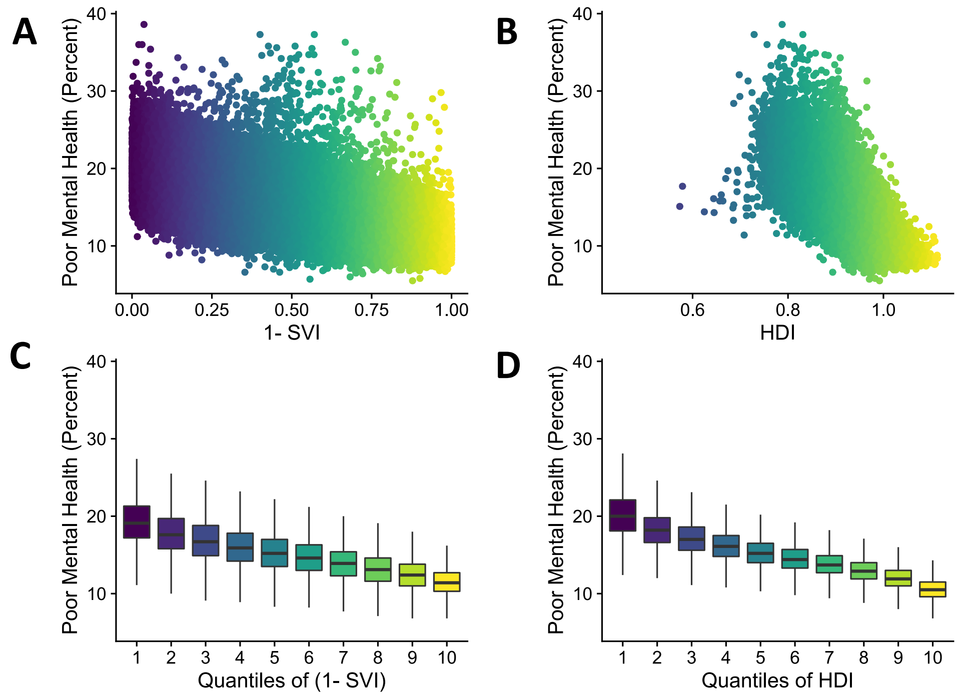


Figure 5.3: Comparison of Census Tract Level (A) 1- SVI and (B) HDI with Percentage of adults reporting 14 or more Poor Mental health Days Per Month. HDI has a closer linear fit to the outcome data, and has decreasing variance at higher values. This is visible in the comparison of the quantile averages of (C) 1- SVI and (D) HDI. Quantiles with the higher HDI values have the lower standard deviation.

survey data to estimate chronic disease rates at the population level, we compare SVI values to disease rates at the census tract level. While SVI roughly correlates with decreasing disease rates, there is a high degree of variation throughout. Based on linear correlation rates, HDI performs as well or outperforms SVI in a number of local health indicators, see Tables 5.3 to 5.8.

In particular, HDI is significantly better at predicting populations' self-rated mental health, with a linear fit R^2 of 0.6199, compared to .4426 for SVI, see Figure 5.3. We previously found that a logistic fit to the averages of HDI bins provides an even closer fit, with an R^2 of 0.949 [22]. HDI also negatively correlates with health risk behaviors, such as obesity, lack of leisure time physical activity and insufficient sleep. Lower community level development was also associated with the percentage of people who reported being in poor physical health. Overall, these results are consistent with previous studies, done on larger spatial scales, that found HDI to be an important metric to analyze factors related to long-term health behaviors and environments, such as quality of nutrition, morbidity from tuberculosis, and, to a lesser degree, vaccination coverage [148, 149, 11].

Health outcomes' variation also decreases as HDI increases. For example, the standard deviation for percent of people considered clinically obese drops from 6.11 for the lowest HDI bin to 4.20 for the highest HDI bins. This means the highest HDI tracts not only have the lowest average rate of obesity, but also the least variation in those rates. SVI shows a similar pattern, but with higher rates of variation at every point, decreasing in standard deviation from 6.67 to 4.52 as Vulnerability scores decrease.

Cancer incidence rates at the tract level was poorly correlated with both HDI and SVI (R^2 values of 0.0729 and 0.0878, respectively) . This differs from previous studies done across countries, which found as HDI and life expectancy increased, the incidence of cancer also increased. This is likely due to the fact that the large gaps in life expectancy in the international scale, combined with varying rates of infectious disease, lead to major differences in the distribution of cancers across development levels [13, 12, 150, 151]. As most census tracts in the US are in the "High" or "Very High" development categories created by the United Nations, the demographic shifts at the root of global cancer differences have already largely occurred. Other studies have found that higher SVI correlates with higher mortality after cancer diagnosis, as well as increased exposure to environmental contaminants linked to cancer, but its connection to cancer incidence rates at the tract level remains relatively unexplored in the literature [152, 153].

5.3 Discussion

Assessing population vulnerability remains a critical challenge for researchers, policymakers, and local governments attempting to mitigate disruptions and build resilient communities over the long-term [132, 154]. Fundamental to addressing this challenge is understanding how populations vary in risk to both disturbances that occur over relatively short time frames, such as natural disasters, and more long-term issues, such as growing rates of chronic disease. This requires building a common knowledge base that facilitates communication and coordination across geographic scales and across countries.

In the United States, the CDC's Social Vulnerability Index, designed for disaster risk management, has been widely used as a population vulnerability metric at both the census tract and county levels. While the value of the metric has been demonstrated in a number of different contexts, it is limited in some aspects due its structure. The SVIs equal weighting to fifteen different variables, some of which may not be relevant in all contexts. It also not designed to be measured longitudinally, preventing communities from measuring growth or change in vulnerability over time. Finally, it is designed to be calculated using a specific set of input variables from the US Census, which prevents it from being used widely in international contexts.

In this work we demonstrate that a subnational, scalable implementation of the Human Development Index can be used as a complementary tool that can address some of these gaps, and create broader discussions around how to address vulnerability across diverse communities in the future. We first show that the community HDI correlates well with the SVI at the census tract level, reflecting a convergence in the predicted vulnerability of neighborhoods. We also show the similarities and differences between the two indices, finding that the HDI overlaps strongly with the SES component of the SVI, partly with the Housing Type, and Housing Composition Components, and minimally with the Minority Status component. We find that HDI predicts higher vulnerability than SVI for rural communities, and that SVI predicts greater vulnerability for urban centers.

Previous studies have found that the HDI is an important metric for assessing the quality of environments for promoting good health [11, 149]. To assess the HDI's utility in predicting chronic disease and long-term health vulnerability , we performed a systematic, comparative analysis of HDI and SVI to health indicators and outcomes from the CDC PLACES dataset. We found that for nearly all indicators, HDI performed as well or better than SVI as a predictive metric, and had lower variation. For certain health risk factors, such as poor mental health and obesity, HDI significantly outperformed the SVI. However, both the HDI and SVI were relatively poor predictors of cancer incidence rates.

Currently, our study is limited to vulnerability analysis in public health and chronic disease metrics. Future studies should examine the utility of the HDI in different risk contexts, including natural disasters such as hurricanes, floods, and wildfires. Longitudinal data from the COVID-19 pandemic should also be sued to compare how well each of the indices performed at different phases of the pandemic.

Our analyses reveal that HDI can function as a population and public health vulnerability metric across spatial scales, serving as a complement to the SVI in the US. This opens the possibilities for many new directions of research, and allows for greater dialogue between the field of human development, based in economics, ethics, and international studies, and the field of risk management. Critically, it also allows for a common base metric that communities across different countries can use to measure vulnerability over time, facilitating discussions around best-practices to reduce risk over time.

5.4 Methods

5.4.1 Data Sources

Social Vulnerability Index data at the census tract and county levels was downloaded from the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry website for the most recent year it was calculated, 2018 [141].

Community Human Development Index Values at the census tract and county scales were calculated using a combination of data sources. Expected and mean education was calculated using school enrollment rates and educational attainment, with data was taken from the US Census Bureau's American Community Survey (ACS) estimates for 2010-2015 [155]. Life expectancy data was taken from the National Center for Health Statistics, the Robert Wood Johnson Foundation, and the National Association for Public Health Statistics and Information Systems' US Small-area Life Expectancy Estimates Project, as well as the Institute for Health Metrics and Evaluation's US Health Map [156, 157]. Income was calculated by down-scaling the 2015 US Gross National Income from the UN Statistics Division using personal income estimates from the American Community Survey estimates for 2010-2015. The down-scaled Gross National Income was then adjusted to reflect real income using the US Bureau of Economic Analysis' regional purchasing power parities [35].

