An Information Theory Based Framework for the Measurement of Population Health

Erik T. Nesson*
Joshua J. Robinson†

Abstract

This paper proposes a new framework for the measurement of population health and the ranking of the health of different geographies. Since population health is a latent variable, studies which measure and rank the health of different geographies must aggregate observable health attributes into one summary measure. We show that the methods used in nearly all the literature to date implicitly assume that all attributes are infinitely substitutable. Our method, based on the measurement of multidimensional welfare and inequality, minimizes the entropic distance between the summary measure of population health and the distribution of the underlying attributes. This summary function coincides with the constant elasticity of substitution and Cobb-Douglas production functions and naturally allows different assumptions regarding attribute substitutability or complementarity. To compare methodologies, we examine a well-known ranking of the population health of U.S. states, America's Health Rankings. We find that states' rankings are somewhat sensitive to changes in the weight given to each attribute, but very sensitive to changes in aggregation methodology. Our results have broad implications for well-known health rankings such as the 2000 World Health Report, as well as other measurements of population and individual health levels and the measurement and decomposition of health inequality.

Keywords: Population Health, Health Rankings, Information Theory, Entropy
JEL Codes: C43, I12, I18

*Corresponding author. Department of Economics, Miller College of Business, Ball State University, Muncie, IN (E-Mail: etnesson@bsu.edu)

†Department of Marketing, Industrial Distribution, and Economics, Collat School of Business, University of Alabama at Birmingham, Birmingham, AL (E-Mail: jjr@uab.edu)

1 Introduction

Since the passing of the Patient Protection and Affordable Care Act in 2010, expanded government involvement in the health care sector has increased the ability of policy makers to influence the health outcomes of the populations they represent. However, the efficient allocation of public resources requires robust measures of the costs and benefits associated with policy. Policy makers often use rankings and other measurements of geographies' population health, such as the Robert Wood Johnson Foundation's *County Health Rankings and Roadmaps*, the Commonwealth Fund's *Health System Scorecards*, and the United Health Foundation's *America's Health Rankings*, in designing public policies (Erwin et al., 2008). In this paper, we show that implicit assumptions embedded in these popular metrics could result in misleading evaluations of health, and we describe an alternative framework that is more flexible and transparent.

Because the health of a population is a latent characteristic, these rankings, as well as many other multidimensional health measurements, aggregate health attributes into a measure of latent health using a weighted arithmetic mean, where the weights are chosen on a normative basis. These methods require assumptions about the relative importance of the attributes used and the relationships between attributes, and the methodology often masks the nature of these assumptions. For example, normatively-chosen weights can place unintended emphasis on highly-correlated dimensions of health. More importantly, linearly aggregating attributes using a weighted arithmetic mean implicitly assumes that the attributes are infinitely substitutable, where the marginal rate of substitution between any two attributes is constant, completely determined by the attribute weights, and independent of the level of each attribute.

We propose a methodology to measure and rank population health based on the concept of multivariate generalized entropy (MGE). Originally developed by Maasoumi (1986), our methodology chooses a summary measure of population health that minimizes the entropic distance between the summary measure and the multivariate distribution of underlying attributes. This preserves as much information as possible from the underlying attribute distribution when constructing the summary measure of population health. While not applied to the measurement of health, MGE has been widely used to measure economic welfare, inequality, and poverty (e.g. Hirschberg et al., 1991, 2001; Maasoumi and Nickelsburg, 1988; Maasoumi and Jeong, 1985; Lugo, 2007; Maasoumi and Lugo, 2008; Decancq and Lugo, 2013).

Unlike the weighted arithmetic mean approach currently used by health rankings, our methodology makes assumptions transparent, changeable, and comparable. The MGE summary functions coincide with the functional forms of popular utility function and production functions. This allows researchers to transparently modify the relative importance of each attribute, the substitutability or complementarity between the attributes, and compare how different weighting methodologies and social preferences change the measurement of population health. Lastly, as we show, the weighted arithmetic mean approach taken by most health rankings is a special, extreme case of our methodology, when health attributes are assumed to be infinitely substitutable.

To create a basis for comparison, we utilize data from a well-known ranking of the health status of U.S. states, America's Health Rankings, which aggregates 24 measures of health to a measure of population health. We find that states' health rankings are somewhat sensitive to changes in the weighting methodology but very sensitive to changes in aggregation methodology. As we move away from the implicit assumption of infinite substitutability in the original America's Health Rankings methodology towards a more complementary relationship between the different health attributes, the correlation coefficient between the original and new rankings falls to below 0.6. The rankings of some Southern states traditionally considered unhealthy improve dramatically while the rankings of many Midwestern states typically regarded as being healthy fall significantly. Wealthy New England states typically remain near the top of the rankings and states commonly considered to be exceptionally unhealthy, like Mississippi, remain poorly ranked. Thus, while linear rankings may accurately describe extreme parts of the population health distribution, they may not accurately characterize other aspects of the distribution. These results demonstrate the advantages of our MGE-based method, which allows for straightforward sensitivity analyses of the aggregation assumptions.

The rest of this paper proceeds as follows. Section 2 reviews different studies that rank health systems and population health. Section 3 describes the general entropy aggregation methodology and its uses in the measurement of economic welfare. Section 4 describes the data source and specifics of the America's Health Rankings methodology. Section 5 summarizes our results, and Section 6 concludes.

2 Background

We focus on the use of health rankings rather than the constructed summary measure itself, although the points we make remain relevant even outside the context of rankings.[1] A major advantage of rankings is that they provide a unit-free, relative metric that makes a complex set of information much easier to understand. We are only able to measure latent characteristics, like health, by using constructed values for which magnitudes have little intuitive meaning. This makes such metrics difficult for policy makers and researchers to understand and utilize. Transforming such measures into rankings provides context by comparing observable conditions of nature.

One of the most well-known rankings is the World Health Organization's 2000 World Health Report on health care system performance, which ranks the health systems of all World Health Organization member countries (World Health Organization, 2000). More recently, the Commonwealth Fund has published a series of reports aimed specifically at assessing the relative performance of the United States health care system. These reports compare the United States to five other, mostly English speaking, countries (Davis and Fund, 2004, 2007; Davis et al., 2010). Within the United States, attention often focuses on ranking the population health of states or counties (Kindig et al., 2008; Kindig and Stoddart, 2003; Erwin et al., 2008; Booske et al., 2010; Kanarek et al., 2011; Peppard et al., 2008). These reports often get very high profile coverage, especially with recent reforms of the United States health care system. For example, the often-cited statistic in the 2000

[1]For example, quality adjusted life expectancy (QALE) is another metric commonly used to assess population health in a multidimensional fashion (see Stewart et al., 2013, for a recent example). A detailed review and comparison of QALE methods with our methodology is beyond the scope of this paper. However, like health rankings, QALE metrics essentially calculate a weighted arithmetic mean. As we detail below, the primary advantage of the MGE-based metrics we propose is a lack of dependency on linear functions to summarize the different dimensions of health. Thus, many of our general critiques of existing health rankings also apply to QALE metrics.

World Health Report that the United States health system ranks 37^{th} in the world, between Costa Rica and Slovenia, was covered in the New York Times, Associated Press, and USA Today, and is often mentioned in other popular media, such as the film *Sicko* (World Health Organization, 2000; Hilts, 2000; Neergaard, 2000; Rubin, 2000). Within the United States, rankings of states are also covered in the national press (Rubin, 2009; Bivins, 2009).

The use of multiple attributes requires the summarization of this information by dimension reduction, and whenever the number of dimensions is reduced information is lost. Thus, depending on the aggregation method, valuable information could be missing from the aggregation metric. Nearly all existing health rankings reduce the number of dimensions by calculating the weighted arithmetic mean of the attributes, which implicitly assumes an infinite degree of substitutability between the different attributes. In the context of a health ranking, this assumption implies that, for example, a rise in the cardiovascular death rate in a state can be offset by a certain level increase in the health insurance rate, and that the amount of this tradeoff is independent of the cardiovascular death rate or health insurance rate. If society exhibits any degree of diminishing marginal utility, then this assumption does not represent society's preferences. Furthermore, a world where insurance rates are increasing at a rate proportional to rate of cardiovascular deaths may be more disconcerting than one with an increase in cardiovascular deaths alone. Yet, the linear aggregator would treat such an event neutrally.

