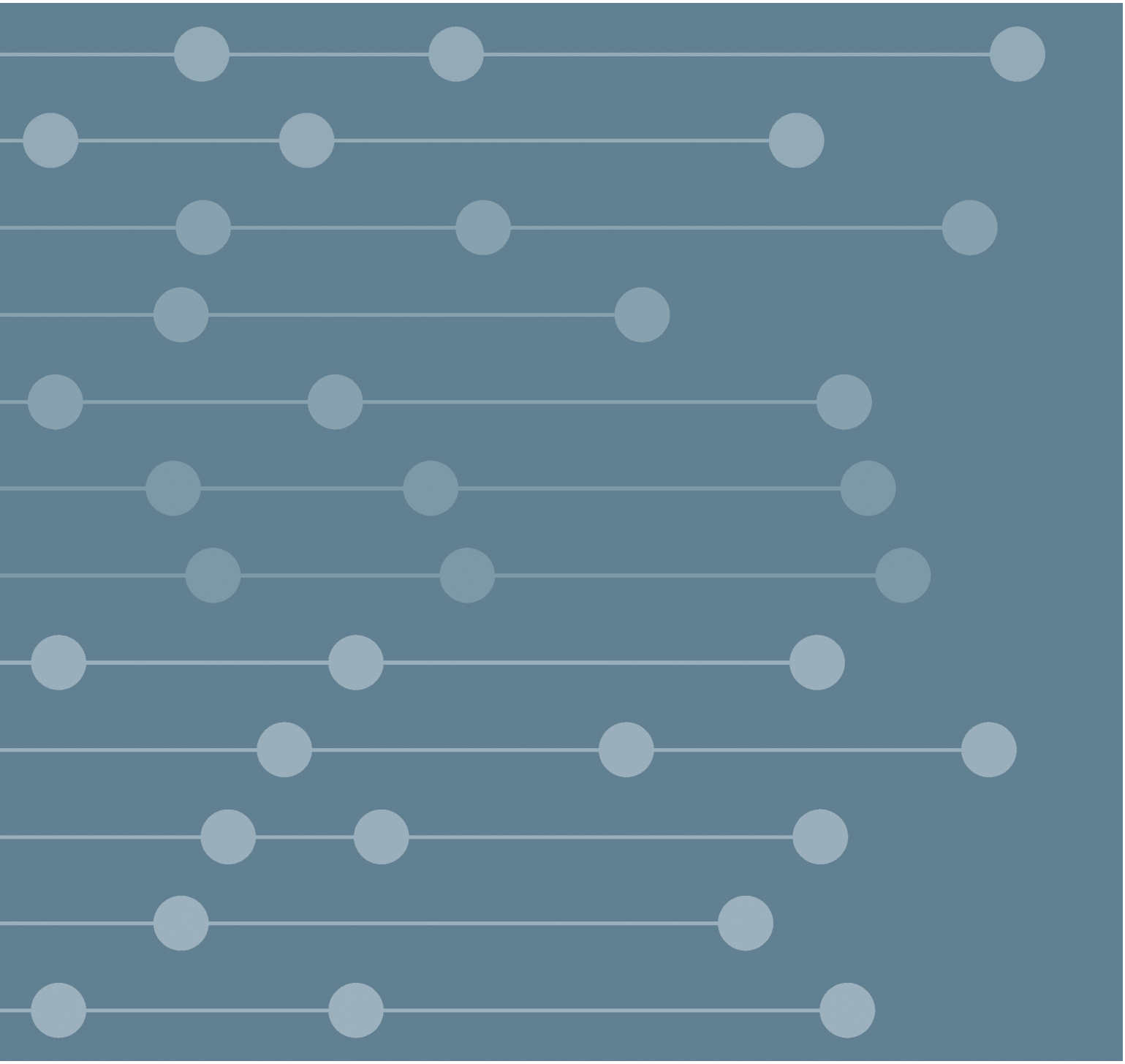
The HOPE Initiative

TECHNICAL SUMMARY

NATIONAL COLLABORATIVE FOR HEALTH EQUITY

TEXAS HEALTH INSTITUTE

VIRGINIA COMMONWEALTH UNIVERSITY CENTER ON SOCIETY AND HEALTH



Support provided by

Robert Wood Johnson Foundation

July 2018

Credits & Acknowledgements

Recommended Citation

Wilson, K., Turner, M., Chapman, D., Woolf, S., Smedley, B., & Andrulis, D. (2018). The HOPE Initiative: Technical Summary.

Lead Authors

Kim Wilson, Matt Turner, Derek Chapman, Steven Woolf, Brian Smedley, and Dennis Andrulis

The HOPE Initiative Staff

NATIONAL COLLABORATIVE FOR HEALTH EQUITY

Brian D. Smedley, PhD, Principal Investigator

Naima Wong Croal, PhD, MPH, Project Director

Jalisa Whitley, MPP

TEXAS HEALTH INSTITUTE

Dennis P. Andrulis, PhD, MPH, Co-Principal Investigator

Nadia Siddiqui, MPH

Kim Wilson, DrPH, MPIA

Anna Stelter, LMSW, MPH

Matt Turner, PhD, MPH

Meghan Cocking, MSSW

VIRGINIA COMMONWEALTH UNIVERSITY – CENTER ON SOCIETY AND HEALTH

Steven H. Woolf, MD, MPH, Co-Principal Investigator

Derek A. Chapman, PhD

Latoya Hill, MPH

Sarah Blackburn, MS

Lauren K. Snellings, MPH, CHES

This work was made possible with the valuable contributions of:

National Advisory Committee (NAC)

David R. Williams, PhD, MPH, Harvard University, Chair of The HOPE Initiative NAC

Dolores Acevedo Garcia, PhD, MPA-URP, Brandeis University

Jeanne Ayers, MPH, Minnesota Department of Health

Rajiv Bhatia, MD, MPH, US Department of Veterans Affairs

Paula Braveman, MD, MPH, University of California - San Francisco

Renee Canady, PhD, MPA, Michigan Public Health Institute

Janet Corrigan, PhD, MBA, Moore Foundation

Rachel Davis, MSW, Prevention Institute

Tom Eckstein, MBA, Arundel Metrics

Ed Ehlinger, MD, MSPH, Minnesota Department of Health (formerly)

Glenn Flores, MD, Connecticut Children's Medical Center

Marjory Givens, PhD, University of Wisconsin, Population Health Institute

Doug Jutte, MD, PhD, University of California - Berkley

Ernest Moy, MD, MPH, Centers for Disease Control and Prevention (CDC), National Center for Health Statistics