Health outcomes data was taken from the US Centers for Disease Control PLACES Project 2021 release. Detailed descriptions of the variables and measure definitions can be found on the PLACES website and documentation [81].

5.4.2 Calculating CHDI and SVI

The Human Development Index is calculated as the geometric mean of three separate subindices- the Education Index (I_E), the Real Income Index (I_{RI}), and Life Expectancy Index(I_{LE}). Calculating the HDI as a geometric mean, $HDI = (I_{\mathrm{LE}} \times I_{\mathrm{E}} \times I_{\mathrm{RI}})^{1/3}$ prevents any one component from dominating the index- to have a high HDI score, a population has to have high values in all three components. The United Nations Development Programme has determined benchmarks for each focus areas, based on global distributions of these values. The values for each geographic area under study are indexed using these benchmarks, producing a score from 0 to 1 for each component, as well as the HDI overall.

The Life Expectancy Index is calculated by indexing life expectancy at birth LE between 20 years and 85 years, $I_{LE} = (LE - 20)/(85 - 20)$.. The Education index was designed to account for both current educational attainment levels (for the population 25 years and older), and predicted educational attainment for the population based on school enrollment rates. These two sub-indices are the Mean Years of Schooling Index ($MYSI$) and the Expected Years of Schooling Index ($EYSI$), respectively. The Mean Years of Schooling Index, which applies to the adult population 25 years and older, has a benchmark value of 15 years, and the Expected Years of Schooling Index has a benchmark of 18 years. The Education Index is then calculated as the mean of these two subindices, $I_{\mathrm{E}} = 1/2(EYSI + MYSI)$.

The Real Income Index is calculated by the United Nations using Gross National Income per capita (gni). To create a subnational equivalent, we calculated the ratio of Gross National Income to total personal income per census tract, and adjusted using local purchasing power parities (PPP) to calculate real incomes, $\hat{\mathrm{gni}}_{PPP} = c_{\mathrm{GNI/I}}\ I_{\mathrm{pc}}/PPP$. The United Nations takes the log of inputs, as income has diminishing returns on development as it increases. The benchmarks for the Income Index are \$ 100 to \$ 75,000, giving the equation $I_{RI} = (\log \hat{gni}_{PPP} - \log 100)/(\log(75,000) - \log(100))$.

Further details on calculating the subnational HDI can be found in our previous work [21, 22].

The SVI is calculated every two years based on data from the American community Survey.It is made up of 15 variables organized into 4 themes, see Figure 5.1A. For each tract, each variable is given a score x, from 0 to 1, based on the percentile rank of the raw value of that variable, u, in comparison to all of the other tracts in the United States, $x = \eta_{.u}$. 0 indicates the least vulnerability, and 1 indicates the most. Each theme area, y is then given a score based on the percentile rank of the sum of the scores of its variables $y = \eta_{\sum x}$. Finally, the SVI is calculated by taking the percentile rank of the sum of the scores of the four theme areas $SVI = \eta_{\sum y}$.

5.5 Supplementary Materials

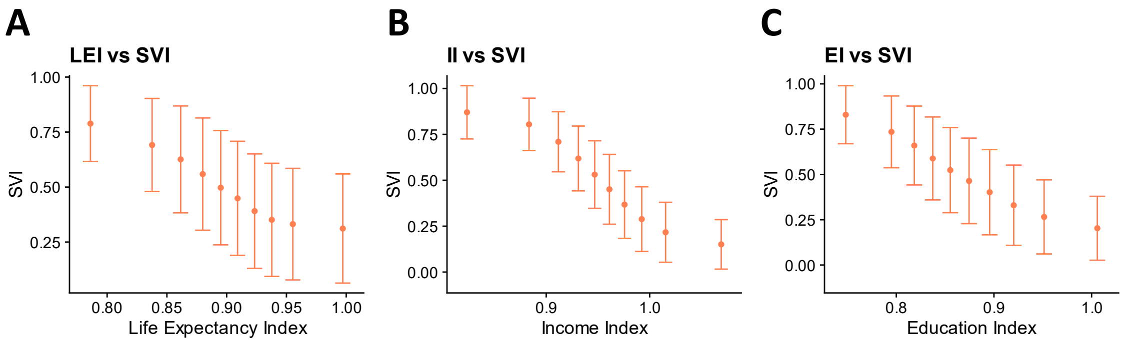


Figure 5.4: Comparison of SVI with CHDI components.

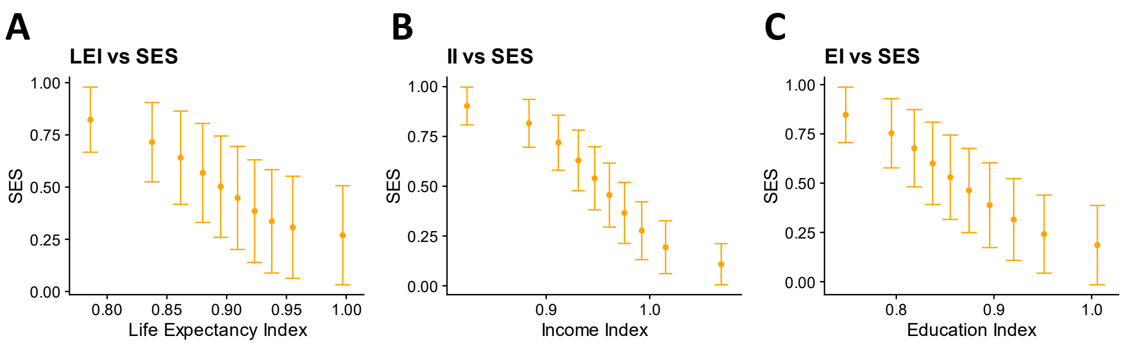


Figure 5.5: Comparison of Socioeconomic Status with CHDI components.

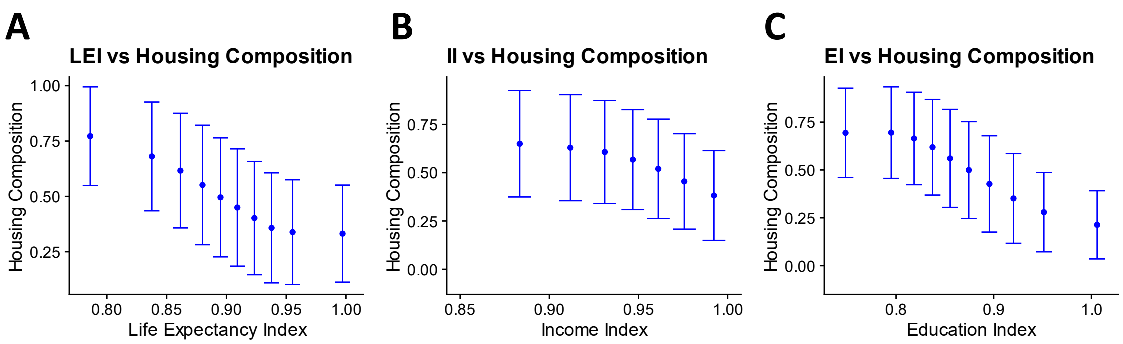


Figure 5.6: Comparison of Housing Composition with CHDI components.

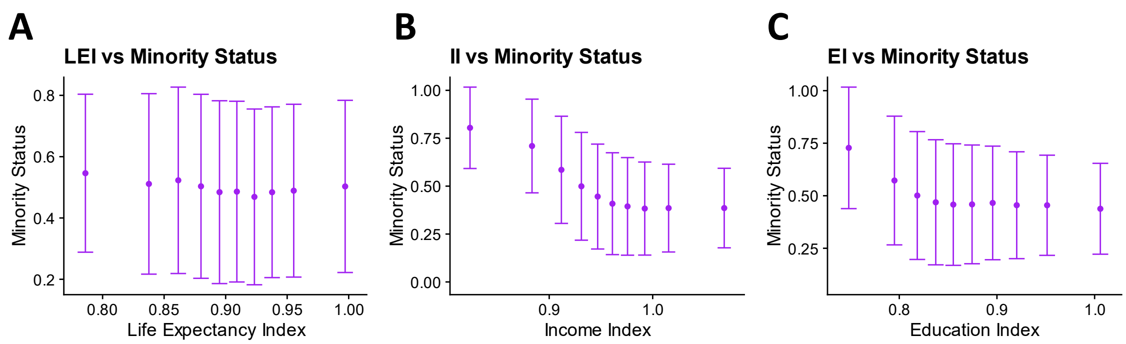


Figure 5.7: Comparison of Minority Topic Area with CHDI components.