More recently, there have been methodological and theoretical improvements to the health ranking literature. For example, treating health as a latent variable that is correlated with observable health measures, Courtemanche et al. (2013) utilize a Bayesian hierarchical model to identify a county's relative rank and assess the overall health of counties in Texas and Wisconsin. They improve on the traditional factor analysis approach by allowing factor weights to depend on spacial correlations across counties. Additionally the Bayesian approach naturally lends to accounting for uncertainty in the ranking, which they further improve by accounting for the uncertainty created through imputing missing values. Although much more sophisticated than previous approaches, the final estimated rankings still depend on what is fundamentally a linear combination of observed attributes, which implicitly assumes perfect substitutability.

Another recent paper, Makdissi and Yazbeck (2014), addresses the issue of combining multiple attributes of health into a single index in the context of measuring inequalities. They show that measures that are not ratio-scale cannot be aggregated by common index procedures because the rankings are not robust to monotonic transformations of the original values.[2] They propose counting the number of health dimensions for which health is below a certain threshold and derive inequality measures and stochastic dominance procedures for these modified indices. Nonetheless, their proposed aggregation method is a linear transformation and, thus, also assumes perfect substitutability.

While making strong contributions to the literature, these papers—like most of the health rankings and measurement literature—focus only on the methods of weighting and scaling variables, leaving the issue of substitutability between health attributes largely unexplored. We show that allowing for some degree of complementarity can have a large effect on health rankings, and that the choice of summary function can have a more dramatic effect on the resulting rankings than the choice of weighting or scaling method. In the next section, we introduce the theory of entropy-based aggregation and further explore issues related to weighting.

3 Methods

In this paper, we focus on entropy-based aggregation methods, which are advantageous both for their desirable information-preserving properties and their intuitive link to the economic theories of production and utility. The concept of entropy comes from the field of information theory and is a measure of the average uncertainty of a random variable. Intuitively, entropy is indicative of how much information is needed to describe a random variable. A random variable has a minimum level of entropy when one event in its sample space is certain and has a maximum level of entropy when all events in the sample space are equally likely.

[2] A cardinal number for which zero indicates the absence of value is ratio scale. Measures that are not ratio scale include nominal numbers, ordinal numbers, and cardinal numbers for which zero is not meaningful, like temperature.

Relative entropy describes the information lost by mischaracterizing the probability distribution of a random variable. If $p(x)$ is the true distribution of a random variable, relative entropy, $D(p||q)$ is a measure of the inefficiency, or information loss, that occurs when the random variable is instead represented by some other distribution $q(x)$:

$$D(p||q) = \sum_{x \in X} p(x) \log \left(\frac{p(x)}{q(x)} \right) \tag{1}$$

This measure, also known as Kullback-Leibler distance, can be thought of as the distance between the distributions $p(x)$ and $q(x)$ (Cover and Thomas, 2006).[3] Theil (1967) is the first to use relative entropy in economic welfare analysis. He uses the entropic distance between a distribution of a population's individual income shares and the uniform distribution of income as a measure of inequality. Several variations on Theil's measures have since been proposed in the literature (see Decancq and Lugo (2013) for a literature review).

Maasoumi (1986) extends the inequality measures proposed by Theil to a MGE-based measure of multidimensional inequality. In addition to using relative entropy to evaluate inequality, he proposes utilizing a generalized formulation of relative entropy for the aggregation of multiple attributes of well-being into an index measure. He proposes choosing a summary function to minimize the entropic distance between the univariate distribution of the summary function and the original multivariate distribution. This ensures that the summary function will preserve as much information from the original data as possible. Rather than use the calculated summary functions to evaluate inequality like Maasoumi (1986), we take the utility function interpretation of the aggregator one step further and directly utilize the resulting ordinal ranking it produces to rank population health.

Formally, if X_{if} is the value of some attribute, $f = 1 \ldots M$, for an observation, $i = 1 \ldots N$, then $X_i = (X_{i1}, X_{i2}, \ldots, X_{iM})$ is the row vector of values for all attributes for observation i and $X^f = (X_{1f}, X_{2f}, \ldots, X_{Nf})$ if the column vector of values of attribute f for all observations. We

[3] Relative entropy is not technically a true distance measure in the mathematical sense because it fails the triangle inequality and is not symmetric.

regard X^f as the sample distribution of an attribute.[4] The optimal summary function transforms the M-vector of attributes into a single value for each observation, $S_i = h(X_i)$, in such a way as to minimize the generalized multivariate relative entropy between S_i and the multivariate distribution of X^f's. That is, S_i minimizes the function:

$$
\begin{aligned}
D_\beta(S, X; \alpha) &= \sum_{f=1}^{M} \alpha_f \left\{ \sum_{i=1}^{N} S_i \left[\left(\frac{S_i}{X_{if}} \right)^{\beta} - 1 \right] \Big/ \beta(\beta+1) \right\} \\
&= \sum_f \alpha_f \left\{ \sum_i S_i \log(S_i / X_{if}) \right\}, \quad \text{if } \beta = 0 \qquad (2) \\
&= \sum_f \alpha_f \left\{ \sum_i X_{if} \log(X_{if} / S_i) \right\}, \quad \text{if } \beta = -1
\end{aligned}
$$

This is essentially a weighted sum of the divergence of S_i from the set of corresponding X_i. Here the α's represent the relative importance (or weight) assigned to each attribute. Maasoumi (1986) shows that the summary functions that minimize Equation (2) are

$$
\begin{aligned}
S_i &\propto \left[\sum_{f=1}^{M} \delta_f X_{if}^{-\beta} \right]^{-1/\beta} \\
&= \prod_{f=1}^{M} X_{if}^{\delta_f}, \quad \text{if } \beta = 0 \qquad (3) \\
&= \sum_{f=1}^{M} \delta_f X_{if}, \quad \text{if } \beta = -1
\end{aligned}
$$

where $\delta_f = {}^{\alpha_f}/_{\sum_f \alpha_f}$ is each attribute's relative weight. By construction, these summary functions are as close to the original multivariate distribution as possible. The functions in Equation (3) are the weighted harmonic mean (for $\beta \neq 0, -1$), the weighted geometric mean (for $\beta = 0$), and the weighted arithmetic mean for (for $\beta = -1$), respectively. Thus, the methodology adopted by current health rankings are synonymous to choosing an entropic summary function with $\beta = -1$.

[4]The values, X_{if}, are standardized such that their distributions have the same support.

This method of dimension reduction has been widely used in the economic literature examining multidimensional well-being. Hirschberg et al. (1991) and Hirschberg et al. (2001) consider clusters of attributes in well-being and measure quality of life in the United States and across countries, Maasoumi and Zandvakili (1990) develop a variation of the entropic aggregation and inequality measurement function to estimate income mobility, and Maasoumi and Trede (2001) apply this technique to comparing income mobility in the United States and Germany. An extension of the entropic aggregation method has been applied to measuring inequality. Maasoumi and Jeong (1985) estimate trends in world inequality measures, Maasoumi and Nickelsburg (1988) analyze trends in inequality in Michigan, and more recently researchers have used these and similar methods to examine inequality and poverty in a number of settings (Justino et al., 2005; Lugo, 2007; Brandolini, 2008; Decancq et al., 2009; Decancq and Ooghe, 2010; Maasoumi and Lugo, 2008; Lugo and Maasoumi, 2009).