Ana Penman-Aguilar, PhD, MPH, CDC, Office of Minority Health and Health Equity

Malia Villegas, EdD, National Congress of American Indians (formerly)

Robert Wood Johnson Foundation

Dwayne Proctor, PhD, Senior Adviser to the President

C. Tracy Orleans, PhD, Senior Program Officer and Senior Scientist

Nicole Collins Bronzan, Communications Officer

J. Michael White, MPH, Program Associate

Elaine Arkin, Consultant, Special Adviser on The HOPE Initiative

Burness Communications

Sara Knoll, Vice President, Health & Social Policy

Janet Firshein, Principal and Co-Director, Health & Social Policy

Shannon Ryan, Vice President and Co-Director, Creative (formerly)

Special Thanks for Additional Review and Guidance

Robiel Abraha, MPH, Episcopal Health Foundation

Laudan Aron, MA, BSc, Urban Institute

Breannon Babbel, PhD, MPH, MPP, National Indian Health Board

Julia Berenson, MSc, Columbia University

Ashley Brown, MPP, Center for Global Policy Solutions

Juanita Chinn, PhD, CDC, National Center for Health Statistics

Stacy Cunningham, MS, National Network of Public Health Institutes

Jennifer Hadayia, Harris County Public Health

Julia Holmes, PhD, CDC, National Center for Health Statistics

David Huang, PhD, MPH, CDC, National Center for Health Statistics

Marjorie Innocent, PhD, NAACP

Anthony Iton, MD, JD, MPH, The California Endowment

Ginger Jacinic, LMSW, MPAff, City of Austin, Texas

Iyanrick John, JD, Asian & Pacific Islander American Health Forum

Michael Knapp, PhD, Green River Consulting

Kathy Ko Chin, MPH, Asian & Pacific Islander American Health Forum

Vincent Lafronza, EdD, National Network of Public Health Institutes

Leandris Liburd, PhD, MPH, CDC, Office of Minority Health and Health Equity

Adrienne McFadden, MD, JD, Virginia Department of Health (formerly)

Ramal Mooneshinghe, PhD, CDC, Office of Minority Health and Health Equity

Linda Rae Murray, MD, MPH, (retired)

Julie Nelson, Government Alliance on Race and Equity / Race Forward

Sarah Norman, MPP, NeighborWorks America

Brion Oaks, MPA, City of Austin, Texas

Mary Pittman, DrPH, Public Health Institute

Eiman Siddiqui, Chief of Staff, Texas State Representative Tom Oliverson

Shao-Chee Sim, PhD, MPA, Episcopal Health Foundation

Makram Tahlih, PhD, CDC, National Center for Health Statistics

Phil Tegler, JD, Poverty and Race Research Action Council

William M. Tierney, MD, Dell Medical School, University of Texas at Austin

Hieu Truong, MUP, NeighborWorks America

Table of Contents

Overview

This technical summary describes the methodological approach for The Health Opportunity and Equity (HOPE) Initiative. Methodologically, HOPE aims to create reliable and precise measures of differences in health outcomes and the determinants of health by race, ethnicity, and socioeconomic status (SES) at the national and state levels. Additionally, because the literature suggests that the health and opportunity benefits of SES are not experienced equally across all racial and ethnic groups, we present national data, where possible, by both race and ethnicity and SES (Braveman, 2005; Shavers, 2007).

Grouping Variables

RACE AND ETHNICITY

Generally, this project expended considerable effort to balance a desire for discrete, homogenous racial and ethnic groups with a practical need to create reliable estimates that are representative of the populations to which they pertain. Thus, HOPE provides data on the following six mutually-exclusive racial and ethnic groups:

- White;
- Black or African American (Black);
- Hispanic or Latino (Hispanic);
- Asian and Pacific Islander (Asian/PI);
- American Indian and Alaska Native (AI/AN); and
- Multiracial.

A catchall category covering "other" racial and ethnic groups was not utilized due to insufficient sample sizes across measures.

In using six groups, rather than the traditional four (Black, White, Hispanic, Other), we suppressed the results of some groups where samples or numbers were too small in their states due to concerns about the reliability of estimates for these groups. The HOPE Initiative, guided by its National Advisory Committee, embraced this tradeoff in order to best represent health outcomes and determinants within relatively discrete groups across most states. At the same time, we recognize that the six groups presented in HOPE analyses are not homogenous and that the life experiences of group members may vary considerably within each state. Unfortunately, given the practical considerations of data availability, creating more granular racial and ethnic groups was not feasible at the state-level.

SOCIOECONOMIC STATUS

When data about socioeconomic status were available, this element was grouped by either educational attainment or household income as a percent of the federal poverty level (FPL). Levels of educational attainment were condensed into four categories:

- Less than high school;
- High school graduate;
- Some college (including technical school and associates degrees); and
- College graduate.

Five income categories were used:

- 0-99% FPL;
- 100-199% FPL;
- 200-299% FPL;
- 300-399% FPL; and
- 400% FPL or greater.

As data were more frequently available for educational attainment, and income data tends to be subject to higher non-random, non-response rates, in most cases, where the data were available, we defaulted to using educational attainment as our SES grouping variable. Because educational attainment is often dependent upon age, many of our indicators were analyzed after restricting data to include adults ages 25 and older. To facilitate cross-group comparisons, these age restrictions were retained for racial and ethnic group analyses for measures that used educational attainment as the SES grouping variable.

Data Sources

The HOPE Initiative includes 28 indicators of health outcomes and the broader determinants of health, organized across five domains: health outcomes; socioeconomic factors; social environment; physical environment; and access to health care. Nine primary data sources were used to develop indicators covering these domains. For a full listing of domains, indicators, definitions, and data sources, see Table 1.

National Survey of Children's Health

The National Survey of Children's Health (NSCH) is a national survey of parents or other adult caregivers of children aged 0 to 17 years. With a sample size smaller than other longer-running federal surveys, the NSCH purports to yield nationally- and state-representative population estimates. TThe NSCH has been carried out four times, in 2003, 2007, 2011-2012, and 2016. The 2011-2012 version of the NSCH was used for these analyses. The NSCH as a data source presents a number of challenges that limits its utility for highlighting state-level racial and ethnic differences in child health outcomes and determinants. First, due to small group size and the survey's periodic (i.e., non-annual) administration schedule, it is not possible to achieve adequate group size by pooling data over multiple continuous years. A second, and related issue is that state-level survey results are suppressed for racial and ethnic groups that make up less than 5% of the child population in that state. Third, the publicly available NSCH data only reports by four standard racial and ethnic groups (White, Black, Hispanic, and Other) at the national level. Finally, due to lack of current data needed to create parental educational attainment population estimates (see below), we could not complete analyses by education for these data points. Thus, for the NSCH data, only analyses by household income were performed at the state-level. We note that 7.6% of respondents did not provide household income data and responses missing this data were removed from the sample. No attempts were made to impute missing income data. Sample sizes were sufficient at the state-level to retain all groups and states in the analyses.