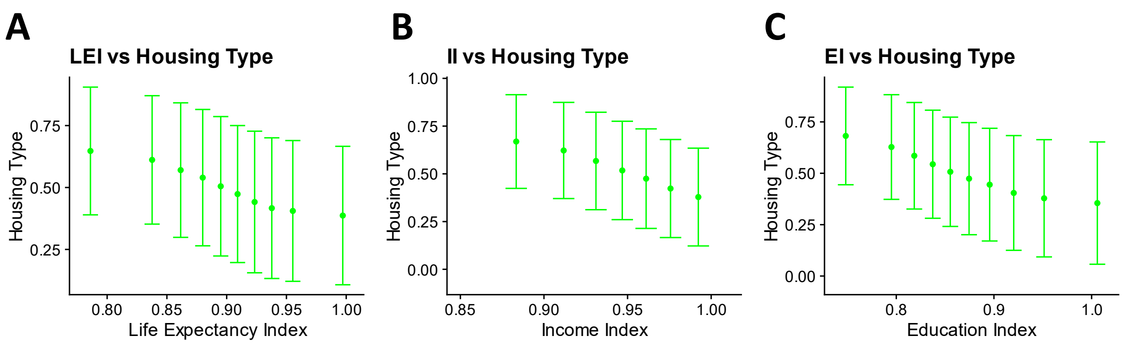


Figure 5.8: Comparison of Housing Type with CHDI components.

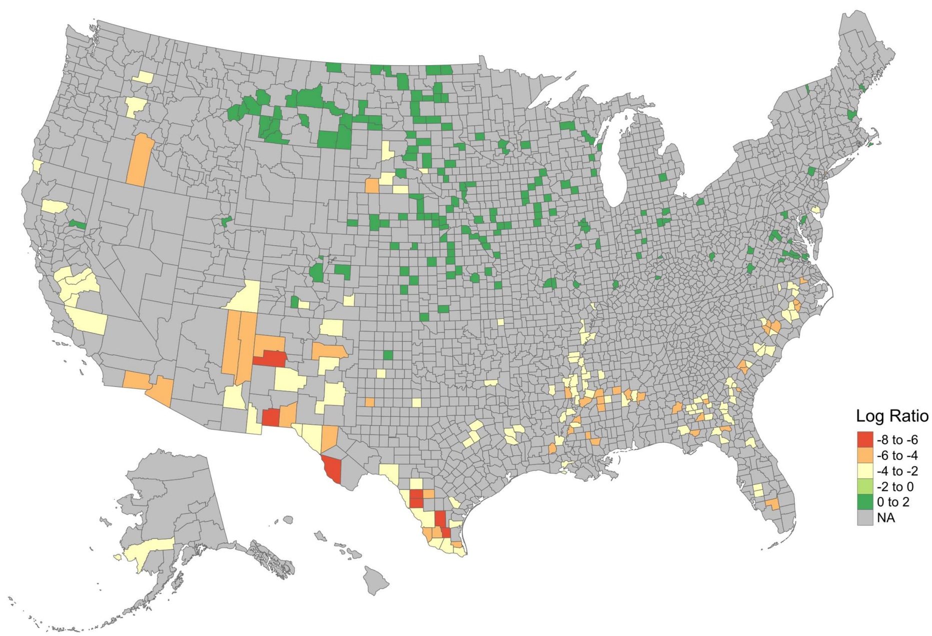


Figure 5.9: Map of counties with top 5 % and bottom 5% of values for $(ln((1-SVI)/HDI))$, representing the counties where the two metrics diverge most from one another. Green counties represent areas where $1 - SVI$ is higher than HDI, indicating that SVI predicts them to be more resilient. Yellow, orange and red counties represent the converse. Overall, the metric diverge most from one another in the Midwest and South.

Table 5.1: Linear Fits for CHDI, SVI and SVI Topic Areas

SVI and Topic Areas	Linear Model $y = a + b \times hdi$
SVI	a = 4.0595 (0.0112) b = -3.929 (0.0123) R^2 = 0.6048
Housing Composition	a = 3.1027 (0.0143) b = -2.8859 (0.0157) R^2 = 0.3365
Housing Type	a = 2.4512 (0.0159) b = -2.1575 (0.0175) R^2 = 0.185
Minority Status	a = 1.9869 (0.0164) b = -1.6184 (0.0180) R^2 = 0.1078
Socioeconomic Status	a = 4.3926 (0.0093) b = -4.3001 (0.0102) R^2 = 0.7262

Table 5.2: Logistic Fits for HDI, SVI and SVI Topic Areas. K is the asymptote on the y-axis, μ is the location parameter (the midpoint of the curve on the x-axis), and s is the scaling parameter. A Logistic fit could not be calculated for Minority Status, as the parameters did not converge.

SVI and Topic Areas	Logistic Model $y = \frac{K}{1+e^{\frac{(\mu - hdi)}{s}}}$
SVI	K = 1.0344 (0.0053)
	μ = 0.9003 (0.0006)
	s = -0.0520 (0.0004)
	R^2 = 0.6182
Housing Composition	K = 0.8821 (0.0068)
	μ = 0.9240 (0.0011)
	s = -0.0607 (0.0008)
	R^2 = 0.3456
Housing Type	K = 0.9387 (0.0183)
	μ = 0.9190 (0.0041)
	s = -0.0984 (0.0025)
	R^2 = 0.1870
Socioeconomic Status	K = 1.0270 (0.0035)
	μ = 0.9016 (0.0004)
	s = -0.0454 (0.0002)
	R^2 = 0.7443

Table 5.3: Part 1: Linear Fits for CHDI and Health Indicators

Health Indicator	Linear Model $y = a + b \times hdi$
Adults Lacking Health Insurance	a = 107.582 (0.3672) b = -100.8855 (0.4051) R^2 = 0.4699
Adults with Arthritis	a = 47.7353 (0.3685) b = -25.0918 (0.4065) R^2 = 0.0516
Binge Drinking in Adults	a = -1.6738 (0.1917) b = 21.2918 (0.2114) R^2 = 0.1266
Adults with High Blood Pressure	a = 88.2952 (0.3858) b = -61.6402 (0.4255) R^2 = 0.2307
Adults Taking Blood Pressure Medication	a = 75.1087 (0.4091) b = -2.432 (0.4513) R^2 = 4e-04
Adults with Cancer (excluding skin cancer)	a = -1.4031 (0.1101) b = 9.0029 (0.1214) R^2 = 0.0729
Adults with Asthma	a = 25.1391 (0.0737) b = -16.989 (0.0813) R^2 = 0.3843
Cervical Cancer Screening Rate	a = 46.2409 (0.2019) b = 41.5233 (0.2226) R^2 = 0.3321
Adults with Coronary Heart Disease	a = 20.9691 (0.1089) b = -16.4099 (0.1201) R^2 = 0.2106
Adults who had Routine Checkup Last Year	a = 66.3644 (0.2937) b = 10.2903 (0.3240) R^2 = 0.0142