The generality and transparency of the methodology allows for a straightforward characterization of three persistent problems in the literature on multidimensional well-being and dimension reduction: how to evaluate different bundles of attributes, how to account for the relative importance of an attribute compared to the others, and how to handle variables with different units of measurement. In the following sections, we discuss each of these issues in greater detail.

3.1 Complementarity, Substitutability, and the Value of β

The value of the summary functions in Equation (3) depend on the choices of β and the vector of relative weights. A major advantage of this framework is the characterization of the parameter β. The general form of S_i ($\beta \neq 0, -1$) has the same functional form as the constant elasticity of substitution utility (or production) function, the second form ($\beta = 0$) is synonymous with a Cobb-Douglas function, and the third form ($\beta = -1$) is a perfect substitutes function. Thus $\sigma = {}^{1}/_{(1+\beta)}$ is the constant elasticity of substitution between attributes across individuals.

With this characterization in mind, one can think of changes in β as altering the degree of substitutability or complementarity between attributes. A β less than zero implies a higher degree of substitutability between attributes, and as $\beta \to -1$, the elasticity of substitution approaches infinity and attributes are combined as though they are perfectly substitutable in the calculation of health. On the other hand, $\beta > 0$ implies a degree of complementarity between attributes. Attributes are considered approaching perfect complements, a preference for all attributes to rise in perfect proportion to one another, as $\sigma \to 0$ and $\beta \to \infty$. That is, the larger the value of β, the more complementarity is assumed between the attributes.

To formally illustrate the importance of the parameter β, consider the marginal rate of substitution (MRS) between any two attributes l and m. The MRS indicates the amount of attribute m that must be given up for a one unit increase of attribute l to preserve the value of S_i,

$$MRS_{lm} = \frac{\partial S_i(X)/\partial X_l}{\partial S_i(X)/\partial X_m} = \left[\frac{\alpha_l}{\alpha_m}\right]\left[\frac{X_m}{X_l}\right]^{(\beta+1)}. \tag{4}$$

The second part of Equation (4) describes how the relative levels of attributes X_m and X_l affect the marginal rate of substitution and how the importance of this ratio is impacted by changes in β. If $\beta = -1$ the MRS becomes the ratio of attribute weights, α_l/α_m, and the value of the summary measure is unaffected as one attribute is exchanged for another in fixed proportion (determined by the attribute weights) regardless of the level of X_l or X_m. For any value of β greater than -1, the MRS declines as X_l increases. This implies a degree of decreasing marginal benefit to increasing the value of any single attribute. As β becomes increasingly large, the decreasing marginal benefit of increasing X_l becomes more pronounced, and the MRS declines more quickly as X_l increases. For very large values of β, an increase in the value of any attribute will have little effect on the summary function unless all other attributes increase in a similar fashion.

These differences are further illustrated by Figure 1. The three points (A, B, and C) represent different combinations of two attributes X_1 and X_2, and each graph shows the relative rankings of A, B, and C under different levels of β. The lines in each graph represent indifference curves, combinations of X_1 and X_2 that lead to the same level of health. Curves further above and to the right of the origin represent higher health levels and better rankings. The top graph depicts perfect substitutability ($\beta = -1$) between X_1 and X_2, in which case point C is the best outcome and point B is the worst. The middle graph represents $\beta = 0$ (a Cobb-Douglas function), and here point C is the worst outcome and point A is the preferred outcome. The bottom graph depicts a case in which X_1 and X_2 have a more complementary relationship ($\beta = 2$) and shows point B as the best combination. As this example indicates, a preference ranking of a set of outcomes can change significantly when the choice of β changes.

The value of β can be chosen based on experience or a particular policy goal. However, the purpose of this procedure is to measure some unobserved characteristic about the population. As such, researchers should consider if the assumptions made by choosing β are both indicative of realistic relationships or tradeoffs in the data and representative of how a decision-maker would evaluate the latent characteristic. For example, if a summary function implies that an increase in infant mortality, no matter how large, has no negative impact on population health as long as there is a proportional decrease in the number of people living sedentary lifestyles (as is the case when $\beta = -1$), then the summary function may not represent realistic tradeoffs in health outcomes or be consistent with microeconomic theory. However, most existing summary measures and rankings aggregate multiple attributes using weighted arithmetic means, implicitly assuming that $\beta = -1$. These summary measures include Multiple Indicators Multiple Causes models, Principal Component Models, and popular policy metrics like the Human Development Index, the World Health Organization rankings, the Commonwealth Fund Rankings, The Robert Wood Johnson Foundation County Health Rankings, and America's Health Rankings.

There are also many papers employing MGE-based summary functions, like the one we propose, in the measurement of latent characteristics such as multidimensional well-being, poverty, and inequality. Often researchers employing MGE-based summary functions test the robustness of aggregation methods to different levels of β. Recently, Maasoumi and Racine (2013) propose a data-driven method for estimating the values of β for different quantiles in the multivariate distribution using nonparametric techniques. Whatever method is used, it is important to explicitly state and justify the choice of β.

3.2 Relative Importance and Attribute Weights

The choice of β allows a researcher to transparently characterize the degree to which an evenly mixed bundle of attributes is preferable. The first part of Equation (4) indicates that the MRS between any two attributes also depends on the relative weight given to each attribute. As α_l grows larger compared to α_m, more units of X_m can be exchanged for a unit increase in X_l, holding the value of the summary function constant, and thus the relative importance of X_l increases.

The methodologies for choosing attribute weights fall into three broad categories: normative weights, hybrid weights, and data-driven weights (For a detailed review of weighting methodologies, see Decancq and Lugo, 2013). Normative weights are often chosen to reflect experts' judgments on the relative importance of each attribute or desirable policy preferences. They may also be assigned by weighting each attribute equally or by constructing weights using the price of each attribute. Normative weights are desirable in that they may reflect experts' analysis on which attributes most affect the underlying latent measure. However, normative weights are inherently subjective, and different experts or policy makers may have different preferences over the different attributes. Additionally, experts' opinions may not reflect realistic relationships between the attributes. For example, behavior considered to be unhealthy by an expert, like excessive drinking, might be common among healthy people of high socioeconomic status. Thus, while the behavior is probably unhealthy on the margin, in practice it may not be negatively correlated with population health. Finally, not accounting for attributes which are highly correlated and possibly represent the same underlying latent information implicitly places more emphasis on these dimensions of the attribute matrix, an issue referred to as double counting.[5]

Hybrid weights are similar to normative weights in that they are based on stated preferences. However, hybrid weights are constructed using survey data to get a more representative picture of social preferences, rather than relying on the opinions of a few experts. Hybrid weights are also often generated by surveying the population at large about the relative importance of different

[5]Normative weighting procedures often attempt to avoid this by assigning weights based on groups of related attributes. However, the underlying relationships between attributes may not be obvious to the researcher. Data-driven weights require no special expertise in the data to calculate.

attributes. For example, the World Health Report rankings of countries' health care systems derived weights by surveying 1006 respondents in 125 countries, half of whom worked for the WHO and half of whom filled out the survey through the WHO website. Respondents were asked to rank the relative importance of different health system goals, and the responses were then used to construct the weights (World Health Organization, 2000; Gakidou et al., 2000).

Data-driven weights solve many of the problems of normative weights by inferring the relative importance of each attribute from the sample data, and they are derived from the distribution of the attribute matrix. This is advantageous because the weights are more representative of the underlying data and because the data-driven weights can avoid the double-counting problem. The disadvantages of data-driven weights are that they might be too sample-specific or they could be based on spurious correlations in the data. Moreover, they may reflect inefficient behavior on the part of decision-makers concerning the latent variable, thus hiding something policy makers might want to change for normative reasons. Common approaches to constructing data-driven weights are to use principal components, factor analysis, or estimate weights for each entity which maximizes that entity's aggregation measure.

