SUMMARY

This technical document outlines the methods underpinning data presented on the HOPE Initiative web site and downloadable resources. At a high level, HOPE Initiative resources aim to provide national and state-by-state information about how people from different backgrounds fare when it comes to health, well-being and quality of life. We do this by segmenting state populations by race and ethnicity (on the web site) and either educational attainment or household income levels (downloadable resources) and comparing how they fare on up to 27 indicators of health and its determinants organized by domain relative to attainable benchmarks. At the national level, because the literature suggests that associations between socioeconomic status (SES) (e.g., education and income) and health outcomes and determinants vary by race and ethnicity, we ran analyses by groups defined by both race and ethnicity *and* education groups where possible (Braveman, 2005; Shavers, 2007). The race and ethnicity analyses are presented on the web site; both sets of analyses are also available as downloadable resources. In addition to comparing how groups perform individually, we also score and rank states based on how far they are from the HOPE goal as well as on how equally opportunity is shared by residents from different backgrounds.

The following document details the following:
1. Domains
2. Data Sources
3. Indicator Development
4. Group Definitions
5. Analytical Methods including:
    a. Assessing precision / reliability
    b. Benchmark development
    c. Calculating distance to goal
    d. Measuring inequity
    e. Indicator and domain summary score creation
    f. State and national population performance categories

DOMAINS

We developed a conceptual model based largely on frameworks proposed in the National Academies of Sciences, Engineering and Medicine's 2017 report *Communities in Action: Pathways to Health Equity* funded by the Robert Wood Johnson Foundation and the World Health Organization's Conceptual Framework for Action on the Social Determinants of Health (2010). The former model proposes that various aspects of individual socioeconomic position (i.e. education, employment and income), access to quality health care, as well as the physical and social environment in which an individual resides directly and indirectly influences health outcomes. In later work published by the Robert Wood Johnson Foundation, physical and social environments were separated into two domains, and the HOPE model follows this precedent. The WHO framework proposes that socioeconomic position influences where individuals live, and thus, the social and physical environments to which they are exposed; this framework also incorporates policies, culture and societal values into the model, as well as individual behaviors, biological and psychosocial factors.

Ultimately, we organized our framework into Health Outcomes and the following domains that influence health outcomes:

- Access to Health Care
- Community and Safety Factors (eg. Social Environment)
- Physical Environment
- Social and Economic Factors

Indicators selected within each of these domains are meant to capture important aspects of the category. For example, the Access to Health Care domain includes indicators related to availability (i.e. Access to Primary Care and Access to Psychiatric Care), affordability (i.e. Health Insurance Coverage and Affordable Health Care), continuity (i.e. Dedicated Health Care Provider) and utilization (i.e. Colorectal Cancer Screening).

DATA SOURCES

Data sources for HOPE indicators include the National Survey of Children's Health, Behavioral Risk Factor Surveillance System (BRFSS), National Vital Statistics System, American Community Survey (ACS), Area Health Resources Files (AHRF), U.S. Census Bureau's County Business Patterns, U.S. Department of Agriculture's (USDA) Food Access Research Atlas. Finally, Neighborhood Change Database (NCD) and the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program.

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. While small compared to other longer-running surveys like BRFSS, the NSCH purports to yield nationally-representative and state-representative population estimates. The NSCH has been carried out four times, in 2003, 2007, 2011-2012, and 2016-2017. The 2016-2017 version of the NSCH was used for these analyses. The NSCH as a data source presents a number of problems that limits its utility for highlighting state-level racial and ethnic differences in child health outcomes and determinants. 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. An additional 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 the four standard race / ethnic groups (White, Black, Hispanic and Other) at the national level. In addition, 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. No attempts were made to impute missing income data (<10% of the sample). Response rates were sufficient at the state-level to retain all groups and states in the analyses. Through a restricted-data use agreement we were also able to create snapshots of the HOPE Measures' six race and ethnic groups and race/ethnicity-by-household income at the national level which are available via the HOPE downloadable resources.

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 (2015-2017) 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.

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 education attainment of the decedent (or the parent, in the case of child measures) is less so. In general, we used 2013-2017 data. However, for the SES analyses (i.e. analyses by educational attainment), we excluded years for individual states if they were missing greater than 20% of decedent (premature mortality) or maternal (infant mortality and low birth weight) education attainment data.

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. To create the HOPE indicators, we utilized the ACS Public Use Microdata System (PUMS). Five-year combined (2013-2017) files were the underlying data sources for many of the income domain variables.