Table 5.4: Part 2: Linear Fits for CHDI and Health Indicators

Health Indicator	Linear Model $y = a + b \times hdi$
Cholesterol Screening Rate	a = 50.2789 (0.2237) b = 39.056 (0.2467) R^2 = 0.2638
Older Adult Colon Screening Rate	a = -19.3422 (0.2811) b = 91.8678 (0.3101) R^2 = 0.5564
Adults with COPD	a = 33.2287 (0.1294) b = -28.7722 (0.1427) R^2 = 0.3674
Preventative Health Screening (Men)	a = -37.7671 (0.3230) b = 76.3467 (0.3563) R^2 = 0.3962
Preventative Health Screening (Women)	a = -31.0779 (0.2911) b = 65.2408 (0.3211) R^2 = 0.3711
Visit to Dentist	a = -89.1809 (0.3292) b = 168.5216 (0.3631) R^2 = 0.7548
Adults with Depression	a = 39.561 (0.2074) b = -20.9236 (0.2288) R^2 = 0.1068
Adults with Diabetes	a = 50.1744 (0.1671) b = -43.3196 (0.1843) R^2 = 0.4412
Self-Rated Health Status	a = 119.3989 (0.2405) b = -109.8575 (0.2653) R^2 = 0.7103
Adults with High Cholesterol	a = 36.7357 (0.2814) b = -5.3735 (0.3104) R^2 = 0.0043

Table 5.5: Part 3: Linear Fits for CHDI and Health Indicators

Health Indicator	Linear Model $y = a + b \times hdi$
Adults with Chronic Kidney Disease	a = 11.5088 (0.0421) b = -9.3752 (0.0465) R^2 = 0.3677
No Leisure Physical Activity	a = 134.2356 (0.2783) b = -117.6942 (0.3070) R^2 = 0.6775
Mammography Use	a = 64.3419 (0.2213) b = 13.9071 (0.2440) R^2 = 0.0444
Poor Mental Health	a = 59.6066 (0.1317) b = -49.0823 (0.1453) R^2 = 0.6199
Adults with Obesity	a = 109.7922 (0.2894) b = -85.2591 (0.3192) R^2 = 0.5049
Poor Physical Health	a = 61.5641 (0.1412) b = -53.1409 (0.1557) R^2 = 0.6247
Poor Sleep	a = 92.7478 (0.2334) b = -61.9095 (0.2574) R^2 = 0.4526
Adults with Stroke	a = 15.6322 (0.0619) b = -13.444 (0.0682) R^2 = 0.3569
Adults with Tooth Loss	a = 115.1719 (0.2542) b = -110.3305 (0.2803) R^2 = 0.6888

Table 5.6: Part 1: Linear Fits for SVI and Health Indicators

Health Indicator	Linear Model $y = a + b \times svi$
Adults Lacking Health Insurance	a = 6.0926 (0.0454) b = 20.2964 (0.0783) R^2 = 0.4898
Adults with Arthritis	a = 23.9949 (0.0474) b = 2.0612 (0.0819) R^2 = 0.009
Binge Drinking in Adults	a = 19.9244 (0.0237) b = -4.6395 (0.0410) R^2 = 0.1548
Adults with High Blood Pressure	a = 27.5007 (0.0509) b = 9.9847 (0.0878) R^2 = 0.1559
Adults Taking Blood Pressure Medication	a = 72.7986 (0.0515) b = 0.2177 (0.0889) R^2 = 1e-04
Adults with Cancer (excluding skin cancer)	a = 7.722 (0.0137) b = -1.9472 (0.0237) R^2 = 0.0878
Adults with Asthma	a = 8.3684 (0.0101) b = 2.7815 (0.0175) R^2 = 0.2653
Cervical Cancer Screening Rate	a = 88.31 (0.0242) b = -8.9449 (0.0417) R^2 = 0.3968
Adults with Coronary Heart Disease	a = 4.6822 (0.0141) b = 2.8614 (0.0243) R^2 = 0.1649
Adults who had Routine Checkup Last Year	a = 76.8831 (0.0369) b = -2.4021 (0.0637) R^2 = 0.0199

Table 5.7: Part 2: Linear Fits for SVI and Health Indicators

Health Indicator	Linear Model $y = a + b \times svi$
Cholesterol Screening Rate	a = 89.1625 (0.0290) b = -7.0494 (0.0500) R^2 = 0.2213
Older Adult Colon Screening Rate	a = 72.5509 (0.0370) b = -17.4387 (0.0638) R^2 = 0.5163
Adults with COPD	a = 4.9221 (0.0179) b = 4.52 (0.0310) R^2 = 0.2335
Preventative Health Screening (Men)	a = 39.5268 (0.0382) b = -16.3346 (0.0660) R^2 = 0.4671
Preventative Health Screening (Women)	a = 34.8632 (0.0351) b = -13.7413 (0.0606) R^2 = 0.424
Visit to Dentist	a = 79.2009 (0.0470) b = -31.619 (0.0812) R^2 = 0.6843
Adults with Depression	a = 19.8274 (0.0274) b = 1.5934 (0.0473) R^2 = 0.0159
Adults with Diabetes	a = 6.8381 (0.0216) b = 8.2327 (0.0373) R^2 = 0.4103
Self-Rated Health Status	a = 9.6376 (0.0336) b = 20.6019 (0.0580) R^2 = 0.6433
Adults with High Cholesterol	a = 31.8881 (0.0355) b = -0.0288 (0.0613) R^2 = 0

Table 5.8: Part 3: Linear Fits for SVI and Health Indicators

Health Indicator	Linear Model $y = a + b \times svi$
Adults with Chronic Kidney Disease	$a = 2.0865$ (0.0053)
	$b = 1.8681$ (0.0091)
	$R^2 = 0.376$
No Leisure Physical Activity	$a = 17.1656$ (0.0411)
	$b = 21.0352$ (0.0709)
	$R^2 = 0.5573$
Mammography Use	$a = 77.6488$ (0.0283)
	$b = -1.4384$ (0.0489)
	$R^2 = 0.0122$
Poor Mental Health	$a = 11.0864$ (0.0201)
	$b = 8.1721$ (0.0347)
	$R^2 = 0.4426$
Adults with Obesity	$a = 26.2965$ (0.0438)
	$b = 12.6301$ (0.0756)
	$R^2 = 0.2853$
Poor Physical Health	$a = 8.7593$ (0.0205)
	$b = 9.3898$ (0.0353)
	$R^2 = 0.5022$
Poor Sleep	$a = 31.2872$ (0.0319)
	$b = 10.825$ (0.0550)
	$R^2 = 0.3563$
Adults with Stroke	$a = 2.2162$ (0.0080)
	$b = 2.489$ (0.0139)
	$R^2 = 0.315$
Adults with Tooth Loss	$a = 5.3574$ (0.0374)
	$b = 19.8565$ (0.0646)
	$R^2 = 0.5746$

CHAPTER 6

CONCLUSION

In this thesis, we created a novel precision public health framework and metric to measure community human development across geographic scales and over time, allowing us to examine the connections between human development, neighborhood effects, and health risk management. In Chapters Two and Three, we first created a novel Community Human Development Index (CHDI) that could be used across geographic scales for multilevel analysis over time. Our analysis of the spatial patterns of CHDI in the US at the census tract level found that communities surrounding colleges concentrate individuals with high education, income, and life expectancy, resulting in very high CHDI values. Policy makers and public health adminstrators can use these towns as examples of positive community human development outcomes to implement best practices for areas with low CHDI values and offer greater support to vulnerable and risk-prone communities around the world, especially in the healthcare arena. Importantly, we also created a bridge to study the relationship between community human development and neighborhood effects, and found that every increase in CHDI at the local scale was associated with systematic decreases in negative outcomes in population health, social and economic realms.

In Chapter Four, we examined longitudinal health development trends at the county level in the United States. While average life expectancy increased overall in the US over this period, this masked a rapid divergence in life expectancy outcomes at the county level, with some counties actually declining in life expectancy over time. This is a trend that has continued over the last eight years, and has only been accelerated by the COVID-19 pandemic. It is a major topic of interest- the National Academies recently released a report that examined some of the root causes of these increasing mortality rates, including the opioid crisis, increasing deaths from suicide, and rising rates of obesity. While this is certainly grim news, our study, which reviewed trends, outcomes and interventions where pertinent, also helped identify some ways we may be able address this moving forward. In particular, the large life expectancy gains rural Alaskan counties made shows the importance of taking a tailored, local approach to health challenges. Focusing on community needs has proven impact on reducing these life expectancy disparities and promoting better health overall.