In our analysis, we test the robustness of the normative weights in America's Health Rankings by using principal component analysis to generate data-driven weights. Here, the weights assigned to each attribute indicate the degree of correlation between the attribute and the constructed vector that explains the most variance in the multivariate distribution of attributes. While this is an excellent way to uncover the underlying relationships between the attributes, it is not without its drawbacks. One difficulty with principal component weights is that the weights do not necessarily have intuitive meaning and can even be negative.[6] Negative weights can theoretically be corrected by transforming the data such that the first principal component is always in the positive orthant of the m-space of the attributes. However, such a transformation may be equally undesirable, depending on the application. Moreover, principal component analysis generates m orthogonal principal

[6]For example, if the summary function is a weighted geometric mean, an odd number of negative weights would lead to a negative value of the summary function while an even number of negative weights would lead to a positive value of the summary function.

components, and there is no guarantee that the first principal component sufficiently explains the latent variable to be used as the only determinant of the attribute weights. Finally, principal components are calculated using only the first two moments of the multivariate distribution, which increases the likelihood of information loss.[7]

3.3 Data Scaling

As the different health attributes are measured in different units, we must transform the attributes into the same units of measurement. The literature on the measurement of economic development proposes multiple possible transformations, for example z-scores—used in the AHR rankings (detailed in Section 4)—or dividing each attribute by the mean value of that attribute (see Decancq and Lugo, 2013, for a detailed outline of these common transformation techniques). When aggregating using non-linear functions, transformations that result in negative values are problematic.[8] In this paper, we follow the transformation utilized in the Human Development Index (HDI) outlined in United Nations Development Programme (2013),

$$X_{if} = \frac{Y_{if} - \min\{Y_f\}}{\max\{Y_f\} - \min\{Y_f\}} \tag{5}$$

where Y_{if} is attribute f for state i in its original units of measurement, and $\min\{Y_f\}$ and $\max\{Y_f\}$ measure the minimum and maximum values for attribute f. This transformation has two main attractive properties. Most importantly, all outcomes now have the same units. Additionally, the values for the variables all lie between zero and one.[9] Since the transformation maps our original data to the interval (0,1), we account for attributes which are considered undesirable (for example, infant mortality) by subtracting the transformed values from one.

[7]A few distributions, like the multivariate normal, are fully characterized by the first two moments.

[8]Here, again, a weighted geometric mean would generate a positive value of the summary function for an even number of negative attributes but a negative value of the summary function for an odd number of negative attributes.

[9]As zero values are problematic in the context of a geometric mean, we change the minimum possible transformed attribute value to be 0.001 and the maximum possible transformed attribute value to be 0.999.

4 Data

We examine a well-known ranking of the health status of U.S. states, America's Health Rankings (AHR), a partnership between the United Health Foundation, the American Public Health Association, and the Partnership for Prevention. According to AHR, they analyze a "comprehensive set of behaviors, public health policies, community and environmental conditions, and clinical care to provide a holistic view of the health of the nation."[10] Since 1990, AHR has collected attributes of health for each state and provided rankings of states' health. We downloaded data for the 2012 version of the AHR Report, which includes 24 attributes of health, organized into Determinants of health and health Outcomes.[11] Table 1 shows summary statistics for the attributes.[12] For Outcomes, AHR uses Diabetes prevalence, Poor Mental Health Days, Poor Physical Health Days, Geographic Disparities, Infant Mortality, Cardiovascular Deaths, Cancer Deaths, and an overall measure of Premature Death. AHR categorizes the health Determinants into Behaviors (Smoking, Binge Drinking, Obesity, and Sedentary Lifestyle), Community and Environment (Violent Crime, Occupational Fatalities, Infectious Disease, Children in Poverty, and Air Pollution), Policy (Lack of Health Insurance, Public Health Funding, and Immunization Coverage), and Clinical Care (Low Birthweight, Primary Care Physicians, and Preventable Hospitalizations).

To construct their rankings, AHR first transforms each individual attribute into a z-score calculated to three decimal places and truncated between $[-2, 2]$ to reduce the effect of outliers.[13] For any attribute that is considered a bad outcome, like infant mortality, the z-score is multiplied by -1. The final health levels for each state are obtained by taking a weighted average of the transformed attributes for each state, and rankings are determined based on the health levels. All weights are chosen normatively by AHR's Scientific Advisory Committee based on their beliefs about the attribute's effect on overall health, the uniqueness of the information given by the attribute, and the

[10]See http://americashealthrankings.com/About (accessed March 5, 2014).

[11]Specifically, we downloaded data from `http://americashealthrankings.com` on October 17, 2013.

[12]Appendix Table A1 lists these variables and their descriptions.

[13]AHR standardizes their attributes using a population mean when it is available, the unweighted mean of the state attributes when the sample mean is not available, and the median when using attributes from the Behavioral Risk Factor Surveillance System. AHR uses the unweighted standard deviation as the denominator in the z-score calculation.

reliability of the attribute.[14] Thus, the AHR can be characterized as a measure of population health in which multiple attributes are combined using normative weights (based on expert opinion) implicitly assuming perfect substitutability between attributes (due to the choice of aggregating with a weighted arithmetic mean).

5 Results

5.1 Replication

We are able to very closely replicate the published AHR health levels and ranks, and our detailed replication results are available upon request. Figure 2 displays a map of the AHR rankings. Lighter shades correspond to states with better health. Better health levels are concentrated in the Northeast, Minnesota, and Utah, and poor health levels are concentrated in the South. We test the robustness of the AHR rankings to variations in methodology, including not truncating each attribute's z-scores to $[-2, 2]$, replacing median levels with population weighted means and unweighted standard deviations with population weighted standard deviations in the calculations of the z-scores, and replacing the z-score transformation with the alternative transformation described in Equation (5) and used in our aggregation methodology. There is very little change in the rankings when applying these different methodologies, and these detailed results are available upon request.[15]

5.2 Aggregation Method

Next, we examine the stability of health rankings to different methods of aggregation and assumptions about the substitutability or complementarity between different health attributes, utilizing the MGE summary functions described in Section 3. As discussed in Section 3.1 the choice of β determines whether attributes are treated as substitutes or complements in the production of health. In the MGE framework, standardized attribute values are combined using a harmonic mean (constant

[14] See http://americashealthrankings.com/About/Weighting (accessed March 5, 2014).

[15] We show our replication of the AHR health levels in Appendix Figure A1 and show the detailed results of our robustness checks to the AHR methodology in Appendix Table A2.

elasticity of substitution function) except for the special cases when β is -1 or 0, which use the arithmetic mean (perfect substitutes function) or geometric mean (Cobb Douglas function). We examine β values of -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 2, and 5, use the normative weights developed by AHR, and use the attribute transformation described in Equation (5).[16] These values for β are values commonly used in the literature examining multidimensional well-being and inequality (Decancq and Lugo, 2013; Maasoumi and Nickelsburg, 1988; Lugo, 2007). Since $\sigma = {}^{1}\!/\!{\scriptstyle(1+\beta)}$ is the constant elasticity of substitution between attributes, the values $\beta = \{-1, -0.5, -0.25, 0, 0.25, 0.5, 1, 2, 5\}$ translate to substitution elasticities of $\sigma = \{\infty, 2, 1.\overline{3}, 1, 0.8, 0.\overline{6}, 0.5, 0.\overline{3}, 0.1\overline{6}\}$, respectively.

Figure 3 shows scatterplots of the AHR rankings and rankings derived from the MGE summary functions. Each specific scatterplot shows the original AHR rankings on the x-axis and the rankings corresponding to the relevant level of β on the y-axis.[17] The original AHR rankings diverge from the MGE rankings as the attributes change from substitutes to complements, and the correlation coefficient decreases to 0.6 when β=2.[18] This is not surprising since as the value of β increases the relative importance of a state's worst health attribute also increases. Thus, even assuming a state ranked first in every category except one for which it ranked last, that state would eventually fall to last place in the rankings as β increased.

