Index of Deep Disadvantage: Technical Documentation

Silvia Robles
Jasmine Simington
H. Luke Shaefer
University of Michigan

In ***Understanding Communities of Deep Disadvantage***, we use methods refined through Kathryn Edin and Luke Shaefer's book *$2.00 a Day*, combining big data with systematic, in-depth qualitative interviews and ethnographic observations to better understand the history and contemporary dynamics in a subset of communities of deep disadvantage. We seek to paint a vivid portrait of the lived-experiences of poor individuals and families in these places, and explore the views and activities of key stakeholders, including politicians, civil servants, faith leaders, and representatives of the business community. From these in-depth conversations, as well as our observations of community events, we seek to uncover community factors that may drive the sharp disparities documented in our big data work.

The first step in this project was to build a measure of "disadvantage" that could be applied consistently to communities all over the country. To this end, we constructed a unique, multidimensional *Index of Deep Disadvantage* (IDD) for all counties and the 500 largest cities in the U.S. We drew on census and administrative data to examine vulnerability in three interconnected domains of high salience to Americans: 1) income, using poverty and deep poverty rates that are official metrics of well-being for the nation; 2) health, using life expectancy and low birth weight, both of which are deeply connected to well-being over the life course; and 3) social mobility, using new estimates for counties and cities. We used principal component analysis to weight these variables (standardized for comparison). Then, we ranked communities on a continuum of disadvantage. We tested the robustness of our findings through inclusion of numerous other factors related to disadvantage. The data we used for this exercise can be accessed here. We invite others to utilize it to build their own index.

Sources of Data

Income: We include two related measures of income poverty: the share of community residents with cash incomes falling below poverty, and the share with cash incomes falling below 50% of poverty, often referred to as deep poverty. The estimates for counties and cities come from the 2017 American Community Survey, 5-year estimates. Poverty is defined according to the Office of Management and Budget (OMB), and historical poverty thresholds are available on the U.S. Census Bureau website. The amount of cash income designated as falling below the poverty level varies by household size, but not by geographic area, and is adjusted for inflation using the Consumer Price Index.

Health: The index also takes into account life expectancy at birth, from the 2019 RWJF County Health Rankings and the City Health Dashboard. Life expectancy was estimated using small area methods from deidentified death records from the National Center for Health Statistics (NCHS), and population counts from the U.S. Census Bureau, NCHS, and the Human Mortality Database.

For counties, life expectancy is measured over the time period of 2015-2017. For cities, life expectancy is a six-year average for the time period 2010-2015.

The measure for low birth weight represents the share of live births weighing less than 2500 grams in each county or city. Information on low birth weight is available from the 2019 RWJF County Health Rankings and the City Health Dashboard, building from data housed at the National Center for Health Statistics Natality Files. The RWJF County Health Ranking documentation notes that "these data are submitted by the vital registration systems operated in the jurisdictions legally responsible for registering vital events (i.e. births, deaths, marriages, divorces, and fetal deaths). Missing values are reported for all counties where fewer than 20 births were considered low birth weight." For cities, low birthweight is a three year average for the time period 2015-2017. For counties, the measure captures the time period of 2011-2017. Life expectancy and low birth weight together constitute the health domain for the indicator.

Social mobility: Data on adult income mobility by county are publicly available from Chetty and Hendren (2018). These observational estimates represent the mean household income rank for children whose parents were at the 25th percentile of the national income distribution. Income is measured as mean earnings in 2014-15 for individuals at age 26 and household income is defined as the sum of own and spouse's income. By focusing on people whose parents were in the 25th percentile of the national income rankings, the measure characterizes intergenerational mobility specifically for low-income populations who may be more vulnerable to variation in local conditions relative to high-income populations who can use their own resources to insulate themselves from differences in community resources. Chetty and Hendren (2018) note that these observational mobility estimates actually represent a large portion of the causal effect of living in a given county on whether a child moved up or down in the national income distribution relative to the average 26-year old. To estimate mobility for the top 500 cities, the index uses the same mean household income rank variable from Opportunity Insights, but starting at the tract level, aggregates up to the city. We verified the legitimacy of the tract to city aggregation by aggregating from tract to county, then comparing the aggregated tract-to-county estimates with the publicly available county estimates. Less than 1% of counties had an aggregate tract-to-county measure that was one standard deviation away from the county estimate, giving us confidence that the tract to city aggregation is sound.

Geographic unit of analysis

Defining "communities" is challenging. In our case, the primary objective is to compare communities across a set of consistent metrics, and so counties and cities become our primary units of analysis. To our knowledge, this is the first such index to compare both counties and cities on the same outcomes. The reason for doing this is to identify cities of deep disadvantage situated in counties where the non-city population has more resources. For instance, Wayne county, Michigan does not appear among the nation's most disadvantaged places according to our index, but the City of Detroit, within Wayne County, does.