In Chapter Five, we demonstrated that CHDI can effectively be used as a health vulnerability metric and risk indicator. We compared it against the current standard of the field of risk management, the Social Vulnerability Index (SVI), and found that the SVI has limitations which can be overcome when used in conjunction with the CHDI. CHDI correlates well with the SVI overall, but also differs in important ways, capturing different information about vulnerability at the community level. We also found that CHDI successfully predicted health outcomes as well or better than the SVI at census tract level. Overall, these two indices are most powerful when used in conjunction with one another, with SVI being more applicable for short-term disaster relief, and CHDI more applicable for long term, sustained efforts to reduce vulnerability.

6.0.1 Future Directions

The CHDI is a framework and metric for systematically characterizing development across spatial scales and over time, and can be used as a metric for health risk and vulnerability analyses at the community level. Progress in understanding development, vulnerability, and resilience at the neighborhood level will depend on continuing to harmonize datasets from a diverse set of sources, in order to capture the multifaceted nature of these processes. Our work examined US neighborhoods in detail, and as more global granular datasets become available, future studies should examine development in communities in a number of nations. In particular, a comparison of neighborhood-level development distributions in high-income countries versus low-income countries may provide some insight into how inequality varies across these categories and across geographies.

Future research should also examine local life expectancies in greater detail to determine the root causes of the large disparities observed at the census tract level. Comparing life expetancies at different ages at the neighborhood scale to relative rates of infectious disease, chronic disease, and access to care can further illuminate how each of these contribute to mortality across the life course. Age-stratified and location-stratified mortality risk can also be used to determine the relative impacts of features of the built environment on population health outcomes.

Our study revealed that CHDI and SVI perform similarly when used to predict vulnerability to negative long-term health outcomes. Future studies should examine and compare these metrics across a wider range of contexts. In particular, further work needs to be done to compare how well each of these metrics predict vulnerability in the face of natural disasters, such as hurricanes, floods, or wildfires. Using Disaster Relief Aid Application data from the US Federal Emergency Management Agency (FEMA), researchers can examine how well the CHDI or SVI of a community before a disaster predicts the degree of aid required after the disaster has struck. This can also be used to determine how relevant each SVI and CHDI input is to prognosticating risk.

As longitudinal data from the COVID-19 pandemic is standardized across the globe, studies should also examine how the impacts of the pandemic varied based on development levels. A comparison of the CHDI and the SVI in predicting COVID-19 cases and deaths, as well as economic resilience, would help reveal how vulnerability varied across different geographies and at different phases in the pandemic. Using the CHDI to study community COVID-19 vulnerability across countries would also reveal how well it performs as a disaster vulnerability metric in international contexts. This would be critical in building community resilience and saving lives in the future.

REFERENCES

[1] Amartya Sen. *Development as Freedom*. Alfred A. Knopf, New York, 2001.

[2] Amartya Sen. Capability and well-being. *The quality of life*, 30:270–293, 1993.

[3] Martha C. Nussbaum. Philosophy and Economics in the Capabilities Approach: An Essential Dialogue. *Journal of Human Development and Capabilities*, 16(1):1–14, January 2015.

[4] Elizabeth A Stanton. The Human Development Index: A History. *PERI Working Papers*, page 37, 2007.

[5] Pedro Conceição, Milorad Kovacevic, and Tanni Mukhopadhyay. Human Development: A Perspective on Metrics. In *Measuring Human Capital*, pages 83–115. Elsevier, 2021.

[6] Mahbub Ul Haq. *Reflections on Human Development*. Oxford University Press, 1995.

[7] Martha C. Nussbaum. *Creating capabilities: the human development approach*. Belknap Press of Harvard University Press, Cambridge, Mass, 2011.

[8] Jeroen Smits and Iñaki Permanyer. The Subnational Human Development Database. *Scientific Data*, 6(1):190038, March 2019.

[9] Iñaki Permanyer and Jeroen Smits. Inequality in Human Development across the Globe. *Population and Development Review*, 46(3):583–601, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/padr.12343.

[10] Alfonso J. Rodríguez-Morales and Diana M. Castañeda-Hernández. Relationships between morbidity and mortality from tuberculosis and the human development index (HDI) in Venezuela, 1998–2008. *International Journal of Infectious Diseases*, 16(9):e704–e705, September 2012.

[11] Grace Maria Antony, K. Visweswara Rao, and N. Balakrishna. Suitability of HDI for Assessing Health and Nutritional Status. *Economic and Political Weekly*, 36(31):2976–2979, 2001. Publisher: Economic and Political Weekly.

[12] M. Ghoncheh, M. Mirzaei, and H. Salehiniya. Incidence and mortality of breast cancer and their relationship with the human development index (HDI) in the world in 2012. *Asian Pacific Journal of Cancer Prevention*, 16(18):8439–8443, 2016. Number: 18.

[13] Freddie Bray, Ahmedin Jemal, Nathan Grey, Jacques Ferlay, and David Forman. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *The Lancet Oncology*, 13(8):790–801, August 2012.

[14] David W. Stewart. Uncertainty and Risk Are Multidimensional: Lessons from the COVID-19 Pandemic. *Journal of Public Policy & Marketing*, 40(1):97–98, January 2021. Publisher: SAGE Publications Inc.

[15] Jonathan Barlow and Irena Vodenska. Socio-Economic Impact of the Covid-19 Pandemic in the U.S. *Entropy*, 23(6):673, June 2021. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

[16] Seble Tadesse and Worku Muluye. The Impact of COVID-19 Pandemic on Education System in Developing Countries: A Review. *Open Journal of Social Sciences*, 8(10):159–170, September 2020. Number: 10 Publisher: Scientific Research Publishing.

[17] UNDP Strategic Plan 2022-2025 | United Nations Development Programme.

[18] Awareness-Based System Change as The Basis for Transforming Systems and Social Norms | United Nations Development Programme.

[19] Global Project for Managing Development Cooperation Effectively: Annual Status Report 2021 | United Nations Development Programme.

[20] Sean Fox and Allan Macleod. Localizing the SDGs in cities: reflections from an action research project in Bristol, UK. *Urban Geography*, 42:1–21, 2021. Publisher: Routledge.

[21] Suraj Sheth and Luís M. A. Bettencourt. The Community Human Development Index (CHDI): Localizing Sustainable Development Goals Across Scales. In *2020 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 1–7, April 2020.

[22] Suraj Sheth and Luís M. A. Bettencourt. Human Development in US Cities and Neighborhoods. SSRN Scholarly Paper 3961750, Social Science Research Network, Rochester, NY, November 2021.

[23] Christa Brelsford, José Lobo, Joe Hand, and Luís M. A. Bettencourt. Heterogeneity and scale of sustainable development in cities. *Proceedings of the National Academy of Sciences*, 114(34):8963–8968, August 2017.

[24] United Nations. The Sustainable Development Goals Report 2022. Technical report, United Nations, 2022.

[25] Diana Mitlin and David Satterthwaite. *Urban Poverty in the Global South: Scale and Nature*. Routledge, London, December 2012.

[26] Pedro Conceição, Milorad Kovacevic, and Tanni Mukhopadhyay. Chapter 4 - Human Development: A Perspective on Metrics. In Barbara Fraumeni, editor, *Measuring Human Capital*, pages 83–115. Academic Press, January 2021.

[27] Martha C. Nussbaum. *Women and Human Development: The Capabilities Approach*. Cambridge University Press, June 2001.

[28] The Human Development Index – what it is and what it is not | Human Development Reports.

[29] Marianne Bitler, Jonah Gelbach, and Hilary Hoynes. What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments, November 2003.

[30] Anuradha Krishnaswamy. The TripleRM sustainability model: Strategic risk, resilience and resource management of cities (Sustainable infrastructure planning and management of resilient cities). In *2015 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 117–124, July 2015.

[31] Suraj Neil Sheth. The TripleRM Global Health Management Model (GHMM): Strategic risk management of vector borne infectious diseases to build healthy, sustainable, adaptable and resilient communities: (Strategic global health security risk assessment, resilience planning and resource management in urban and rural environments). In *2017 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 1–7, November 2017.

[32] National Library of Medicine. "What is precision medicine?".

[33] Daraius Irani, M. SIERS, M. RICE, and E. BAST. Human Development Index Disparities in Baltimore City. *Regional Economic Studies Institute*, 2018.

[34] Emi Suzuki, Tariq Khokar, and Neil Fantom. What does "life expectancy at birth" really mean?, 2013.