To illustrate the divergences and similarities between the original AHR rankings and the MGE rankings, Figure 4 shows a map of each state's rankings calculated under a subset of the β values used above: $\beta = -1, 0, 1, 2$. Some interesting regional patterns emerge. The Northeast largely stays in good health. In fact, Vermont remains the healthiest state throughout all our chosen values for β. Additionally, Colorado, Utah, Oregon and Washington maintain relatively high health rankings. However, some parts of the South and upper Midwest see large changes in their health rankings. In the South, Louisiana, Alabama, Georgia and Florida see large rankings increases, while in the Midwest, Minnesota and Wisconsin see large rankings decreases. Mississippi, West Virginia, and Kentucky remain among the unhealthiest states for all values of β. We further compare the

[16]In Section 5.3 below we examine the effects of using data-driven weights in the entropic aggregation framework.

[17]Appendix Table A3 shows the full ranking results.

[18]The rankings are not identical when $\beta = 1$ since the entropy rankings use the HDI transformation rather than the z-score transformation.

entropic aggregation results with some specific examples in Table 2, which shows the five states with the largest rankings increases, decreases, and smallest rankings changes as the aggregation procedure changes from perfectly substitutable to higher degrees of complementarity. Minnesota falls rapidly in health rankings as health attributes are more complementary, going from 7th to 45th. Minnesota does better than the national average in all but three of the 24 health attributes. However, Minnesota also has the worst score for infectious disease by a wide margin. As the aggregation method becomes more complementary in nature, increasing emphasis is placed on this one poor outcome. Similarly, Wisconsin does better than the national average in 19 out of 24 health attributes, but has the nation's worst rate of binge drinking and lowest rate of public health funding. Conversely, Alabama's and South Carolina's rankings each increase by 19 spots. Alabama and South Carolina are below average in 18 and 20 attributes, respectively, but are not near the bottom in any particular outcome. This trait becomes increasingly favorable as β increases. Vermont and Mississippi do not change rankings, although they possess very different health levels. Mississippi falls below the national average in 19 out of 24 attributes, and is at or near the minimum attribute levels for roughly a third of the attributes. Vermont is above average in all but two attributes, and Vermont's outcomes for the two attributes which are below average, binge drinking and cancer deaths, are not far below the national average.

The role of complementarity is not trivial. From a practical point of view, a state with one or two dimensions of very poor health could go largely unnoticed by policy makers because a linear aggregator masks these dimensions. This is especially problematic if the few poor dimensions generally only affect specific minority populations or those of low socioeconomic status, which is not uncommon for health outcomes. For example, Idaho ranks very well on many of the physical health dimensions, which is partially due to its demography.[19] However, it has the nation's worst primary care physician coverage, which may disproportionately affect people of low socioeconomic status and those living in rural areas. Thus, the recognition of a degree of complementary between attributes can help uncover health inequalities and disparities within states.

[19] Idaho has a relatively young population and a disproportionally low percentage of minorities.

5.3 Weights

Finally, we test the robustness of the aggregation methodology to data-driven weights. We generate data-driven weights from the first principal component of the variance/covariance matrix of our data. To ensure that the weights add to one, we divide each element of the first principal component by the sum of all the the first principal components. Table 3 displays the AHR and principal component weights for each health attribute. There are some discrepancies between AHR's normative assessment of attribute importance and the principal component weights. The two weighting methodologies particularly diverge in the Outcome variables. The AHR weights underestimate the importance of cardiovascular deaths, cancer deaths, diabetes, poor mental and physical health days, and premature death in the multivariate distribution of attributes. Relative to the principal component weights the AHR weights significantly overemphasize immunization coverage, infectious disease, and public health funding.

The principal component weights for binge drinking and geographic disparities are negative. This suggests, counterintuitively, that binge drinking and geographic disparities are concentrated in states with higher relative population health.[20] To make the weights suitable for the entropy calculations, we reclassify these two variables as good attributes and subtract the transformed values from one. This contradiction highlights the tradeoffs between normative and data-driven weights. The normative AHR weights emphasize attributes of population health that are inversely related to the other attributes of health and do not accurately represent the underlying relationships in the data.[21] However, transforming the data to get positive principal component weights is equally counter-intuitive, as binge drinking and geographic health disparities are not desirable attributes. Thus, principal component weights are less likely to misrepresent the data, while normative weights are less likely to misrepresent the desirability of each attribute. We think this example illustrates the benefits of this type of sensitivity analysis when evaluating multidimensional outcomes.[22]

[20] The AHR weight for geographic disparity is also 10 times larger than the calculated principal component weight.

[21] The negative weight for drinking is possibly due to the positive correlation between drinking and socioeconomic status investigated by van Ours (2004).

[22] One interpretation of this result is that the assessment of multidimensional outcomes should de-emphasize misleading attributes. In this analysis, binge drinking and geographic disparities appear to be noisy attributes of

Figure 5 shows scatterplots of the MGE rankings using AHR weights on the x-axis and MGE rankings using principal component weights on the y-axis. When $\beta = -1$ and the summary function is linear, the rankings using AHR's weights and principal component weights show a fairly strong relationship, with a Spearman rank correlation coefficient of 0.95. There are, however, some states which are ranked fairly differently under the two weighting schemes. For example, Rhode Island falls 11 spots from 10^{th} to 21^{st}, Maine falls 11 spots from 9^{th} to 20^{th}, and South Dakota rises eight spots from 27^{th} to 19^{th}.[23] Both Rhode Island and Maine have low measures of geographic health disparities, a trait that hurts their rankings when geographic health disparities is treated as a positive attribute under the principal component weights. South Dakota's high rate of geographic disparities and binge drinking turn into positives when the principal component weights are used. The rankings diverge more as attributes are treated as more complementary in nature. Interestingly, much of the divergence seems to occur among the healthiest states in each weighting framework. This is due to the interaction between differently-weighted attributes and the increasing emphasis placed on the states' worst health outcomes.

6 Conclusion

This paper proposes a new framework for the measurement and ranking of population health. Many well-known rankings of population health use a weighted average of different health attributes to compute a summary measure of population health. However, a weighted average implicitly assumes an infinite degree of substitutability between the different attributes. In the context of a health ranking, this assumption implies that a decrease in one attribute can be compensated for by a proportional increase in another attribute, where the proportion is independent of the levels of the attributes. We instead utilize a summary function derived by Maasoumi (1986) which minimizes the entropic distance between the summary metric and the underlying distribution of the attributes. This methodology has a number of attractive properties. First, the MGE-based summary functions

population health and are possibly more representative of some other characteristic of the population, such as the demographic composition.

[23] These changes can also been seen in column 1 of Appendix Table A3.

preserve as much information as possible from the underlying multivariate distribution of health attributes. Second, our methodology naturally incorporates variable degrees of substitutability or complementarity between the attributes. Finally, the summary functions mirror functional forms often utilized in economic theory, such as the constant elasticity of substitution and Cobb-Douglas utility or production functions, which makes the choice and interpretation of parameters more intuitive. In fact, rankings derived from weighted averages are a special case of our methodology, in which attributes are combined using a perfect substitutes utility function.

We compare methodologies using America's Health Rankings, which aggregates 24 attributes of health to a measure of state-level population health using a normatively-weighted arithmetic mean. We find that states' rankings are fairly stable when we change the original normative attribute weights to weights derived from principal components. However, as we move away from the infinite substitutability implicitly assumed in the original America's Health Rankings methodology towards a more complementary relationship between the different health attributes, the correlation coefficient between the AHR and our rankings falls from above 0.9 to 0.6. We see some large changes in state rankings, with some states commonly associated with good health, such as Minnesota, dropping rapidly in the health rankings. Research indicates that officials in state health departments are aware of America's Health Rankings and use America's Health Rankings in designing public policies (Erwin et al., 2008).[24] If policy makers do not understand the implicit assumptions embedded in the AHR rankings, it may lead to a mis-allocation of public resources.

It is interesting that we find very little sensitivity due to changes in scaling technique or weighting technique given that these have been the primary methodological focus in the health rankings literature. Meanwhile, while changes in the summary functional form lead to large changes in the health rankings, research into summary functional form is limited in the measurement and ranking of multidimensional health.