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 2016 American Medical Association's Physician Master File.

The U.S. Census Bureau's County Business Patterns (CBP) provides subnational economic data by industry, updated annually. Data includes 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 comes from a variety of sources, including the Economic Censuses, the Annual Survey of Manufactures, 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). 2016 annual data was used for this round of analyses.

The U.S. Department of Agriculture's 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, the same year of data utilized for the original HOPE analyses.

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 years used in this analysis (2015-2017), 74 to 77% of counties provided data that was 90 to 100% complete.

Table 1. Indicator Summary

Indicator Name	Description	Data Source[1]	Population	Segments[2,3]	Benchmark
Health Outcomes					
Adult Health Status	Portion of adults reporting excellent or very good health	BRFSS, 2015-2017	Adults 25 yo+	R/E, Educ	73%
Mental Health	Portion of adults reporting 14 days or more poor mental health in the past 30	BRFSS, 2015-2017	Adults 25 yo+	Educ	5%
Child Health Status	Portion of children in excellent or very good health, as reported by parents	NSCH, 2016-2017	Children 0-17 yo	Income	98%
Premature Mortality	All-cause mortality rates for 25-44, 45-64 year olds	NVSS, 2013-2017	Adults 25-64 yo	R/E, Educ	41 / 241
Infant Mortality	Number of infant deaths (prior to first birthday) per every 1,000 live births	NVSS, 2013-2017	Infants 0-1 yo	R/E, Educ	2.5
Low Birthweight	Portion of live births that are low birth weight	NVSS, 2013-2017	Newborns	R/E, Educ	5%
Social and Economic Factors					
Livable Income	Portion of adults with income greater than 250% of FPL	ACS, 2013-2017	Adults 25yo+	R/E, Educ	88%
Affordable Housing	Portion of households spending ≤ 30% of monthly household income on housing and related expenses	ACS, 2013-2017	Households headed by adult 25yo+	R/E, Educ	87%
Post-Secondary Education	Portion of adults with post-high school education	ACS, 2013-2017	Adults 25 yo+	R/E, Income	83%
Youth in School or Working	Portion of 16-24 year olds in school or working, including military service	ACS, 2013-2017	Youth 16-24 yo	R/E	100%
Preschool Enrollment	Portion of 3-4 year olds enrolled in preschool	ACS, 2013-2017	Children 3-4 yo	Income	73%
Employment	Portion employed	ACS, 2013-2017	Individuals aged 16 yo+	R/E, Income	99%
Access to Health Care					
Access to Primary Care	Portion living in a county with fewer than 2000:1 population to primary care physician ratio	AHRF, 2016	All	R/E, Income	100%
Access to Psychiatric Care	Portion living in a county with fewer than 3000:1 population to psychiatrist ratio	AHRF, 2016	All	R/E, Income	100%

Health Insurance Coverage	Portion under 65 years with health insurance (any kind)	ACS, 2013-2017	Ages 0 - 64 yo	R/E, Income	98%
Affordable Health Care	Portion reporting they did not delay care due to cost in the past year	BRFSS, 2015-2017	Adults 25 yo+	R/E, Educ	96%
Dedicated Health Care Provider	Portion reporting they have usual source of care	BRFSS, 2015-2017	Adults 25 yo+	R/E, Educ	92%
Colorectal Cancer Screening	Portion of adults getting recommended colorectal screenings	BRFSS, 2015-2016	Adults 50-75 yo	Educ	81%
Physical Environment					
Home Ownership	Portion of households owning home	ACS, 2013-2017	Households headed by adult 25yo+	R/E, Educ	81%
Housing Quality	Portion not living in houses with severe housing problems	ACS, 2013-2017	Households headed by adult 25yo+	R/E, Educ	92%
Low Liquor Store Density	Portion not living in a county that in is the 80th percentile or higher in terms of the number of liquor stores per 10,000 population	U.S. Census Bureau County Business Patterns, 2016	All	R/E, Income	100%
Food Security	Portion of the population living in census tracts that are not food deserts (i.e. not designated low food access and low income)	USDA Food Access Research Atlas, 2015	All	R/E, Income	97%
Community and Safety Factors					
Low Poverty Concentration	Portion of the population living in low-poverty concentration neighborhoods	ACS, 2010 from the Neighborhood Change Database	All	R/E	100%
Low Homicide Rate	Portion living in counties with fewer than 5.1 murders per 100,000 persons annually	UCR, 2015-2017	All	R/E, Income	100%
Low Physical Assault Rate	Portion living in counties with fewer than 283 cases of aggravated assault per 100,000 persons annually	UCR, 2015-2017	All	R/E, Income	100%
Low Sexual Assault Rate	Portion living in counties with fewer than 36.9 cases of	UCR, 2015-2017	All	R/E, Income	100%

	reported rape per 100,000 persons annually				
Low Robbery Rate	Portion living in counties with fewer than 52.1 cases of robbery per 100,000 persons annually	UCR, 2015-2017	All	R/E, Income	100%