Behavioral Risk Factor Surveillance System

The Behavioral Risk Factor Surveillance System (BRFSS) is a large national survey of adults (individuals 18 years of age and older) that utilizes sampling methods that enable the creation of state and national population estimates. Three years of data (2012-2014) were used for these analyses to ensure reliable estimates. Analyses were performed by race and ethnicity and educational attainment. They were not performed by household income due to missing data and difficulty in structuring income responses to align household income as a percent of the FPL. Pooling of data over three years resulted in sufficient group size for all states and groups except for Blacks in two states – Idaho and Montana – resulting in these states being left out of the race and ethnicity Distance to Goal rankings and state rates (explained below). While no group size issues were present in the single year educational attainment groups, we used 2012-2014 data for these analyses as well to facilitate comparisons between the two sets of analyses.

National Vital Statistics System

The National Vital Statistics System (NVSS) administered by the Centers for Disease Control and Prevention (CDC) collects and disseminates data on the nation's vital events including births, deaths, marriages, divorces, and fetal deaths. While data related to race and ethnicity is always available, data on educational attainment of the decedent (or the parent, in the case of child measures) is less so. On average (with the exception of Rhode Island), about 1.5% of mortality data was missing education information in each state. Rhode Island did not report educational attainment in the majority (about 95%) of deaths; thus, the state was excluded from analyses by level of education. There were also high rates of missing parental education data for infant indicators after 2010 due to a change in how the National Center for Health Statistics gathered parental education data. Therefore, while data from 2010 to 2014 were used for the premature mortality indicator, 2006 to 2010 data were used for the low birth weight and infant mortality measures. Because the NVSS data covers the population, as opposed to survey data, and cases of low birth weight and infant mortality are relatively rare events, we had an insufficient number of live births (specified as a minimum of 40 for the low birth weight measure and 400 for the infant mortality measure over the 5-year period) to reliably estimate rates of infant mortality and low birth weight for minority groups in some states. This resulted in the exclusion of the Multiracial group entirely from our analyses and excluding nine states (Delaware, District of Columbia, Hawaii, Kentucky, Montana, New Hampshire, Vermont, West Virginia, and Wyoming) from the state rankings for infant mortality and Vermont from the rankings for low birth weight. We encountered related issues with the premature mortality indicator when we restricted deaths to those aged 25-64 years old (see methods below) and split them into two age groups. In this case, annual population estimates for 25-44 year olds and 45-64 year olds needed to be at least 4,000 and 700, respectively, for the group's estimate to be considered reliable. As such, rather than dropping groups or states from the analyses, we retained all and identified potentially unreliable estimates using footnotes.

American Community Survey

The American Community Survey (ACS) is an ongoing statistical survey by the U.S. Census Bureau, which is sent to approximately 295,000 addresses monthly (or 3.5 million per year) making it the largest survey after the decennial census that the Census Bureau administers. Complete ACS data is publicly accessible via FactFinder. However, when one needs to simultaneously segment by two variables (e.g., education and race), as we did for many of our indicators, a subset of ACS data is accessed via the Public Use Microdata System (PUMS). One year (2010 or 2014) and five-year combined (2010-2014, 2011-2015 for one variable) files were the underlying data sources for income-related variables. For indicators that were to be presented by both race and ethnicity and SES, we used ACS PUMS data. For those that were just presented by race and ethnicity or SES, we used the larger ACS file. The latter resulted in ample within-group sample size to report all states and groups; use of the former occasionally resulted in within-group sample sizes for Blacks and AI/ANs in a few states that were insufficient for creating reliable estimates. In these cases, the rates for the groups were suppressed and the states in question were excluded from state rankings. Generally, we used pooled data from 2010 to 2014. Exceptions include the Health Insurance Coverage variable for which we used data from 2011 to 2015 to better reflect changes in coverage due to the implementation of the Affordable Care Act; and the Low Poverty Concentration measure accessed from the Neighborhood Change Database, which relies on 2006 to 2010 ACS data. In addition, data sourced from the Neighborhood Change Database was not age-restricted and was available only by race and ethnicity, and not by SES.

Area Health Resources File

The Area Health Resources File (AHRF) is a data warehouse that includes county-, state-, and national-level files in eight broad areas: Health Care Professions, Health Facilities, Population Characteristics, Economics, Health Professions Training, Hospital Utilization, Hospital Expenditures, and Environment. The AHRF data are obtained from more than 50 sources. For this project the underlying data source used was the 2014 American Medical Association's Physician Master File for data on access to primary care physicians and psychiatrists.

CDC Wonder

CDC Wonder's Outdoor Air Quality-Particulate Matter data are geographically aggregated daily measures of fine particulate matter ($PM_{2.5}$) in the outdoor air, spanning the years 2003-2011. $PM_{2.5}$ particles are air pollutants with an aerodynamic diameter less than 2.5 micrometers. Reported measures are the daily measure of fine particulate matter in micrograms per cubic meter (μg/m^3), the number of observations, minimum and maximum range value, and standard deviation. Data are available by place, time and specified fine particulate matter value. County-level and higher data are aggregated from 10-kilometer square spatial resolution grids. Estimates were developed by scientists at the National Aeronautics and Space Administration (NASA) Marshall Space Flight Center/Universities Space Research Association using data from two sources: the Environmental Protection Agency's (EPA) Air Quality System and NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth remotely sensed data. Data were not available for Alaska or Hawaii.

County Business Patterns

The U.S. Census Bureau's County Business Patterns (CBP) provides subnational economic data by industry, updated annually. Data include the number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll. CBP data are extracted from the Business Register, the Census Bureau's file of all known single and multi-establishment companies. Data come from a variety of sources, including the Economic Censuses, the Annual Survey of Manufacturers, and Current Business Surveys, as well as from administrative records of the Internal Revenue Service (IRS), the Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS). For this project, the 2014 CBP data served as the primary source on liquor store density.