[35] Regional Price Parities by State and Metro Area | U.S. Bureau of Economic Analysis (BEA).

[36] Asma Khalid. The Perfect State Index: If Iowa, N.H. Are Too White To Go First, Then Who? *NPR*, January 2016.

[37] Robert J. Sampson. *Great American city: Chicago and the enduring neighborhood effect*. Univ. of Chicago Press, Chicago, Ill., 2012.

[38] VerDate Mar. 2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas. *Federal Register*, 75(123):8, 2010.

[39] Alanna M. Chamberlain, Lila J. Finney Rutten, Patrick M. Wilson, Chun Fan, Cynthia M. Boyd, Debra J. Jacobson, Walter A. Rocca, and Jennifer L. St Sauver. Neighborhood socioeconomic disadvantage is associated with multimorbidity in a geographically-defined community. *BMC public health*, 20(1):13, January 2020.

[40] All of Us Research Program Overview, June 2020.

[41] UNDP, editor. *Beyond income, beyond averages, beyond today: inequalities in human development in the 21st century*. Number 2019 in Human development report / publ. for the United Nations Development Programme (UNDP). United Nations Development Programme, New York, NY, 2019.

[42] Laura Dwyer-Lindgren, Amelia Bertozzi-Villa, Rebecca W. Stubbs, Chloe Morozoff, Johan P. Mackenbach, Frank J. van Lenthe, Ali H. Mokdad, and Christopher J. L. Murray. Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA Internal Medicine*, 177(7):1003, July 2017.

[43] Martha Craven Nussbaum. *Creating capabilities: the human development approach*. Belknap Press of Harvard University Press, Cambridge, Mass, 2011.

[44] Amartya Sen and James E. Foster. *On economic inequality*. Clarendon Press ; Oxford University Press, Oxford : New York, 1997.

[45] Martha C. Nussbaum. Philosophy and Economics in the Capabilities Approach: An Essential Dialogue. *Journal of Human Development and Capabilities*, 16(1):1–14, January 2015.

[46] Mahbub Haq. *Reflections on human development*. Oxford University Press, Oxford, UK, 1995.

[47] Pedro Conceição. *The next frontier: human development and the Anthropocene*. Human development report. United Nations Development Programme, New York, NY, 2020.

[48] Stephen J. Agostini and Sandra J. Richardson. A Human Development Index for U.S. Cities: Methodological Issues and Preliminary Findings. *Real Estate Economics*, 25(1):13–41, 1997. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1540-6229.00706.

[49] Robert J. Sampson. Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure. *American Journal of Sociology*, 114(1):189–231, July 2008.

[50] Patrick Sharkey and Jacob W. Faber. Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects. *Annual Review of Sociology*, 40(1):559–579, 2014. _eprint: https://doi.org/10.1146/annurev-soc-071913-043350.

[51] William Julius Wilson. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press, 1987.

[52] Robert J. Sampson, Jeffrey D. Morenoff, and Thomas Gannon-Rowley. Assessing "Neighborhood Effects": Social Processes and New Directions in Research. *Annu. Rev. Sociol.*, 28(1):443–478, August 2002.

[53] Steven N. Durlauf. Chapter 50 Neighborhood effects. In *Handbook of Regional and Urban Economics*, volume 4, pages 2173–2242. Elsevier, 2004.

[54] Eric Chyn and Lawrence F Katz. Neighborhoods Matter: Assessing the Evidence for Place Effects. Working Paper 28953, National Bureau of Economic Research, June 2021. Series: Working Paper Series.

[55] Ana V. Diez Roux. Investigating Neighborhood and Area Effects on

Health. *American Journal of Public Health*, 91(11):1783–1789, 2001. _eprint: https://doi.org/10.2105/AJPH.91.11.1783.

[56] Katherine King and Christin Ogle. Negative Life Events Vary by Neighborhood and Mediate the Relation between Neighborhood Context and Psychological Well-Being. *PLOS ONE*, 9(4):e93539, April 2014.

[57] Patrick Sharkey, Gerard Torrats-Espinosa, and Delaram Takyar. Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime. *American Sociological Review*, 82(6):1214–1240, 2017. _eprint: https://doi.org/10.1177/0003122417736289.

[58] Charles C. Branas, Eugenia South, Michelle C. Kondo, Bernadette C. Hohl, Philippe Bourgois, Douglas J. Wiebe, and John M. MacDonald. Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *PNAS*, 115(12):2946–2951, March 2018.

[59] Patrick Sharkey. *Uneasy peace: the great crime decline, the renewal of city life, and the next war on violence*. W. W. Norton & Company, New York, first edition edition, 2018.

[60] Finn Hedefalk and Martin Dribe. The social context of nearest neighbors shapes educational attainment regardless of class origin. *PNAS*, 117(26):14918–14925, June 2020.

[61] Sara B. Heller, Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *The Quarterly Journal of Economics*, 132(1):1–54, February 2017.

[62] J. Ludwig, G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu. Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science*, 337(6101):1505–1510, September 2012.

[63] Raj Chetty, John Friedman, Nathaniel Hendren, Maggie Jones, and Sonya Porter. The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. Technical Report w25147, National Bureau of Economic Research, Cambridge, MA, October 2018.

[64] Raj Chetty and Nathaniel Hendren. The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, 133(3):1107–1162, August 2018.

[65] Martha A Gephart. Neighborhoods and Communities as Contexts for Development. In Aber JL Brooks-Gunn J, Duncan GJ, editor, *Neighborhood Poverty: Context and Consequences for Children*, pages 1–43. Russell Sage Foundation, New York, NY, 1997. Section: 1.

[66] Luis M. A. Bettencourt. *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems*. MIT Press, August 2021.

[67] Mandi Yu, Zaria Tatalovich, James T. Gibson, and Kathleen A. Cronin. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer Causes Control*, 25(1):81–92, January 2014.

[68] UNESCO Institute for Statistics. *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics, Montreal, Quebec, 2012. OCLC: 879577639.

[69] Georges Nguefack-Tsague, S. Klasen, and W. Zucchini. On Weighting the Components of the Human Development Index: A Statistical Justification. *Journal of Human Development and Capabilities*, 2011.

[70] Dylan Shane Connor and Michael Storper. The changing geography of social mobility in the United States. *PNAS*, 117(48):30309–30317, December 2020.

[71] Max Besbris, Jacob William Faber, Peter Rich, and Patrick Sharkey. Effect of neighborhood stigma on economic transactions. *PNAS*, 112(16):4994–4998, April 2015.

[72] Robert Manduca and Robert J. Sampson. Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *PNAS*, 116(16):7772–7777, April 2019.

[73] S. V. Subramanian, Jarvis T. Chen, David H. Rehkopf, Pamela D. Waterman, and Nancy Krieger. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *American Journal of Public Health*, 95(2):260–265, 2005.

[74] Mario L. Small, Robert A. Manduca, and William R. Johnston. Ethnography, Neighborhood Effects, and the Rising Heterogeneity of Poor Neighborhoods across Cities. *City & Community*, 17(3):565–589, September 2018.

[75] Robert J. Sampson and Alix S. Winter. The Racial Ecology of Lead Poisoning: Toxic Inequality in Chicago Neighborhoods, 1995-2013. *Du Bois Review: Social Science Research on Race*, 13(2):261–283, 2016. Publisher: Cambridge University Press.

[76] Joanna Maselko. Social Epidemiology and Global Mental Health: Expanding the Evidence from High-Income to Low- and Middle-Income Countries. *Curr Epidemiol Rep*, 4(2):166–173, June 2017.

[77] Elizabeth O. Ananat, Anna Gassman-Pines, Dania V. Francis, and Christina M. Gibson-Davis. Linking job loss, inequality, mental health, and education. *Science*, 356(6343):1127–1128, June 2017.

[78] Matthew Desmond and Bruce Western. Poverty in America: New Directions and Debates. *Annu. Rev. Sociol.*, 44(1):305–318, July 2018.

[79] David Manley, Maarten van Ham, and Lina Hedman. Inherited and Spatial Disadvantages: A Longitudinal Study of Early Adult Neighborhood Careers of Siblings. *Annals of the American Association of Geographers*, 110(6):1670–1689, November 2020.

[80] US Census Bureau. American Community Survey (ACS), 2021. Publication Title: The United States Census Bureau.