Our framework involves a set of choices which researchers measuring population health or developing health rankings should state explicitly. First, researchers must choose a set of attributes

[24]A list of policies and initiatives were related to the results of America's Health Rankings can be found on their "Success Story Archive": `http://americashealthrankings.org/Stories/`.

to use in the measurement of health. Second, researchers need to choose a set of weights for the attributes. These weights can be chosen normatively or through data-driven methods. Since our results demonstrate that different weighting methodologies can result in some nontrivial changes in the levels of population health and health rankings, researchers should examine how their measurements change under different weighting methodologies.

Finally, researchers must choose which values of β to use. This is also a subjective choice and the reasons for the choice should be made explicit. For example, the Cobb-Douglas production function is the most widely used production function in economics and coincides with $\beta = 0$, which represents neither substitutability nor complementarity between the attributes. Alternatively, economic studies have estimated production function elasticities between a number of goods, and researchers could use these estimates to justify choices of β. For example, Jensen and Morrisey (1986) estimate substitution elasticities between different hospital production inputs. On the low end, they find an elasticity of 0.16 between medical staff and hospital beds ($\beta = 5.25$), and on the higher end they find an elasticity of 2.13 between nurses and residents ($\beta = -0.53$). Hamermesh (2008) finds substitution elasticities of 0.2 to 0.4 between household production factors (β between 1.5 and 4), and Antras (2004) finds substitution elasticities between capital and labor in the United States significantly below one ($\beta > 0$). Thus, there is an empirical precedent for complementarity in production inputs in a variety of contexts, providing evidence against the assumption that inputs are perfectly substitutable in the production of population health. While any single choice of β can be disputed on empirical or normative grounds, the advantage of the methodology we propose is that it is a relatively simple exercise to compare results from several different β values. This can be used to construct a metric or ranking range—similar to a confidence interval—for reasonable ranges of β. Additionally, values of β could be estimated from individual-level data in some contexts.

Our study has a number of limitations which suggest areas for future research. First, although we constructed data-driven weights and compared them to the weights developed by America's Health Rankings, there are other weighting methodologies which we have not considered. Some recent economics research suggests using Bayesian econometrics to develop weights which account for spatial correlations, and other papers have proposed using weights which maximize the summary metric for each observation (Courtemanche et al., 2013; Decancq and Lugo, 2013; Maasoumi and Xu, 2013). Second, our current methodology does not account for sampling errors in the underlying metrics. Thus, some of the differences we find between state rankings may not be statistically significantly different from the original rankings. This is also an issue in the original *America's Health Rankings*. Thirdly, our study does not estimate heterogeneity in population health within states. One extension of this study could examine smaller geographies, perhaps counties or zip codes, to achieve a more finely tuned analysis of health levels in the United States. Relatedly, our study, like most previous work, does not account for differences in outcomes that are related to differing state demographic characteristics.

Our methodology could be extended to analyze the level, trend, and decomposition of health inequality, a topic that continues to receive much attention in the literature (Allanson et al., 2010; Clark, 2011; Wagstaff et al., 1991). In addition to developing an aggregation metric based on MGE, Maasoumi (1986) develops a measure of inequality which has been widely used in the measurement of inequality in multidimensional well-being. This metric of inequality can also be decomposed to determine the relative inequalities within groups versus across groups. Thus, our methodology could be extended to measure health inequalities within and across states, counties, or groups of individuals, as well as measuring trends across time in these measures of inequality.

References

Allanson, P., Gerdtham, U.-G., Petrie, D., 2010. Longitudinal analysis of income-related health inequality. Journal of Health Economics, 29 (1), 78–86.

Antras, P., 2004. Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. Contributions in Macroeconomics, 4 (1).

Bivins, L., 2009. Wis. health care system gets high marks. USA Today, p. ARC.

Booske, B. C., Rohan, A. M., Kindig, D. A., Remington, P. L., 2010. Peer Reviewed: Grading and Reporting Health and Health Disparities. Preventing Chronic Disease, 7 (1).

Brandolini, A., 2008. On applying synthetic indices of multidimensional well-being: Health and income inequalities in selected EU countries. Working Paper, Bank of Italy, Department for Structural Economic Analysis, 1–41.

Clark, R., 2011. World health inequality: Convergence, divergence, and development. Social Science & Medicine, 72 (4), 617–624.

Courtemanche, C., Soneji, S., Tchernis, R., 2013. Modeling Area-Level Health Rankings. NBER Working Paper, 1–39.

Cover, T. M., Thomas, J. A., 2006. Elements of Information Theory. Hoboken, Wiley-Interscience.

Davis, K., Fund, C., 2004. Mirror, mirror on the wall: looking at the quality of American health care through the patient's lens. Commonwealth Fund.

Davis, K., Fund, C., 2007. Mirror, mirror on the wall: an international update on the comparative performance of American health care. Commonwealth Fund.

Davis, K., Schoen, C., Stremikis, K., Fund, C., 2010. Mirror, mirror on the wall: How the performance of the US health care system compares internationally: 2010 update. Commonwealth Fund.

Decancq, K., Decoster, A., Schokkaert, E., 2009. The evolution of world inequality in well-being. World Development, 37 (1), 11–25.

Decancq, K., Lugo, M. A., 2013. Weights in Multidimensional Indices of Wellbeing: An Overview. Econometric Reviews, 32 (1), 7–34.

Decancq, K., Ooghe, E., 2010. Has the world moved forward? A robust multidimensional evaluation. Economics Letters, 107 (2), 266–269.

Erwin, P., Myers, C., Myers, G., Daugherty, L., 2008. States' responses to America's Health Rankings: The search for meaning, utility, and value. Journal of Public Health Management & Practice, 17 (5), 406–412.

Gakidou, E., Murray, C. J., Frenk, J., 2000. Measuring preferences on health system performance assessment. World Health Organization, GPE Discussion Paper Series No. 20.

Hamermesh, D. S., 2008. Direct estimates of household production. Economics Letters, 98 (1), 31–34.

Hilts, P. J., 2000. Europeans perform highest in ranking of world health. New York Times, p. A12.

Hirschberg, J. G., Maasoumi, E., Slottje, D. J., 1991. Cluster analysis for measuring welfare and quality of life across countries. Journal of Econometrics, 50 (1), 131–150.

Hirschberg, J. G., Maasoumi, E., Slottje, D. J., 2001. Clusters of attributes and well-being in the USA. Journal of Applied Econometrics, 16 (3), 445–460.

Jensen, G. A., Morrisey, M. A., 1986. The role of physicians in hospital production. The Review of Economics and Statistics, 432–442.

Justino, P. et al., 2005. Empirical applications of multidimensional inequality analysis. Poverty Research Unit at Sussex Working paper.

Kanarek, N., Tsai, H.-L., Stanley, J., 2011. Health Ranking of the Largest US Counties Using the Community Health Status Indicators Peer Strata and Database. Journal of Public Health Management and Practice, 17 (5), 401–405.

Kindig, D. A., Asada, Y., Booske, B., 2008. A population health framework for setting national and state health goals. Journal of the American Medical Association, 299 (17), 2081–2083.

Kindig, D., Stoddart, G., 2003. What is population health? American Journal of Public Health, 93 (3), 380–383.

Lugo, M. A., 2007. Comparing multidimensional indices of inequality: Methods and application. Research on Economic Inequality, 14, 213–236.

Lugo, M. A., Maasoumi, E., 2009. Multidimensional poverty measures from an information theory perspective. Working Paper.

Maasoumi, E., 1986. The Measurement and Decomposition of Multi-Dimensional Inequality. Econometrica, 54 (4), 991–997.

Maasoumi, E., Jeong, J. H., 1985. The trend and the measurement of world inequality over extended periods of accounting. Economics Letters, 19 (3), 295–301.