Food Access Research Atlas

The USDA Food Access Research Atlas (FARA) maps food access indicators for census tracts using ½-mile and 1-mile demarcations to the nearest supermarket for urban areas, 10-mile and 20-mile demarcations to the nearest supermarket for rural areas, and vehicle availability for all tracts. It also includes data on whether census tracts are "low-income" as individual resources affect accessibility. A census tract is identified as low income if (a) its poverty rate is 20 percent or greater, (b) its median family income is less than or equal to 80 percent of the statewide median family income, or (c) it is in a metropolitan area and has a median family income less than or equal to 80 percent of the metropolitan area's median family income. Population data are from the 2010 Census of Population and Housing. Data on food outlets is derived from merging the 2015 STARS directory of stores authorized to accept SNAP benefits and the 2015 Trade Dimensions TDLinx directory of stores.

Uniform Crime Reporting

The FBI's Uniform Crime Report (UCR) Program has collected statistics from local law enforcement agencies on the number of known offenses of the following types - murder and non-negligent homicide, rape, robbery, aggravated assault, burglary, motor vehicle theft, larceny-theft, and arson – since 1930. Participation is voluntary and not all law enforcement agencies provide data for complete reporting periods. The FBI computes estimates for participating agencies not providing 12 months of complete data. For agencies supplying 3 to 11 months of data, the UCR Program creates estimates for the missing data by following a standard estimation procedure using the data provided by the agency. If an agency has supplied less than 3 months of data, the FBI computes estimates by using the known crime figures of similar areas within a state and assigning the same proportion of crime volumes to non-reporting agencies. The estimation process considers the following: population size covered by the agency; type of jurisdiction (e.g., police department versus sheriff's office); and geographic location. For the period and offenses used for this project, 94% of counties provided data that was 90% to 100% complete.

Table 1. HOPE Indicators

Indicator Name	Description	Data Source[1]	Age Restrictions	Segments[2]
Health Outcomes				
Adult Health Status	Portion of adults reporting excellent or very good health	BRFSS, 2012-2014	Adults 25 yo+	R/E, Educ
Mental Health Status	Portion of adults reporting 14 days or more of poor mental health in the past 30 days	BRFSS, 2012-2014	Adults 25 yo+	R/E, Educ
Child Health Status	Portion of children in excellent or very good health, as reported by parents	NSCH, 2011-2012	Children 0-17 yo	Income
Premature Mortality	Number of annual deaths due to any cause per 100,000 population age 25-64	NVSS, 2010-2014	Adults 25-64 yo	R/E, Educ
Infant Mortality	Number of infants who die before their first birthday annually per 1,000 live births	NVSS, 2006-2010	Infants 0-1 yo	R/E, Educ
Low Birth Weight	Portion of infants weighing less than 2,500 grams at birth	NVSS, 2006-2010	Newborns	R/E, Educ
Socioeconomic Factors				
Livable Income	Portion of people living in households with incomes greater than 250% of FPL	ACS, 2010-2014	Adults 25 yo+	R/E, Educ
Affordable Housing	Portion of households spending no more than 30% of monthly household income on housing and related expenses	ACS, 2010-2014	Households headed by adult 25 yo+	R/E, Educ
Post-secondary Education	Portion of adults with at least some college education after graduating from high school	ACS, 2010-2014	Adults 25 yo+	R/E, Income
Connected Youth	Portion of young people age 16-24 enrolled in school or working, including military enlistment	ACS, 2010-2014	Youth 16-24 yo	R/E
Preschool Enrollment	Portion of 3-4 year olds enrolled in preschool	ACS, 2010-2014	Children 3-4 yo	Income
Employment	Portion of labor force that is employed	ACS, 2010-2014	Individuals 16 yo+	R/E, Income
Access to Health Care				
Access to Primary Care	Portion of people living in counties with a population-to-primary case physician ratio of less than 2,000:1	AHRF, 2014	n/a	R/E, Income
Access to Psychiatric Care	Portion of people living in counties with a population-to-psychiatrist ratio of less than 30,000:1	AHRF, 2014	n/a	R/E, Income
Health Insurance Coverage	Portion of people under 65 years with any kind of health insurance	ACS PUMS, 2011-2015	Individuals ages 0-64 yo	R/E, Income
Affordable Health Care	Portion of adults reporting they did not delay or forgo care due to cost in the past year	BRFSS, 2012-2014	Adults 25 yo+	R/E, Educ
Usual Source of Care	Portion of adults who have someone they consider their personal health care provider	BRFSS, 2012-2014	Adults 25 yo+	R/E, Educ
Colorectal Cancer Screening	Portion of adults age 50-75 receiving recommended colorectal cancer screenings	BRFSS, 2012, 2014	Adults 50-75 yo	Educ

Table 1. HOPE Indicators

Indicator Name	Description	Data Source[1]	Age Restrictions	Segments[2]
Physical Environment				
Home Ownership	Portion of households living in owner-occupied homes	ACS, 2010-2014	Households headed by adult 25 yo+	R/E, Educ
Housing Quality	Portion living in households with no severe housing problems	ACS, 2010-2014	Households headed by adult 25 yo+	R/E, Educ
Air Quality – Particulate Matter	Portion of people living in counties with average daily fine particulate matter ($PM_{2.5}$) below 12 micrograms per cubic meter	CDC Wonder, 2011	n/a	R/E, Income
Low Liquor Store Density	Portion of people living in counties with fewer than 1.736 liquor stores per 10,000 people	U.S. Census Bureau County Business Patterns, 2014	n/a	R/E, Income
Food Security	Portion of people living in census tracts that are not food deserts (i.e., census tracts not designated low income and low food access)	USDA Food Access Research Atlas, 2015	n/a	R/E, Income
Social Environment				
Low Poverty Concentration	Portion of people in neighborhoods with less than 20% of residents living in poverty	ACS, 2010 from the Neighborhood Change Database	n/a	R/E
Low Murder Rate	Portion of people living in counties with fewer than 5.1 murders per 100,000 people annually	UCR, 2010-2012	n/a	R/E, Income
Low Assault Rate	Portion of people living in counties with fewer than 283 reported cases of aggravated assault per 100,000 population annually	UCR, 2010-2012	n/a	R/E, Income
Low Rape Rate	Portion of people living in counties with fewer than 36.9 cases of rape per 100,000 annually	UCR, 2010-2012	n/a	R/E, Income
Low Robbery Rate	Portion of people living in counties with fewer than 52.1 reported cases of robbery per 100,000 population annually	UCR, 2010-2012	n/a	R/E, Income

[1]Data Source abbreviations: ACS – American Community Survey; AHRF – Area Health Resource Files; BRFSS – Behavioral Risk Factor Surveillance System; USDA – U.S. Department of Agriculture; CDC – U.S. Centers for Disease Control and Prevention; NSCH – National Survey of Children's Health; NVSS - National Vital Statistics System, National Center for Health Statistics, U.S. Centers for Disease Control and Prevention; UCR – Federal Bureau of Investigation's Uniform Crime Reporting program.