[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; NSCH – National Survey of Children's Health; NVSS - U.S. Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System; 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 / ethnicity; Educ – four groups defined by educational attainment; Income – five groups defined by household income as a percent of the federal poverty level. R/E results are presented on the HOPE web site. Results by education and income (and nationally by R/E & SES combined) can be found in the downloadable resources.*

[3]*Indicators only being available by Educ or Income at the state level are available at the national level by R/E, Educ or Income, and R/E x SES.*

INDICATOR DEVELOPMENT

HEALTH OUTCOMES

Adult Health Status is defined as the portion of adults reporting excellent or very good health. It was created using a BRFSS item that asks respondents to rate their health as excellent, very good, good, fair or poor. We dichotomized responses as excellent or very good health vs. good, fair or poor health.

Child Health Status is defined as the portion of children in excellent or very good health, as reported by parents. It is derived from the NSCH item that asks parent / adult respondents to rate their child's health as excellent, very good, good, fair or poor. Due to NSCH data suppression criteria related to race and ethnicity, this variable is only presented for SES (household income) groups at the state level. Results for race and ethnic groups are presented at the national level. Due to changes in the NSCH sampling methodology, the current findings should not be compared to those from the original round, which used 2011-2012 data.

Infant Mortality is defined as, for every 1,000 live births, the number of infants who die before their first birthday. It is derived from birth and death certificate data that is submitted by states to the NVSS and made available as linked birth-death files. Rhode Island was excluded from the education analyses due to missing more than 20% of maternal education data for all but one year. For New Jersey and Connecticut only 2016 and 2017 data were used for the analyses by maternal education attainment due to missing more than 20% of relevant education data for the first three years. Finally, for seven other states (Alaska, Alabama, Arkansas, Arizona, Hawaii, Maine and West Virginia) 2013 data was excluded from rate calculations by education attainment due to not meeting the less than 20% missing education data threshold.

Low Birthweight is defined as the portion of live births for which infants weighed 2,499 grams or less at birth, and is derived from birth certificate data submitted by states to the NVSS. As with the infant mortality indicator, missing maternal education was a problem in some states and some years and resulted in less than five years of data being utilized for these states. Specifically, only 2016 and 2017 data could be used for New Jersey and Connecticut; 2015 – 2017 data for Rhode Island; and 2014-2017 data for Alabama, Arkansas, Arizona, Hawaii, Maine and West Virginia.

The Mental Health item is defined as the portion of adults reporting 14 or more days of poor mental health in the past 30 days. It is derived from a BRFSS item that 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?" Due to large relative standard errors for some race and ethnic groups in many states (indicative of potentially unreliable estimates), results for this indicator are presented only by education attainment.

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/ethnic and education groups within states) varies, mortality data was aggregated across 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 across five years (2013 through 2017) 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/ethnic) group-specific mortality rates. No data was missing with respect to race and ethnicity; education data was missing from less than 10% of records except for Rhode Island which was missing over 45% of education data for both age groups and New York which was missing over 10% for the 45-64 age group. These two states were excluded from the education attainment group analyses.

SOCIAL & ECONOMIC FACTORS

Affordable Housing is defined as the percent of those spending 30% or less of monthly household income on rent, mortgage, and housing-related expenses. It should be noted that this last 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. This indicator is derived from ACS data.

Employment rates are 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. ACS is the data source used to create this indicator.

Livable Income is defined as the portion of adults 25 years old or older who earn income above 250% FPL (using the U.S. Census FPL definitions). This indicator was calculated based on reported household income and number of household occupants using ACS data.

Post-Secondary Education is 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. ACS data is the data source used.

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

Youth in School or Working is defined as the portion 16 to 24 year-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) this 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).