[81] CDC. PLACES: Local Data for Better Health, April 2022.

[82] Xingyou Zhang, James B. Holt, Shumei Yun, Hua Lu, Kurt J. Greenlund, and Janet B. Croft. Validation of Multilevel Regression and Poststratification Methodology for Small Area Estimation of Health Indicators From the Behavioral Risk Factor Surveillance System. *American Journal of Epidemiology*, 182(2):127–137, July 2015.

[83] Leann M. Tigges, Irene Browne, and Gary P. Green. Social isolation of the urban poor: Race, class, and neighborhood effects on social resources. *Sociological Quarterly*, 39(1):53–77, 1998.

[84] Chris Hess. Residential Segregation by Race and Ethnicity and the Changing Geography of Neighborhood Poverty. *Spatial Demography*, 9(1):57–106, April 2021.

[85] Erik Schokkaert. The capabilities approach. *Available at SSRN 1084821*, 2008.

[86] Jeffrey D Sachs. From Millennium Development Goals to Sustainable Development Goals. *The Lancet*, 379(9832):2206–2211, June 2012.

[87] Sabina Alkire. Dimensions of Human Development. *World Development*, 30(2):181–205, February 2002.

[88] Amartya Sen. A Decade of Human Development. *Journal of Human Development*, 1(1):17–23, February 2000.

[89] Frances Stewart, Gustav Ranis, and Emma Samman. *Capabilities and Human Development*, volume 1. Oxford University Press, March 2018.

[90] Elizabeth A. Stanton and Frank Ackerman. Florida and climate change: the costs of inaction. *Florida and climate change: the costs of inaction.*, 2007. Publisher: Global Development and Environment Institute, Tufts University.

[91] Dominik Paprotny. Convergence Between Developed and Developing Countries: A Centennial Perspective. *Social Indicators Research*, 153(1):193–225, 2021.

[92] John R. Campbell and Jeremy Holland. Development research: convergent or divergent approaches and understandings of poverty? An introduction. *Focaal*, 2005(45):3–17, 2005. Publisher: Berghahn Journals.

[93] Robert T. Barro and Xavier Sala-I-Martin. Regional growth and migration: A Japan-United States comparison. *Journal of the Japanese and International Economies*, 6(4):312–346, December 1992.

[94] Khalid Malik. *The rise of the South: human progress in a diverse world.* Human development report. United Nations Development Programm, Ney York, NY, 2013.

[95] Mark McGillivray and Nora Markova. Global Inequality in Well-being Dimensions. *The Journal of Development Studies*, 46(2):371–378, February 2010. Publisher: Routledge _eprint: https://doi.org/10.1080/00220380903033280.

[96] George Gray Molina and Mark Purser. Human development trends since 1970: A social convergence story. *UNDP-HDRO Occasional Papers*, 2(2010/2), 2010.

[97] Vanesa Jordá and José María Sarabia. Well-Being Distribution in the Globalization Era: 30 Years of Convergence. *Applied Research in Quality of Life*, 10(1):123–140, March 2015.

[98] Carlos Mendez and Felipe Santos-Marquez. Economic and Social Disparities across Subnational Regions of South America: A Spatial Convergence Approach. *Comparative Economic Studies*, January 2022.

[99] Anand Sahasranaman and Luís MA Bettencourt. Economic geography and the scaling of urban and regional income in India. *Environment and Planning B: Urban Analytics and City Science*, 48(3):540–554, March 2021. Publisher: SAGE Publications Ltd STM.

[100] Daniel Zünd and Luís M. A. Bettencourt. Growth and development in prefecture-level cities in China. *PLOS ONE*, 14(9):e0221017, September 2019. Publisher: Public Library of Science.

[101] Alaska Native Epidemiology Center. Alaska Native Health Status Report: Third Edition. Technical report, Alaska Native Epidemiology Center, Anchorage, AK, 2021.

[102] Michelle Theriault Boots Updated: December 2, 2017 Published: May 9, and 2017. Alaska sees the most dramatic increases in life expectancy in the nation, new study says.

[103] Appalachian Counties Served by ARC.

[104] Gopal K. Singh, Michael D. Kogan, and Rebecca T. Slifkin. Widening Disparities In Infant Mortality And Life Expectancy Between Appalachia And The Rest Of The United States, 1990–2013. *Health Affairs*, 36(8):1423–1432, August 2017.

[105] Steven H. Woolf, Heidi Schoomaker, Latoya Hill, and Christine M. Orndahl. The Social Determinants of Health and the Decline in U.S. Life Expectancy Implications for Appalachia: Implications for Appalachia. *Journal of Appalachian Health*, 1(1):6–14, 2019. Publisher: The University Press of Kentucky.

[106] Leah R. Abrams, Mikko Myrskylä, and Neil K. Mehta. The growing rural–urban divide in US life expectancy: contribution of cardiovascular disease and other major causes of death. *International journal of epidemiology*, 50(6):1970–1978, 2021. Publisher: Oxford University Press.

[107] Yana C. Vierboom, Samuel H. Preston, and Arun S. Hendi. Rising geographic inequality in mortality in the United States. *SSM - Population Health*, 9:100478, December 2019.

[108] National Academies of Sciences, Engineering, and Medicine. *High and Rising Mortality Rates Among Working-Age Adults*. The National Academies Press, March 2021.

[109] Kristen Pettrone. Urban–rural Differences in Suicide Rates, by Sex and Three Leading Methods: United States, 2000–2018. *NCHS Data Brief*, 373(373):8, 2020.

[110] N. Douthit, S. Kiv, T. Dwolatzky, and S. Biswas. Exposing some important barriers to health care access in the rural USA. *Public Health*, 129(6):611–620, June 2015.

[111] George A. Erickcek and Hannah McKinney. "Small Cities Blues:" Looking for Growth Factors in Small and Medium-Sized Cities. *Economic Development Quarterly*, 20(3):232–258, August 2006. Publisher: SAGE Publications Inc.

[112] Health Planning Council of Southwest Florida, Inc. Collier County Community Health Status Comprehensive Report, 2022.

[113] Arthur J. Hallinan Jr. A review of the Weibull distribution. *Journal of Quality Technology*, 25(2):85–93, 1993. Publisher: Taylor & Francis.

[114] Abdisalam Hassan Muse, Samuel M. Mwalili, and Oscar Ngesa. On the Log-Logistic Distribution and Its Generalizations: A Survey. *International Journal of Statistics and Probability*, 10(3):93, April 2021.

[115] Dejian Lai, Shan P. Tsai, and Robert J. Hardy. Impact of HIV/AIDS on life expectancy in the United States. *AIDS*, 11(2):203–207, February 1997.

[116] Dejian Lai and Robert J. Hardy. An update on the impact of HIV/AIDS on life expectancy in the United States. *AIDS*, 18(12):1732–1734, August 2004.

[117] Majid Ezzati, Ari B. Friedman, Sandeep C. Kulkarni, and Christopher J. L. Murray. The Reversal of Fortunes: Trends in County Mortality and Cross-County Mortality Disparities in the United States. *PLOS Medicine*, 5(4):e66, April 2008. Publisher: Public Library of Science.

[118] Linda S. Geiss, Karen Kirtland, Ji Lin, Sundar Shrestha, Ted Thompson, Ann Albright, and Edward W. Gregg. Changes in diagnosed diabetes, obesity, and physical inactivity prevalence in US counties, 2004-2012. *PLOS ONE*, 12(3):e0173428, March 2017. Publisher: Public Library of Science.

[119] Alaska and Native Epidemiology Center;. Alaska Native Mortality: 1980-2018. Technical report, Alaska Native Epidemiology Center;, Anchorage, AK, 2021.

[120] Christine Golnick, Elvin Asay, Ellen Provost, Dabney Van Liere, Cora Bosshart, Jean Rounds-Riley, Katie Cueva, and Thomas W. Hennessy. Innovative primary care delivery in rural Alaska: a review of patient encounters seen by community health aides. *International Journal of Circumpolar Health*, 71(1):18543, January 2012. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/ijch.v71i0.18543.

[121] Ali Roghani. The relationship between macro-socioeconomics determinants and COVID-19 vaccine distribution. *AIMS Public Health*, 8(4):655–664, September 2021.