[2]The segments column denotes what groups were compared for each variable. R/E – six groups defined by race and ethnicity; Educ – four groups defined by educational attainment; Income – five groups defined by household income as a percent of the federal poverty level.

Indicator Development

This section describes the process for developing HOPE's 28 indicators of health outcomes, socioeconomic factors, social and physical environment, and access to health care. Discussion of indicator development is organized by their primary source of data.

National Survey of Children's Health

One item related to overall child health was used from this data source. Specifically, parents/adult respondents were asked to rate their child's health as excellent, very good, good, fair or poor. In line with published research, we dichotomized children into two groups, those who have excellent or very good health and those whose health is good, fair or poor. The final indicator, **Child Health Status**, was defined as the portion of children in excellent or very good health, as reported by parents.

Behavioral Risk Factor Surveillance System

Two items related to adult health were used from BRFSS: a) overall health status and b) number of days in past 30 when mental health was not good. Similar to the NSCH overall health status question for children, BRFSS asks adults to rate their health as excellent, very good, good, fair or poor. Reponses were dichotomized as excellent or very good health and good, fair, or poor health. The resulting indicator, **Adult Health Status**, was defined as the portion of adults reporting excellent or very good health.

BRFSS' mental health item asks: "thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" Responses were dichotomized, in line with previous research and practice, as less than 14 and 14 and more days. The final indicator, **Mental Health Status,** was defined as the portion of adults reporting 14 days or more of poor mental health in the past 30 days.

Several variables from BRFSS were also used as indicators of health care access, namely: affordable health care, usual source of care, and colorectal cancer screening. **Affordable Health Care** was defined as the percentage of adults who did not report a time in which they needed to see a doctor but could not because of cost at least once in the past twelve months. Of note, this indicator includes those who did not need medical care in the past 12 months; therefore, states' results may be influenced by having younger or healthier populations (i.e., greater portions not needing to seek care in the past 12 months).

Usual Source of Care was defined as the percentage of adults who report having one or more people they think of as their personal doctor or health care provider. Finally, **Colorectal Cancer Screening** was defined as the percentage of adults aged 50 to 75 who report having received colorectal cancer screening per U.S. Preventive Services Task Force (USPSTF) guidelines in place between 2008 and 2016. While the first two indicators were derived for both race and ethnicity and SES, due to the narrow age range of the third and the resultant reduction in within-group sample size, this final indicator was available only by SES.

National Vital Statistics System

Premature mortality, infant mortality, and low birth weight are the three health outcomes measures that were derived from NVSS data. Measures are expressed as the rate per 100,000 persons for the all-cause indicator, as rate per 1,000 live births for the infant mortality indicator, and as portion of live births for low birth weight measures. Data from 2006-2010 were used for infant outcomes due to mother's education not being available after 2010 and for 2010-2014 for premature mortality.

Premature Mortality used all-cause mortality data for adults ages 25-64. Because likelihood of death increases geometrically with age and the age structures of states (and race and ethnic and education groups within states) varies, mortality data was segmented into two age groups (25-44, 45-64). To create age-specific mortality rates for each group of interest we: (1) segmented death counts for race and ethnic and education groups by age (25-44, 45-64) creating 8 groups for education analyses (4 education groups x 2 age groups) and 12 groups for race and ethnicity analyses (6 racial and ethnic groups x 2 age groups); (2) summed death counts from the years 2010 to 2014 to smooth year-to-year variations in death counts; and (3) divided by 5-year population totals for each group and then divided by 100,000 to arrive at age and education (or race and ethnic) group-specific mortality rates. No data was missing with respect to race and ethnicity; education data was missing from less than 5% of records except for Rhode Island which was missing over 90% of data and therefore was excluded from the SES analyses.

For the **Infant Mortality** and **Low Birth Weight** measures, state differences exist in how maternal education is reported on birth certificates. Data on maternal education for live births varies non-systematically in completeness of reporting across states and years. For the five-year period and 50 states and the District of Columbia (n=255 records), data on maternal education was completely suppressed in 43.5% of records and partially suppressed for one state in two years (in New York in 2006 and 2007, approximately 50% of records reported maternal education in both years). The process used to estimate number of live births by education in states and years where maternal education data was completely suppressed was to calculate the average percent live births by education level for years in which maternal education was not suppressed, and multiply by total number of births in that year. In no instance was there a high degree of variability in missing (not to be confused with "suppressed by NVSS") data across years that could potentially impact these results. In the one state with partially suppressed data, reported percentages of births by education level in years with partial suppression notably deviated from the narrow variation in percentages in the three years where data was not suppressed; therefore, HOPE Measures followed the same procedure as for states and years with complete suppression to calculate number of births for the two years (2006 and 2007) with partial suppression.

Number of States with Partial / Complete Suppression By Number of Years

	Partial	Complete
4 years	0	17
3 years	0	6
2 years	1	1
1 year	0	23
0 years	0	4

American Community Survey

The 2010-2014 ACS main file was used to create most of the state-level population estimates by race and ethnicity, education, and household income needed for HOPE's Distance to Goal analyses described in this document. Relevant files were pulled for the entire U.S. population and the state, race and ethnicity, education or household income, and age variables used to create annual average population percent and number estimates. Exceptions include state-level race and ethnicity estimates and state-level educational attainment estimates for adults ages 25 years old and over, both of which were derived using U.S. Census Population Projections, and where explicitly stated in the indicator description.

In addition, ten indicators were developed from the 2010-2014 ACS/ACS PUMS files: (a) livable income, (b) affordable housing, (c) adults with post-secondary education, (d) connected youth, (e) preschool enrollment, (f) rate of employment, (g) health insurance coverage (2011-2015), (h) home ownership, and (i) housing quality.

Livable Income was defined as the portion of households earning income above 250% FPL (using the U.S. Census FPL definitions). This indicator intends to measure what is considered a sustainable income, and was calculated based on reported household income and number of household occupants. The results were average annual population estimates for 2010-2014. The sample was restricted to households headed by adults 25 years of age or older since educational attainment was the indicator used for the SES analyses.