ACCESS TO HEALTH CARE

Access to Primary Care is defined as the portion of (grouped) 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 binary fashion as either having a sufficient supply of primary care doctors (1) or not (0). Then, using 2013-2017 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, ethnic, and income group to arrive at the portion of each group having sufficient access to primary care physicians in each state.

Access to Psychiatric Care is defined as the portion of (grouped) state populations 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 binary fashion as either having a sufficient supply of psychiatrists (1) or not (0). Then, using 2013-2017 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, ethnic, and income group to arrive at the portion of each group having sufficient access to psychiatric care in each state.

Affordable Health Care is defined as the portion of adults who did not delay or forego any necessary medical care due to cost in the past year. It is derived from BRFSS data. Of note, the denominator for this rate's calculation 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).

Colorectal Cancer Screening is defined as the percentage of adults aged 50 to 75 who report having received colorectal cancer screening per USPSTF guidelines in place between 2008 and 2016. It is derived using several BRFSS variables (*bldstool, lstblds3, hadsigm3, lastsig3*) related to home blood stool test, sigmoidoscopy and colonoscopy administration. Due to the narrow age range used in creating this variable and the fact that this data is only collected for all states in even years (with a small portion of the sample coming from 10 states that collected this data in 2015), sample sizes for certain minority race and ethnic groups in many states were too small to create reliable estimates. Thus, findings for this indicator are provided only by education attainment (in the downloadable resources).

Dedicated Health Care Provider is defined as the percentage of adults who report having one or more people they think of as their personal doctor or health care provider and is derived from the BRFSS variable.

Health Insurance Coverage is defined as the portion of populations under the age of 65 who report having any kind of health insurance coverage (eg. public, private or mixed). It is calculated using ACS data.

PHYSICAL ENVIRONMENT

Food Security is defined as the portion of state populations living in census tracts that are not food deserts. Specifically, the indicator was derived using 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 2013-2017 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 group to arrive at the portion of each group living in areas that are not food deserts (i.e. are not low income and low food access). We acknowledge that census track population estimates by race/ethnicity or education are potentially unreliable but did not attempt to address it. Thus, we made 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.

Home Ownership is defined as the portion of households living in owner-occupied homes using ACS data. The samples were restricted to households headed by an adult who was 25 years old or older.

Housing Quality is the 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 burdensome) using ACS data. 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 the *County Health Rankings* in deriving their severe housing problems indicator except that whereas they present portion of households *with* severe housing problems, we present the inverse in line with a positive opportunity-oriented frame. The samples are restricted to households headed by adults aged 25 years or older.

Low Liquor Store Density is the portion of (grouped) state populations living in counties having less than 1.77 liquor stores per 10,000 people (i.e. being at or below the 80th percentile for liquor store concentration) using liquor store counts from the 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 nationwide. Counties below the 80th percentile (i.e. with fewer than or equal to 1.77 liquor stores per 10,000 people) were coded as 1; those above the 80th percentile were coded as 0. Then, using 2013-2017 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.

COMMUNITY & SAFETY FACTORS

Low Homicide Rate is defined as the portion of (grouped) state populations living in counties having less than 6.02 homicides per 100,000 people (i.e. the 80th percentile for county homicide rates) using data from the UCR files. The number of homicides 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. Counties below the 80th percentile were coded as 1; those at or above the 80th percentile were coded as 0. Then, using 2013-2017 ACS county population estimates by race/ethnicity and income, respectively, the populations of counties with codes as 1 were summed to the state-level and divided by the total state population for each race/ ethnic (and income) group to arrive at the portion of each group living in areas with low homicide rates in each state.

Low Physical Assault Rate is defined as the portion of (grouped) state populations living in counties having less than 308.8 reported incidents of aggravated assault per 100,000 people (i.e. the 80th percentile for county aggravated assault rates) using data from the UCR files. 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. Counties below the 80th percentile were coded as 1; those at or above the 80th percentile were coded as 0. Then, using 2013-2017 ACS county population estimates by race/ethnicity and income, respectively, the populations of counties with codes as 1 were summed to the state-level and divided by the total

state population for each race/ ethnic (and income) group to arrive at the portion of each group living in areas with low physical assault rates in each state.

Low Poverty Concentration is defined as the portion of (grouped) state populations living in neighborhoods in which less than 20% of residents live in poverty. We used most recently available data from the 2006-2010 NCD to identify the portion of individuals who live in low poverty concentration neighborhoods by race and ethnicity by state and nationally. Data from the NCD is not available by any measure of SES.