One might argue that some cities with extremely high rates of inequality will not appear on our index, such as New York City or Los Angeles, because the disadvantage in these places is shrouded by their proximity to areas of affluence. Yet the same argument could be made for all

counties. We welcome ideas and debate about what the right unit of community is. And we would argue that while some might make the case that there are places missing from the IDD's list of communities of deepest disadvantage, the ones that do appear on our list are clearly so, given that they emerge even despite all these factors.

A novel aspect of our index, in fact, is that it ranks cities and counties together, with the goal of understanding the differences in vulnerability across place. The analysis excludes a handful of cities and counties for various reasons, most commonly:
(1) The county was a U.S. territory, and not a part of the continental U.S., Hawaii or Alaska; or
(2) The city was not one of the top 500 most populated cities; or
(3) The city and county were spatially equivalent units, therefore we excluded the county. Note in these cases that we find that the city and county estimates are virtually identical

The full list of geographic areas excluded, and the reason why, can be found in Appendix B.

Urban, Rural, and Tribal Lands
We classify counties as urban or rural using definitions from the National Center for Health Statistics, which offers more nuance than census definitions. Definitions drawn from this source are below:

Urban counties include: (1) Counties in MSAs of 1 million or more population that: contain the entire population of the largest principal city of the MSA; or have their entire population contained in the largest principal city of the MSA; or contain at least 250,000 inhabitants of any principal city of the MSA; (2) Other counties in MSAs of 1 million or more population but that did not qualify as large central metro counties; (3) Counties in MSAs of populations of 250,000 to 999,999; and (4) all other counties in metropolitan MSAs. Rural counties are all other non-urban counties, including counties that are in micropolitan statistical areas or nonmetropolitan counties that did not qualify as micropolitan.

Furthermore, counties that include any form of federally-recognized tribal land are flagged as such, using standard census and OMB definitions.

According to the census, the various types of Tribal land include: (1) reservations—areas with boundaries established by treaty, statute, and/or executive or court order. Reservations can cross state boundaries; (2) Tribal trust land—real property, held in trust by the Federal Government, that is associated with a specific American Indian reservation or tribe, or, in some cases, individual American Indians. Tribal trust land may be located inside or outside of a Reservation; (3) Tribal jurisdiction statistical areas—land associated with Federally recognized tribes in Oklahoma that no longer have a reservation; (4) Tribal designated statistical areas (TDSAs)— geographic entities delineated by Federally and State recognized tribes that have no reservation or trust lands. TDSAs consist of territory that contains the American Indian population over which a Federally recognized tribe has jurisdiction or territory within which a State-recognized tribe provides benefits and services to its members; and (5) Alaska Native Regional Corporations—There are 12 ANRCs that cover the entire state of Alaska.

Methodology

The IDD is the first principal component from a principal component analysis (PCA) of the five features. Principal component analysis is a technique often chosen to reduce the dimensions being considered in prediction-style analysis. This made it a natural choice to summarize the information carried across the five characteristics we chose. The five factors, while highly correlated (see Table A1), also highlight different aspects of the sources and consequences of deep disadvantage. PCA yields a weighted average of the five features where the weightings are chosen to capture as much of the variation in the observed data as possible.

The first step was to gather the variables. Table A2 lists the source and a URL for each piece of information included in the index. PCA is sensitive to the magnitude of each feature included. Therefore, after merging all five factors into one dataset, each feature was normalized by subtracting its mean and dividing by its standard deviation.

Missing data was a challenge for this procedure. The first approach to reduce the potential for bias due to missing information was to choose the county and city as the units of analysis. Many metrics are available nation-wide at the county (or city) level, but are not measured at a more fine-grained unit of analysis such as census tracts. In addition, the analysis focused on variables that were available for a large fraction of counties in the U.S. If a county was missing any of the features, the principal component analysis was completed with the remaining features. The analysis examined data only from mainland U.S. counties, Alaska, and Hawaii. U.S. territories were excluded.

The first principal component represents over sixty percent of the variation (or information) in the data. The weights on each variable are fairly even, with a slightly higher loading on the share of community residents in poverty.

Sensitivity Analyses

The project undertook two types of sensitivity analysis. The first checked for stability of the measure to the addition of different or additional characteristics. The second checked for face validity by examining whether the most disadvantaged communities (those with the lowest index rankings) displayed compound disadvantage. It would be less credible for a county to be ranked as extremely disadvantaged if most of the measures were not demonstrating disadvantage.

The results of testing for stability are available here. The first tab lists the top 50 most disadvantaged counties according to seven different specifications of the index. Each specification adds one or more additional feature to the core index. The additional features were either county arrest rates from the National Archive of Criminal Justice Data, the share of adult county residents with a high school diploma or less education from the ACS, the share of adult county residents who are underemployed (defined as percent working part-time or less than 34 hours per week) or combinations of these three additional features. The rankings are remarkably stable. The second tab describes the original rankings of any county that broke into the top 50 most disadvantaged counties under the different specifications. A maximum of seven counties move into the top 50 in any single alternative specification, and across all alternative

specifications, 10 unique counties break into the top 50. Only four of the ten have substantially different rankings in the core index. Overall, we conclude that the index is generally stable to the addition of other metrics.