Affordable Housing was defined as the portion of households spending no more than 30% of monthly household income on rent, mortgage, and housing-related expenses (i.e. those not experiencing a housing cost burden). It should be noted that this variable, by its calculation methods, excludes households that do not report monthly income and/or who do not pay rent or mortgage. Thus, the results can only be generalized to households that have income and a rent or mortgage payment. In our assessments of portions of ACS PUMS respondents excluded from analyses due to these requirements, we found that AI/ANs were more likely to be excluded than other race and ethnic groups (7.6% of AI/ANs nationally compared to 3.0% average across all groups). In four states, more than 10% of AI/AN households were excluded: Alaska, Arizona, Iowa and New Mexico. As with the indicator above, the sample was restricted to households headed by adults aged 25 years and older.

Post-secondary Education was calculated as the average annual portion of adults with more than a high school education and was restricted to adults aged 25 years and older to control for varying age distributions (and their impact on educational attainment) of different race and ethnic (and income) groups.

Connected Youth was defined as the average annual percent of 16 to 24 years olds who are in school or working, including service in the military. These analyses were only run by race and ethnicity for two reasons: (1) connected youth is a measure of both education and employment (which influences income), and (2) both of our measures of SES would be confounded by age within this group (i.e. age effect on educational attainment and the effects of living at home versus independently on household income). Nine states were excluded from the analyses due to insufficient within-group sample size because of the limited age range used for one or more groups despite using five years of data. These include: Connecticut, Delaware, Hawaii, Montana, New Hampshire, Rhode Island, Vermont, West Virginia, Wyoming. The District of Columbia was excluded for the same reason.

Preschool Enrollment was defined as the portion of 3 and 4 years olds who are enrolled in preschool. Due to the small age range, within-group sample sizes for analyses by race and ethnicity were too small to create reliable estimates. For this reason, analyses are restricted to differences between SES groups, in this case using household income as the SES indicator.

Employment was defined as the portion of those 16 years and older in the labor force who are employed. Employment rates were derived by summing those 16 years old and over who (1) worked at any time during the reference week; (2) were on temporary layoff and were available for work; (3) who did not work during the reference week but who had jobs or businesses from

which they were temporarily absent (excluding layoff), and dividing by (1) + (2) + (3) + (4) where (4) are people who did not work during the reference week, but who were looking for work during the last four weeks and were available for work during the reference week. People not in the labor force were not included in the calculation of this indicator.

Health Insurance Coverage was defined as the portion of all individuals under the age of 65 with all types of insurance coverage. The data is from the 2011-2015 ACS PUMS dataset, a year later than other ACS data we have used in an effort to capture impacts of the Affordable Care Act. The SES variable used was household income as a percentage of FPL.

Home Ownership was defined as the average annual percent of households living in owner-occupied homes between 2010-2014. Because educational attainment was used as the SES grouping variable, the samples were restricted to households headed by an adult who was 25 years old or older.

The **Housing Quality** indicator is the average annual portion of households that do not have any serious problems (i.e. they have a complete kitchen, functioning plumbing, are not overcrowded and are not severely cost-burdened). Overcrowding is defined as having more than 1.5 people per room. Severe cost burden is defined as monthly housing costs exceeding 50% of monthly household income. These definitions are comparable to those used by *County Health Rankings & Roadmaps* in deriving its severe housing problems indicator except that whereas *County Health Rankings & Roadmaps* presents portion of households with severe housing problems, HOPE presents the inverse in line with the project's positive "opportunity" frame. The sample was restricted to households headed by adults aged 25 years or older.

Low Poverty Concentration neighborhoods are defined as those where less than 20% of residents live in poverty. Data from the 2006-2010 Neighborhood Change Database was used to identify the portion of individuals who live in low poverty concentration neighborhoods by race and ethnicity both by state and nationally. Data from the Neighborhood Change Database is not available by any measure of SES. In addition, analyses were not age-restricted and thus included the entire population.

Area Health Resources File

Access to Primary Care was defined as the portion of state populations living in counties with a ratio of 2,000 people per primary care physician or fewer, using AHRF data on the number of primary care physicians in each county. To create this indicator, counties were categorized in a binary fashion as either having a sufficient supply of primary care doctors (1) or not (0). Then, using 2010-2014 ACS county population estimates, the populations of counties having a sufficient supply were summed to the state-level and divided by the total state population for each race and ethnic and income group to arrive at the portion of each group having sufficient access to primary care physicians in each state. New Hampshire was excluded from analyses by race and ethnicity because two of its counties had highly variable AI/AN population estimates; similarly, Wyoming was excluded from rankings because one county had highly variable Asian/PI population estimates. No states were excluded from the SES group analyses.

Access to Psychiatric Care was defined as the portion of state populations (by SES) living in counties with a ratio of 30,000 people per psychiatrist or fewer, using AHRF data on the number of psychiatrists in each county. To create this indicator, counties were categorized in a binary fashion as either having a sufficient supply of psychiatrists (1) or not (0). Then, using 2010-2014 ACS county population estimates, the populations of counties having a sufficient supply were summed to the state-level and divided by the total state population for each race and ethnic and income group to arrive at the portion of each group having sufficient access to primary care physicians in each state. New Hampshire was excluded from analyses by race and ethnicity because two of its counties had highly variable AI/AN population estimates; similarly, Wyoming was excluded from rankings because one county had highly variable Asian/PI population estimates. No states were excluded from the SES group analyses.

CDC Wonder

Air Quality – Particulate Matter is defined as average daily concentration of fine particulate matter ($PM_{2.5}$) below 12 micrograms per cubic meter, a widely used standard nationally. $PM_{2.5}$ refers to fine inhalable particles, with diameters that are generally 2.5 micrometers or smaller. This data was pulled from the 2012 CDC WONDER dataset but its source is 2011 data from the Environmental Protection Agency's Air Quality System. Counties were coded as meeting national $PM_{2.5}$ standards on average in 2011 (1) or not (0). Then, using 2010-2014 ACS county population estimates, the populations of counties meeting national $PM_{2.5}$ standards were summed to the state-level and divided by the total state population for each income group to arrive at the portion of each race and ethnic (and income) group living in areas meeting air quality standards in each state. New Hampshire was excluded from race and ethnicity analyses because two of its counties had highly variable AI/AN population estimates; similarly, Wyoming was excluded from rankings because one county had highly variable Asian/PI population estimates. No states were excluded from the SES group analyses. Moreover, Alaska and Hawaii were excluded from both sets of analyses because air quality data was not collected for them.