Low Robbery Rate is defined as the portion of (grouped) state populations living in counties having less than 51.07 reported robberies per 100,000 people (i.e. the 80th percentile for county robbery rates) using data from the UCR files. 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. Counties below the 80th percentile were coded as 1; those at or above the 80th percentile were coded as 0. Then, using 2013-2017 ACS county population estimates by race/ethnicity and income, respectively, the populations of counties with codes as 1 were summed to the state-level and divided by the total state population for each race/ ethnic (and income) group to arrive at the portion of each group living in areas with low robbery rates in each state.

Low Sexual Assault Rate is defined as the portion of (grouped) state populations living in counties having less than 56.26 reported incidents of rape per 100,000 people (i.e. the 80th percentile for county rape rates) using data from the UCR files. 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. Counties below the 80th percentile were coded as 1; those at or above the 80th percentile were coded as 0. Then, using 2013-2017 ACS county population estimates by race/ethnicity and income, respectively, the populations of counties with codes as 1 were summed to the state-level and divided by the total state population for each race/ ethnic (and income) group to arrive at the portion of each group living in areas with low sexual assault rates in each state.

GROUP DEFINITIONS

Our decision to create six mutually-exclusive racial and ethnic groups balances a desire for discrete, homogenous groups with a practical need to create estimates that are reliable and representative of the populations to which they pertain. These groups include: Whites, Blacks, Hispanics, Asians and Pacific Islanders (Asian/PI), American Indians and Alaska Natives (AI/AN), and Multiple Race. The Other Race category was excluded due to low numbers. While using six groups (rather than the frequently used four - Black, White, Hispanic, Other) resulted in our excluding some groups who were extreme minorities in a few states from some of the analyses due to estimate reliability concerns, HOPE Initiative leadership, including its National Advisory Committee, embraced the tradeoff in order to be able to represent what was happening within relatively discrete groups across most states. We recognize that our six groups are not homogenous within themselves and that the life experiences of group members do vary. Unfortunately, given the practical considerations of data availability, creating more granular groups was not feasible at the state-level. Data by race and ethnicity is presented on the HOPE web site.

When data about socioeconomic status was available, we grouped by either education attainment or household income as a percent of the federal poverty level (FPL). The education attainment categories were: Less Than High School Education, High School Graduation, Some College (including technical school and associate degrees), College Graduate and Over. Income

categories were: < 100%, 100-199%, 200-299%, 300-399%, and ≥ 400% FPL. <u>As data were more frequently available for education attainment, and income data tends to be subject to higher non-random non-response rates, in most cases we defaulted to using education attainment as our SES grouping variable when both data points were available</u>. Because education attainment is somewhat dependent upon age, many of our indicators were developed restricting data to adults ages 25 and older; we retained the age restrictions for the race and ethnicity group analyses for measures that used educational attainment as the SES grouping variable to facilitate cross-group comparisons.

ANALYTICAL METHODS

Assessing Reliability and Precision

The HOPE project is forging new ground in the assessment of race, ethnic, education and income group differences in health outcomes and their determinants. Because we want decision-makers to find the data useful, it is paramount that we be able to stand behind our results as accurate state-level estimates. Thus, we developed a set of criteria results must meet in order to be included on our web site or downloadable resources. These criteria are:

1. For survey data, there needed to be at least 50 respondents within a group. We combined up to five years of data where possible to reach a level at which a minimal number of states were excluded due to small group sizes.
2. Additionally, for survey data where we had access to individual responses and / or standard errors, relative standard errors of estimates have to be less than 30%.

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

   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. For non-survey/population data sources (e.g., NVSS), our primary concern was the creation of an estimate that would withstand year-to-year fluctuations in death or low birthweight counts and population. Where that was not possible due to small population size and low rates, we removed states from the analyses. Because cases of low birth weight and infant mortality are relatively rare events, we set thresholds for the total number of live births over the five-year period (minimum of 40 for low birth weight and 400 for the infant mortality) a group needed to have at the state level in order for its rate to be considered reliable. This affected the race and ethnicity analyses only. Similarly, for premature mortality, annual population estimates for 25-44 year-olds and 45-64 year-olds needed to be at least 4,000 and 700, respectively, in order for group estimates to be deemed reliable. Potentially unreliable estimates were excluded.