Maasoumi, E., Lugo, M. A., 2008. The information basis of multivariate poverty assessments. Quantitative Approaches to Multidimensional Poverty Measurement, 1–29.

Maasoumi, E., Nickelsburg, G., 1988. Multivariate measures of well-being and an analysis of inequality in the Michigan data. Journal of Business & Economic Statistics, 6 (3), 327–334.

Maasoumi, E., Racine, J. S., 2013. A Solution to Aggregation and an Application to Multidimensional "Well-being" Frontiers. Emory Working Paper Series #1306, 1–29.

Maasoumi, E., Trede, M., 2001. Comparing income mobility in Germany and the United States using generalized entropy mobility measures. Review of Economics and Statistics, 83 (3), 551–559.

Maasoumi, E., Xu, T., 2013. Weights and Substitution Degree in Multidimensional Well-Being in China. Working Paper, 1–16.

Maasoumi, E., Zandvakili, S., 1990. Generalized entropy measures of mobility for different sexes and income levels. Journal of Econometrics, 43 (1), 121–133.

Makdissi, P., Yazbeck, M., 2014. Measuring socioeconomic health inequalities in presence of multiple categorical information. Journal of Health Economics, 34 (0), 84 – 95.

Neergaard, L., 2000. France Rated No. 1 in Health Care. The Associated Press News Service.

van Ours, J. C., 2004. A pint a day raises a man's pay; but smoking blows that gain away. Journal of Health Economics, 23 (5), 863 – 886.

Peppard, P. E., Kindig, D. A., Dranger, E., Jovaag, A., Remington, P. L., 2008. Ranking community health status to stimulate discussion of local public health issues: the Wisconsin County Health Rankings. American Journal of Public Health, 98 (2), p. 209.

Rubin, R., 2000. WHO's assessment: U.S. health system is inefficient. USA Today, p. 10D.

Rubin, R., 2009. 'Shockingly wide' health gaps among states - Insurance data, access to care compared. USA Today, p. 6D.

Stewart, S. T., Cutler, D. M., Rosen, A. B., 2013. US Trends in Quality-Adjusted Life Expectancy From 1987 to 2008: Combining National Surveys to More Broadly Track the Health of the Nation. American Journal of Public Health, 103 (11), e78–e87.

Theil, H., 1967. Economics and Information Theory. Chicago, Rand McNally and Company.

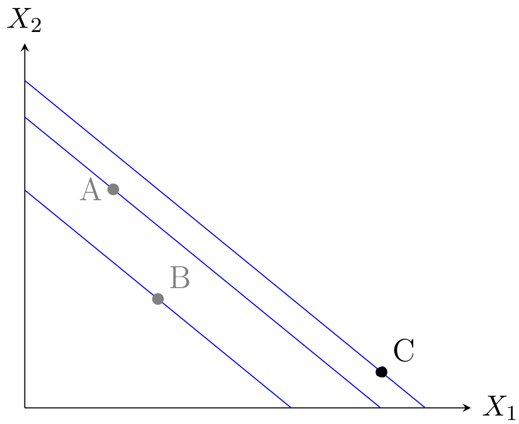
United Nations Development Programme, 2013. Human Development Report 2013. Oxford University Press, New York.

Wagstaff, A., Paci, P., Van Doorslaer, E., 1991. On the measurement of inequalities in health. Social Science & Medicine, 33 (5), 545–557.

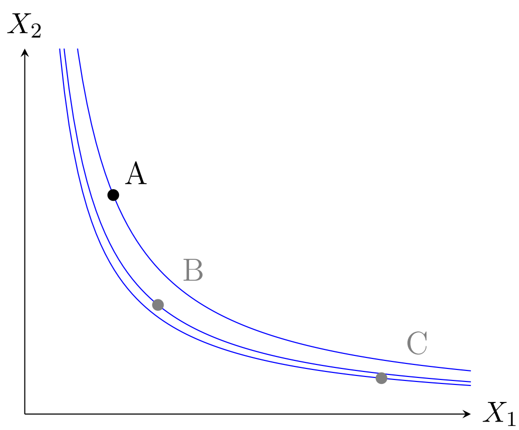
World Health Organization, 2000. The world health report 2000: Health systems: improving performance.Technical report, World Health Organization.