Defining compounded disadvantage
To check for face validity of the index we define a measure of compound disadvantage where a place must rank in the bottom quartile of at least one characteristic in two of the three areas of focus. The areas of focus are poverty, health (life expectancy and low birth weight), and social mobility.

One of the outputs of the analysis is a ranked list of the 300 most disadvantaged localities. To qualify as a city or county in the top 300, localities had to meet the compound disadvantage criteria above. In the original ranking, 299 of 300 places met this criteria. When a place did not fit the definition of compounded disadvantage, it was bumped down in the ranking and replaced by the next highest ranked geography that did.

We consider the Index of Deep Disadvantage to be the start of a conversation. We welcome feedback. Visit the map that shows the results of this work, and the underlying dataset can also be found on the same page.

Appendix A. Supplemental Tables
Table A1: Correlation between any two components of the index

	Mobility	Poverty	Deep Poverty	Life Expectancy	Low Birth Weight
Mobility	1.00				
% Pov.	-0.53	1.00			
% Deep Pov.	-0.52	0.94	1.00		
Life Exp	0.45	-0.68	-0.56	1.00	
Low Birth Wt.	-0.59	0.56	0.51	-0.67	1.00

Table A2: List of component features of the index with links to their sources

Feature	**Data Source**

Chetty and Hendren. estimates	Opportunity Insights
Life expectancy	Robert Wood Johnson Foundation County Health Rankings (RWJ); City Health Dashboard
Percent of births that are low birth weight	Robert Wood Johnson Foundation County Health Rankings (RWJ); City Health Dashboard
Percent of residents below the poverty line	American Community Survey (ACS)
Percent of residents living in deep poverty	American Community Survey (ACS)

Appendix B. Geographic Sample

Dropped County	**Reason for dropping**
All 78 Puerto Rican counties	Dropped b/c U.S. territory
Hoonah-Angoon Census Area, Alaska	Boundaries have changed and has been annexed to other counties/ functions as a part of other counties
Prince of Wales-Hyder Census Area, Alaska	Boundaries have changed and has been annexed to other counties/ functions as a part of other counties
Wade Hampton Census Area, Alaska	Boundaries have changed and has been annexed to other counties/ functions as a part of other counties
Wrangell – Petersburg Census Area, Alaska	Boundaries have changed and has been annexed to other counties/ functions as a part of other counties
Kalawao County, Hawaii	Does not function like other counties: it has a very a small population & it is a judicial district of Maui County ,which includes the rest of the island of Moloka‘i. Kalawao county has no elected government
Bedford, Virginia	Bedford county added an independent city in 2013 that was previously separate. So , we have a record for an independent Bedford City at the FIPS level and for Bedford County at the FIPS level. SOLUTION: remove Bedford City and just use the FIPS code that represents the whole (Bedford) county
Clifton Forge, Virginia	The city of Clifton Forge (FIPS=51560) was incorporated into Allegheny County (FIPS=51005)
Anchorage, Alaska	In Alaska, municipalities, rather than counties, are the primary legal divisions. The municipality of Anchorage, Alaska is considered a consolidated city-county (shows up as both an independent city and county). Because of the overlap, this county is dropped.
Denver, Colorado	Denver is in our top 500 cities && is considered a consolidated city-county (shows up as both an independent city and county). Because of the overlap, the county is dropped.
Honolulu, Hawaii	Honolulu is in our top 500 cities && is considered a consolidated city-county (shows up as both an independent city and county). *However*, all Honolulu data from the ACS and RWJ were county-level metrics, so the city is dropped and the county is retained.

Orleans Parish, Louisiana	New Orleans is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of New Orleans and Orleans Parish). Because of the overlap, the county is dropped.
San Francisco, California	San Francisco is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of San Francisco and San Francisco County). Because of the overlap, the county is dropped.
Philadelphia County, Pennsylvania	Philadelphia is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of Philadelphia and Philadelphia County). Because of the overlap, the county is dropped.
New York City, New York	NYC is complex--each borough is considered its own county, but together the five boroughs represent New York City. is in our top 500 cities && because of the overlap, the county is dropped.
Washington, D.C.	D.C. is in our top 500 cities && is considered a consolidated municipal government, and the county appears to be comprised primarily of D.C. the city an da few unincorporated areas. Because of the overlap, the county is dropped.
Baltimore, Maryland	Baltimore is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of Baltimore and Baltimore County). Because of the overlap, the county is dropped.
Durham, North Carolina	Durham is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of Durham and Durham County). Because of the overlap, the county is dropped.
St. Louis, Missouri	St. Louis is in our top 500 cities && is considered a consolidated city-county (shows up as both the independent city of St. Louis and St. Louis County). Because of the overlap, the county is dropped.
Alexandria, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Chesapeake, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.

Hampton, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Lynchburg, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Norfolk, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Portsmouth, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Richmond, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Roanoke, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Suffolk, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.
Virginia Beach, Virginia	This is a consolidated city. The city is in our top 500 cities, and also shows up as a county. Because of the overlap, the county is dropped.