Benchmark Development

Because the overarching goal and framing of this project is to identify gaps in "opportunity," 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* on the web site and in the downloadable resources. In order to steer the conversation away from the

"model minority" theme, we set benchmarks relative to the performance of high SES groups. For most of our data, SES was operationalized as educational attainment as discussed in the Group Definitions section 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 the College Graduate or 400%+ FPL groups. 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), we created separate HOPE Goals for each age group using the same process.

In a few cases, where data was only available by race and ethnicity, we set 100% as the benchmark (i.e. Low Poverty Concentration, and Youth in School or Working). In both cases, the 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.

Calculating Distance to Goal

Distance to Goal (DTG) shows how much progress the nation and each state must make to achieve the HOPE Goal on a particular indicator, overall and by race and ethnicity *and* SES. DTG is presented in two ways: (a) the absolute number of people whose status would need to improve in order for a group or state as a whole to meet the HOPE Goal, and (b) the portion of the group or state population that would need to improve its status for the HOPE Goal to be met. The first, DTG_Num is generally calculated as:

$$\text{DTG\_Num}_{Group} = (\text{Rate}_{Benchmark} - \text{Rate}_{Group}) * \text{Pop}_{Group}$$

where the Group may be a race/ethnic group at the state level or at the national level. Where indicators are presented as something other than percentages (e.g., Infant Mortality which is the rate per 1,000 live births), the equation is adjusted to turn the difference between the benchmark and group rates into a percentage. The DTG_Num for all groups are summed to arrive at the DTG_Num for the state as a whole.

The second, Percent Distance to Goal (DTG_%) is calculated at the group level simply as the difference between the HOPE Goal and the group's Rate. DTG_% at the state level is calculated by dividing the DTG_Num$_{State}$ by the applicable state population (e.g., adults 25 years and older).

Summarizing Inequity

DTG measures summarize the magnitude / scope of indicator-related issues but do not provide insights into the extent to which rates differ for groups (i.e. within-state inequity). Based on extensive literature review, the HOPE Measures 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).

4. Finally, inequity categories were created using the standardized scores. Where the mean and median z-scores differed by less than 0.10, states were categorized as follows:

   a. Less than -1.0: Very Low Inequity
   b. -1.0 to -0.2499: Below Average Inequity
   c. -0.25 to 0.2499: Average Inequity
   d. 0.25 to 1.0: Above Average Inequity
   e. Greater than 1.0: High Inequity

When means and medians differed by more than 0.10, these ranges were adjusted by the median (eg. if the mean=0 but the median was -0.20, states would be included in the very low inequity category if they had z-scores less than -1.20). Doing this effectively adjusts for the effects of outliers on the mean.

The standardized inequity scores were used in creating state summary scores as described in the following section. The equity categories (and DTG categories created using similar methods) were used on the graphics on the State pages associated with the question, "How does [state] compare with other states?

Indicator and Domain Summary Score Creation

Indicator summary scores were created for all states. The process was as follows:

1. Create inequity z-scores as described above.
2. Created DTG z-scores using the same process described above for standardizing inequity scores but using the Percent DTG results (described in the Distance to Goal section).
3. Averaged the DTG and Inequity z-scores.
4. Rescaled the results to vary from 0 to 100 (ideally the mean is 50, but this is not always the case due to outliers at one or both ends of the spectrum).
5. In order to create indicator scores where numbers closer to 100 are better (very low inequity and DTG) and those closer to 0 are less favorable, we subtracted the total in #4 from 100 to arrive at an indicator summary score that incorporates both DTG and Inequity.

Domain summary scores were created by averaging the state indicator summary scores for all indicators within the domain. Where states are missing scores for one or more indicators due to rate suppression, scores were averaged for available indicators.

State and National Population Performance Categories

States are shaded on the population cartogram maps based on how individual groups perform relative to ALL groups across all states on Percent Distance to Goal. Specifically, groups were categorized into deciles which match to particular shades with more saturated shades signifying lower levels of inequity / less distance to the HOPE goal, and less saturated shades signifying greater inequity / further distance from the HOPE goal.

These same deciles are used for applying shading to groups at a national level on the national domain summary tables / graphics.

References

Asada Y, Yoshida Y, Whipp AM (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 SC, Breen N, Davis WW, Reichman ME (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., . . . Weissman, J. S. (2005). Methodological issues in measuring health disparities. Vital and health statistics. Series 2, Data evaluation and methods research; (141): 1.

Levy JI, Chemerynski SM, Tuchmann JL (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.