County Business Patterns

Low Liquor Store Density is defined as counties having less than 1.736 liquor stores per 10,000 people (i.e. being at or below the 80th percentile for liquor store concentration) in a county using liquor store counts from the Census CBP. The number of liquor stores per 10,000 people was calculated for each county and the 80th percentile was identified from the distribution of all counties within the U.S. Counties below the 80th percentile (i.e. with fewer than or equal to 1.736 liquor stores per 10,000 people) were coded as (1); those above the 80th percentile were coded as (0). Then, using 2010-2014 ACS county population estimates, the populations of counties with low liquor store density were summed to the state-level and divided by the total state population for each income group to arrive at the portion of each race and ethnic (and income) group living in areas with low liquor store density in each state. New Hampshire was excluded from the race and ethnicity analyses because two of its counties had high variability across years in the estimates of their AI/AN population. Similarly, Wyoming was excluded from rankings because one county had high variability across years in its Asian/PI population estimate. No states were excluded from the SES group analyses.

Food Access Research Atlas

Food Security is measured as the portion of state population living in census tracts that are not food deserts. Specifically, the indicator was derived using USDA FARA data that identifies census tracts across the U.S. as being low income and low food access (LILA), or not. LILA census tracts are coded differently for urban and rural areas by this data source to account for the effects of population density on food retail outlet supply. A distance of greater than one-mile from the nearest food retail outlet was used to define "low access" in urban areas while a distance of 10-miles was used for census tracts designated as rural. Counties were coded as being food deserts (0) or not (1). Then, using 2010-2014 ACS census tract population estimates by race and ethnicity (and income), the populations of census tracts that were not food deserts were summed to the state-level and divided by the total state population for each income group to arrive at the portion of each race and ethnic (or SES) group living in areas that are not food deserts (i.e. are not low income and low food access). Despite our county-level population estimates indicating possible reliability issues in population estimates for certain race and ethnic groups in a couple of states, we did not attempt to address this issue with the census tract estimates. Thus, there is an underlying assumption that estimation error varies randomly around a mean of zero and is not associated with whether census tracts are LILAs or not; if this is the case, then summing less-than-reliable census tract-level estimates by race and ethnicity up to the state level should result in reliable state-level estimates of food access.

Uniform Crime Reporting Program

Low crime rate indicators were derived for **murder, assault, rape, and robbery** using data from the FBI's UCR Program. For each, the indicators measure the portion of state population living in counties at or below the 80th percentile for that county's crime rates (per 100,000 population). The same method was used to derive each indicator. Average annual crime counts per county were derived using annual counts from 2010 to 2012, the latest year data is available. The number of cases per 100,000 residents was calculated for each county and the 80th percentile was identified from the distribution of all counties within the U.S. For murder the 80th percentile was 5.1 murders/100,000; for aggravated assault, it was 283; reported rape, 36.9; and robbery, 52.1. Counties below the 80th percentile (e.g., with fewer than or equal to 5.1 murders per 100,000 residents) were coded as (1); those above the 80th percentile were coded as (0). Then, using 2010-2014 ACS county population estimates, the populations of counties with low crime rates were summed to the state-level and divided by the total state population for each race and ethnic (and income) group to arrive at the portion of each group living in areas with low murder rates in each state. As with the other physical and social environment measures, New Hampshire and Wyoming were excluded from the race and ethnicity analyses due to highly variable county populations estimates for AI/AN in two populous counties (New Hampshire) and Asian/PIs in one county (Wyoming). No states were excluded from the SES group analyses.

Analytical Methods

Assessing Reliability and Precision

The HOPE Initiative forges new ground in the assessment of racial, ethnic, and socioeconomic differences in health outcomes and their determinants. For decision-makers to find the data useful, it is paramount that the results serve as accurate state-level estimates. Thus, the following set of rules was used in assessing data prior to adopting it for analyses:

1. For survey data, there needed to be at least 50 respondents within a group. Multiple years of data were pulled where possible to reach a level where a minimal number of states were excluded from analyses due to small group sizes.

2. For survey data, where access to individual responses and/or standard errors was available, relative standard errors (RSEs) of less than 30% were required.

$$RSE = \frac{\text{Standard Error}}{\text{Estimate}}$$

   RSEs were available in cases where we had individual responses to survey data (e.g., NSCH and BRFSS) and are driven by overall group size and the portion reporting (e.g., the estimate).

3. Where sample size and/or RSEs were not available but the data source was a large, nationally-representative survey (e.g., ACS complete file), it was assumed that the U.S. Census rules for data suppression were at least, if not more, conservative than those developed for The HOPE Initiative. Thus, if the data was publicly available through the ACS FactFinder interface, the estimates met reliability and precision standards.

4. For non-survey/population data (e.g., NVSS mortality), the primary concern was the creation of an estimate that would withstand year-to-year fluctuations in death counts and population. Where that was not possible due to small population size and low rates, states were removed from the analyses. This affected nine states for infant mortality and one for low birth weight. As well, potentially unreliable results for extreme minority-minority groups (i.e. < 4,000 annual population for 25-44 year olds and/or less than 700 annual population for 45-64 year olds) in the premature mortality analyses have been indicated throughout the HOPE Measures products through the use of footnotes.

5. For indicators using county-level population estimates by race and ethnicity and SES, highly variable counties (HVCs), those with an RSE greater than 30% for one or more groups were identified and excluded from analyses where they contained HVCs and the group's county population estimate made up more than 10% of the group's total state population (Wyoming and New Hampshire). Moreover, this choice is footnoted on estimates for states where HVC population estimates make up > 35% of the state's group population estimate (Kentucky and Georgia) as being potentially unreliable.

Benchmark Setting

Because the overarching goal and framing of this project is to identify progress toward achieving equity, aspirational, yet achievable benchmarks - as evidenced by some groups already meeting them - were identified for each measure. These benchmarks are referred to as *HOPE Goals* in associated reports and data displays. In order to steer the conversation away from reinforcing "model" racial or ethnic stereotypes, benchmarks were set relative to the performance of high SES groups. For most data, SES was operationalized as educational attainment (see discussion of SES Grouping above).

The process for benchmark creation was to identify the top performing SES group in each state, rank states by those top-performing groups, identify the top five states, and take the average of their top-performing groups' scores, rates, or outcomes. In the vast majority of cases, the top performers were college graduates or those with incomes 400%+ FPL. These benchmarks were then applied to both the SES *and* race and ethnicity analyses. For the NVSS premature mortality analyses, which were segmented into two age groups (25-44 and 45-64 year olds), there are separate benchmarks for each age group using the general process described above.

In a few cases, where data was only available by race and ethnicity, the benchmark is 100% (e.g., low poverty concentration and connected youth). In both cases, The HOPE Initiative team decided that a reasonable aspirational goal was that everybody would live in neighborhoods with low poverty concentration and that all youth would either be in school or gainfully employed.