[122] Andreas Y. Troumbis. Testing the socioeconomic determinants of COVID-19 pandemic hypothesis with aggregated Human Development Index. *J Epidemiol Community Health*, 75(4):414–415, April 2021. Publisher: BMJ Publishing Group Ltd Section: PostScript.

[123] Andrew J. Stier, Marc G. Berman, and Luis M. A. Bettencourt. COVID-19 attack rate increases with city size. preprint, Epidemiology, March 2020.

[124] Homi Kharas. The impact of COVID-19 on global extreme poverty, October 2020.

[125] R. B. Hawkins, E. J. Charles, and J. H. Mehaffey. Socio-economic status and COVID-19–related cases and fatalities. *Public Health*, 189:129–134, December 2020.

[126] Milorad Kovacevic and Admir Jahic. COVID-19 and Human Development: Exploring Global Preparedness and Vulnerability. *Human Development Reports*, page 5, 2020.

[127] Jonathan Heathcote, Fabrizio Perri, and Giovanni L. Violante. Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics*, 13(1):15–51, January 2010.

[128] Michael Berlemann and Max Friedrich Steinhardt. Climate Change, Natural Disasters, and Migration—a Survey of the Empirical Evidence. *CESifo Economic Studies*, 63(4):353–385, December 2017.

[129] Noah S. Diffenbaugh and Marshall Burke. Global warming has increased global economic inequality. *Proceedings of the National Academy of Sciences*, 116(20):9808–9813, May 2019. Publisher: Proceedings of the National Academy of Sciences.

[130] Suraj Sheth. Risk Index Spatial Clustering (RISC): Identifying High Risk Counties Using Local Moran's I and Spatial Statistics for Natural Disaster Risk Management :

Leveraging Spatial Tools for Dynamic Risk Assessment, Resilience Planning And Resource Management Across Spatial Scales. In *2022 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 39–43, April 2022.

[131] Grant M. A. Wyper, Ricardo Assunção, Sarah Cuschieri, Brecht Devleesschauwer, Eilidh Fletcher, Juanita A. Haagsma, Henk B. M. Hilderink, Jane Idavain, Tina Lesnik, Elena Von der Lippe, Marek Majdan, Milena S. Milicevic, Elena Pallari, José L. Peñalvo, Sara M. Pires, Dietrich Plaß, João V. Santos, Diane L. Stockton, Sofie Theresa Thomsen, and Ian Grant. Population vulnerability to COVID-19 in Europe: a burden of disease analysis. *Archives of Public Health*, 78(1):47, May 2020.

[132] Ari-Veikko Anttiroiko and Arto Haveri. Local Actions to Combat Covid-19 Crisis: Contextual Insights into Local Institutional Responses to Covid-19 in Europe and the United States. In Carlos Nunes Silva, editor, *Local Government and the COVID-19 Pandemic: A Global Perspective*, Local and Urban Governance, pages 59–83. Springer International Publishing, Cham, 2022.

[133] Rafael H. Forero H and Remy Sietchiping. Metropolises Overcoming the COVID-19 Pandemic: An Urgent Call for Territorializing Global Agendas at Subnational Levels. In Carlos Nunes Silva, editor, *Local Government and the COVID-19 Pandemic: A Global Perspective*, Local and Urban Governance, pages 85–102. Springer International Publishing, Cham, 2022.

[134] Suhasini Ramisetty-Mikler, Armin R Mikler, Martin O'Neill, and Jared Komatz. Conceptual framework and quantification of population vulnerability for effective emergency response planning. *Journal of emergency management (Weston, Mass.)*, 13(3):227–238, May 2015.

[135] Eric Tate. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, 63(2):325–347, September 2012.

[136] Shirley Laska and Betty Hearn Morrow. Social Vulnerabilities and Hurricane Katrina: An Unnatural Disaster in New Orleans. *Marine Technology Society Journal*, 40(4), 2006.

[137] Elizabeth Fussell. The Long-Term Recovery of New Orleans' Population After Hurricane Katrina. *American Behavioral Scientist*, 59(10):1231–1245, September 2015. Publisher: SAGE Publications Inc.

[138] Barry E. Flanagan, Elaine J. Hallisey, Erica Adams, and Amy Lavery. Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers for Disease Control and Prevention's Social Vulnerability Index. *Journal of environmental health*, 80(10):34–36, June 2018.

[139] Barry E. Flanagan, Edward W. Gregory, Elaine J Hallisey, Janet L. Heitgerd, and Brian Lewis. A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, 8(1), January 2011.

[140] Hannah Andrea Tarling. Comparative Analysis of Social Vulnerability Indices: CDC's SVI and SoVI®. *Lund University Dissertation*, 2017.

[141] CDC SVI Documentation 2018, February 2022.

[142] Samuel Rufat, Eric Tate, Christopher T. Emrich, and Federico Antolini. How Valid Are Social Vulnerability Models? *Annals of the American Association of Geographers*, 109(4):1131–1153, July 2019.

[143] Eric Tate. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, 63(2):325–347, September 2012.

[144] Sudhir Anand and Amartya Sen. Human Development Index: Methodology and Measurement. *UNDP Human Development Report Office*, 1994.

[145] Carlos Dornels Freire de Souza, Michael Ferreira Machado, and Rodrigo Feliciano do Carmo. Human development, social vulnerability and COVID-19 in Brazil: a study of the social determinants of health. *Infectious Diseases of Poverty*, 9(1):124, August 2020.

[146] T C Quinn. Population migration and the spread of types 1 and 2 human immunodeficiency viruses. *Proceedings of the National Academy of Sciences*, 91(7):2407–2414, March 1994.

[147] Paul R Flora. Regional Spotlight: Poverty in Philadelphia, and Beyond. *Federal Reserve Bank of Philadelphia Research Department*, page 7, 2021.

[148] Márcia de Cantuária Tauil, Ana Paula Sayuri Sato, and Eliseu Alves Waldman. Factors associated with incomplete or delayed vaccination across countries: A systematic review. *Vaccine*, 34(24):2635–2643, May 2016.

[149] Alfonso J. Rodríguez-Morales and Diana M. Castañeda-Hernández. Relationships between morbidity and mortality from tuberculosis and the human development index (HDI) in Venezuela, 1998–2008. *International Journal of Infectious Diseases*, 16(9):e704–e705, September 2012.

[150] Miranda M. Fidler, Freddie Bray, and Isabelle Soerjomataram. The global cancer burden and human development: A review. *Scandinavian Journal of Public Health*, 46(1):27–36, February 2018. Publisher: SAGE Publications Ltd STM.

[151] S. Hassanipour-Azgomi, Abdollah Mohammadian-Hafshejani, Mahshid Ghoncheh, Farhad Towhidi, Saeid Jamehshorani, and Hamid Salehiniya. Incidence and mortality of prostate cancer and their relationship with the Human Development Index worldwide. *Prostate International*, 4(3):118–124, September 2016.

[152] J. Madison Hyer, Diamantis I. Tsilimigras, Adrian Diaz, Rayyan S. Mirdad, Rosevine A. Azap, Jordan Cloyd, Mary Dillhoff, Aslam Ejaz, Allan Tsung, and Timothy M. Pawlik. High Social Vulnerability and "Textbook Outcomes" after Cancer Operation. *Journal of the American College of Surgeons*, 232(4):351–359, April 2021.

[153] Akhil Mehta, Won Jin Jin Jeon, Christopher Hino, Eric Lau, Hamid R. Mirshahidi, and Gayathri Nagaraj. Association of U.S. county social vulnerability with cancer mortality. *Journal of Clinical Oncology*, 40(16_suppl):6534–6534, June 2022. Publisher: Wolters Kluwer.

[154] Pytrik Dieuwke Oosterhof. Localizing the SDGs to Accelerate the Implementation of the 2030 Agenda for Sustainable Development:: The Current State of Sustainable Development Goal Localization in Asia and the Pacific. Governance Briefs, Asian Development Bank, Manila, Philippines, December 2018.

[155] Jennifer D Williams. The American Community Survey: Development, Implementation, and Issues for Congress. *Congressional Research Service*, 0(0):27, June 2013.

[156] Institute for Health Metrics and Evaluation (IHME). US Health Map, March 2014.

[157] United States Small-Area Life Expectancy Estimates Project, June 2020.