Absolute Magnitude

To identify gaps in opportunity for various groups we calculated absolute and relative magnitudes. As explained by Asada (2005), Keppel (2005), Braveman (2006), Penman-Aguilar (2016) and others, it is important to calculate and include both absolute and relative magnitudes when a focus of the project is to compare rates across groups, time and measures - all of which are potential uses of our findings.

$$AbsMag = Rate_{Group} - Rate_{Bench}$$
$$RelMag = AbsMag / Rate_{Bench}$$

For communications purposes, absolute magnitude is referred to as the *Distance to Goal*, or the progress that must be made by the nation, state, or populations to reach HOPE Goals. The *Distance to Goal* data are presented in the Appendix of *The HOPE Initiative: Data Chartbook*. Data on relative magnitudes are available by request only.

Distance to Goal

While magnitude analysis highlights differences in outcomes across groups, it does not adjust for differences in the size of these groups to describe the magnitude of the problem within a geography (states). This is the purpose of the *Distance to Goal* analyses.

Excess cases were calculated by multiplying the absolute value of the absolute magnitude (thus they are always a positive number) for states and groups where the rate is worse than the benchmark by the estimated number of people in the group of interest within a state. Where groups were already achieving or doing better than the benchmark, excess cases were set to zero.

$$\text{Excess Cases}_{\text{Group}} = \text{AbsMag} * \text{StatePop}_{\text{Group}}$$

For the age-segmented NVSS data, excess cases were calculated for each racial and ethnic (and education) group at each age range.

Excess cases, overall and for each group, are referred to in our data displays and the Appendix as the *Number to Goal* (# to Goal). This represents the number of people whose health or opportunity would need to improve for that state to meet the HOPE Goal.

The *Distance to Goal* analyses also includes a measure of *Percent to Goal* (% to Goal), which is the portion of populations within a state whose health or opportunity would need to improve for that state to meet the HOPE Goal. This is calculated by summing excess cases across all groups within a state to attain a total number of excess cases (e.g., individuals within the state not meeting the benchmark). This total number is then divided by the estimated applicable state population (e.g., adults 25 years old and older) to arrive at *excess cases as a portion of the state population*. These portions were ranked lowest to highest to rank states according to their *Distance to Goal*. The *Distance to Goal Rankings* were then divided into quartiles. States ranked in the first quartile for a particular indicator are defined as being closest to the HOPE Goal. Lower ranked states are farther from the goal, with those in the fourth or last quartile being farthest.

Results of all *Distance to Goal* calculations are available in the Appendix of *The HOPE Initiative: Data Chartbook*.

Summarizing Inequity

Distance to Goal measures summarize the magnitude or scope of indicator-related issues, but do not provide insights into the extent to which rates differ for groups (i.e. within-state equity). Based on extensive literature review, The HOPE Initiative team selected the "variation from total population rate" method for calculating level of state inequity (Asada, 2013; Harper, 2008; Keppel, 2005; Levy, 2006; Wagstaff, 1991). Other methods considered did not lend themselves to being easily understood and/or used for ordering states by level of inequity. Inequity scores were derived as follows:

1. Calculated rate for the relevant state population as a whole.

2. Calculated un-weighted variance from the state rate using the equation below:

$$\text{Un-weighted Variance} = \frac{\Sigma(X-\mu)^2}{N}$$

Where X = group rate, μ = the mean state rate, and N = the number of groups (5 or 6 in this case). The variance from the state rate is calculated for each group, summed and the total divided by the number of groups.

3. Results were standardized to facilitate cross-indicator inequity comparisons. Standardization was necessary to adjust for the differing scales used (e.g., results ranging from 0 to 1 for indicators reported as portion of the population, but rates per 1,000 used for some health outcomes).

Future phases of this work will more fully utilize the inequity scores. This brief description of the methods used is included in this document to explain the methods behind some key takeaways noted in the final report.

State Rates

For the purposes of mapping state variation in overall prevalence of each indicator, state rates were calculated as the population weighted average of individual race and ethnicity group rates. The equation below was used for all variables with the exception of two infant health variables for which the Multiracial race group was excluded.

$$\text{State rate} = \%_{\text{Black}} * \text{Rate}_{\text{Blacks}} + \%_{\text{White}} * \text{Rate}_{\text{Whites}} + \%_{\text{Hispanic}} * \text{Rate}_{\text{Hispanics}} + \%_{\text{Asian/PI}} * \text{Rate}_{\text{Asian/PI}} + \%_{\text{AI/AN}} * \text{Rate}_{\text{AI/AN}} + \%_{\text{Multi}} * \text{Rate}_{\text{Multi}}$$

The population percentages utilized were specific to the age range and race and ethnicity groups used to derive the indicator. As such, the state rates derived will not, in general, match all-population state prevalence data published elsewhere. In addition, state rates were excluded if data for one or more groups was excluded from the Distance to Goal analyses (as described above). State rates were then ranked such that a rank of 1 was assigned to the highest-performing state.

References

Asada Y., Yoshida Y., Whipp A.M. (2013). Summarizing social disparities in health. Milbank Quarterly; 91(1): 5-36.

Braveman, P., Cubbin, C., Egerter, S., Posner, S. (2005). Socioeconomic status in health research. One size does not fit all. Journal of the American Medical Association; 294: 2879-2888.

Braveman, P. (2006). Health disparities and health equity: concepts and measurement. Annual Review of Public Health; 27: 167-194.

Harper S., Lynch J., Meersman S.C., Breen N., Davis W.W., Reichman M.E. (2008). An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992–2004. American journal of epidemiology; 167(8): 889-899.

Keppel, K., Pamuk, E., Lynch, J., Carter-Pokras, O., Kim, I., Mays, V., Pearcy, J., Schoenbach, V., Weissman, J.S. Methodological issues in measuring health disparities. Vital and health statistics. Series 2, Data evaluation and methods research; (141): 1.

Levy J.I., Chemerynski S.M., Tuchmann J.L. (2006). Incorporating concepts of inequality and inequity into health benefits analysis. International journal for equity in health; 5(1):2.

Penman-Aguilar, A., Talih, M., Huang, D., Moonesinghe, R., Bouye, K., Beckles, G. (2016). Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. Journal of Public Health Management and Practice; 22: S33-S42.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. Journal of the National Medical Association; 99(9): 1013.

Wagstaff A., Paci P., Van Doorslaer E. (1991). On the measurement of inequalities in health. Social science & medicine; 33(5):545-557.