Figure 1: Aggregation Functional Forms and Rankings Changes

(a) Perfect Substitutes Ranks: $S = X_1 + X_2$



(b) Cobb-Douglas Ranks: $S = X_1^{0.5} X_2^{0.5}$



(c) CES Ranks: $S = \left(0.5X_1^{-2} + 0.5X_2^{-2}\right)^{-\frac{1}{2}}$

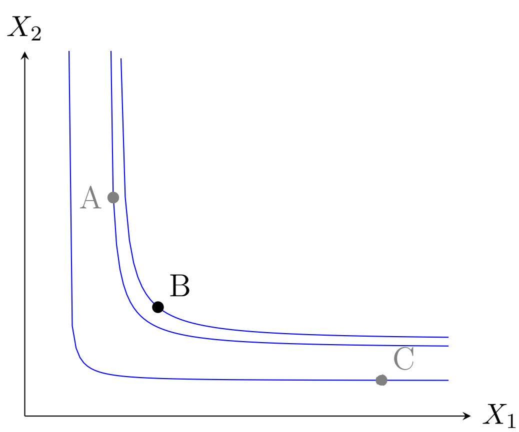
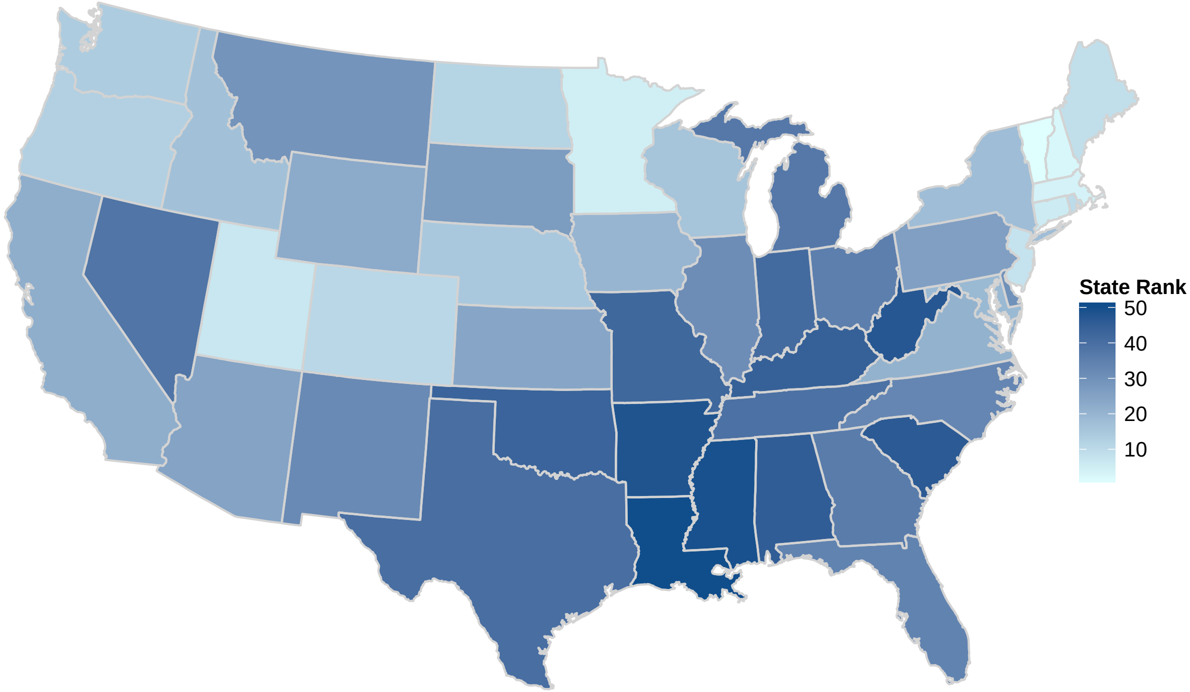
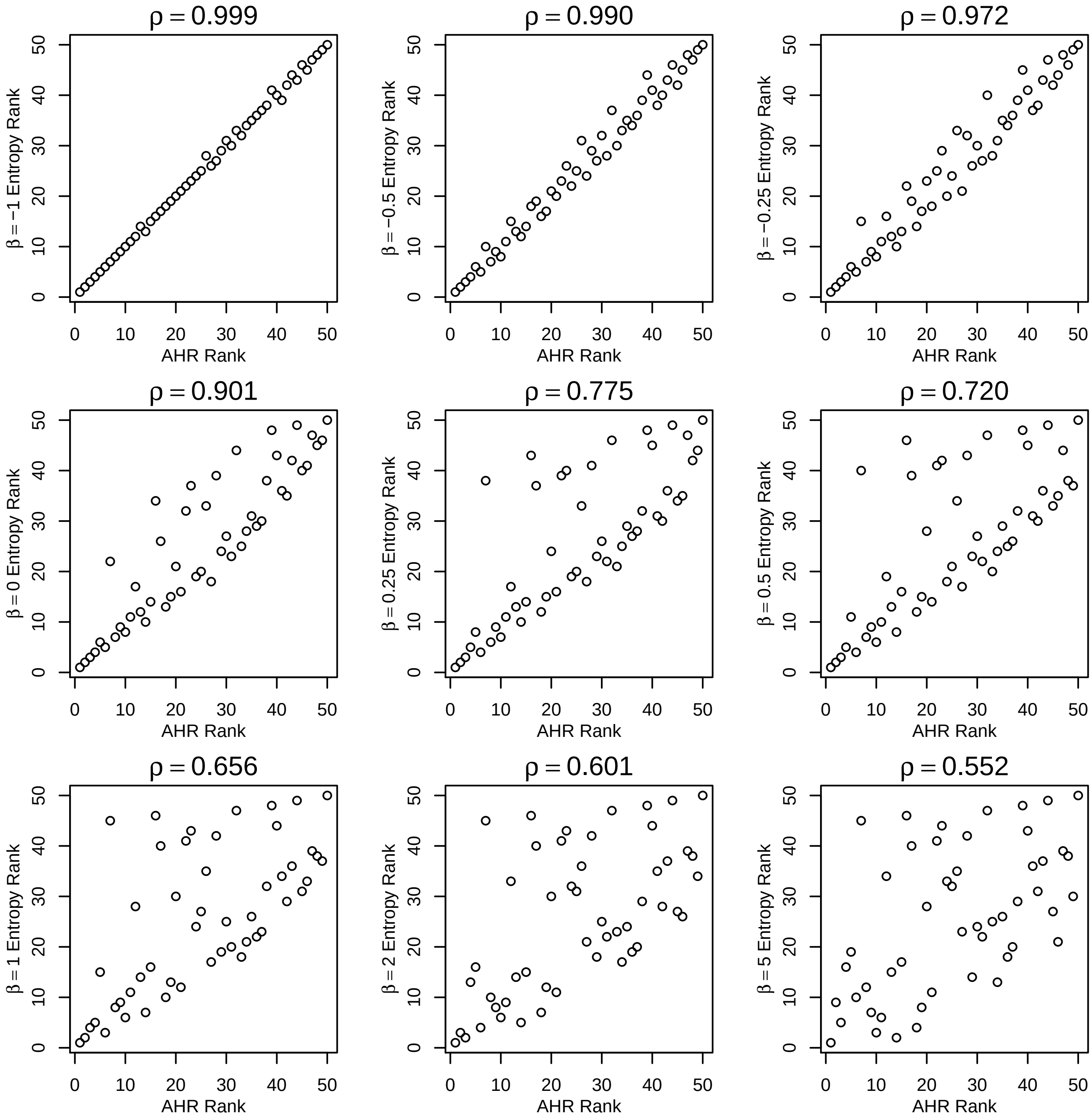


Figure 2: Map of State Population Health Ranks from AHR



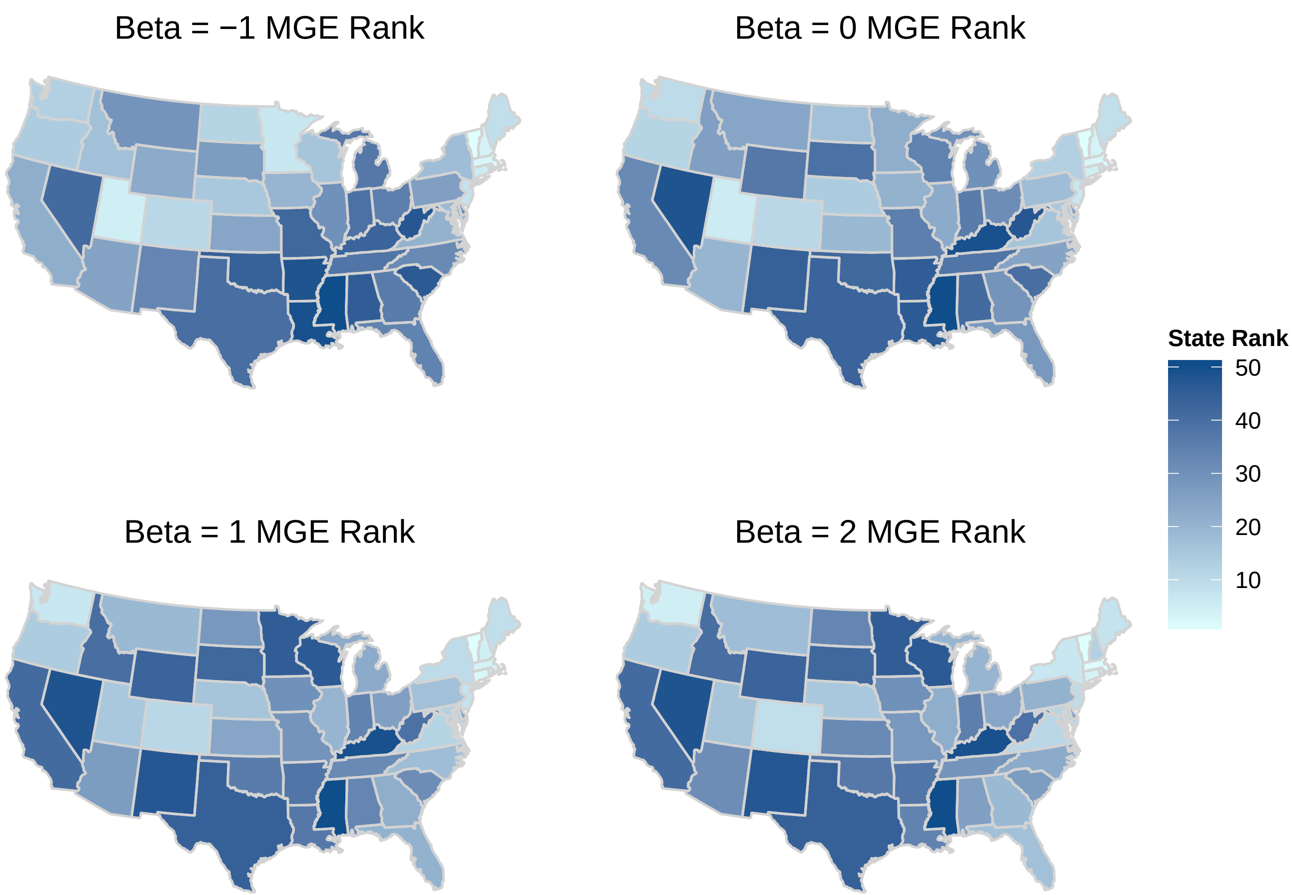
Notes: Data from the 2012 version of America's Health Rankings. Lighter shades correspond to lower health rankings. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one.

Figure 3: Scatterplot of State Population Health Rankings
America's Health Rankings vs. MGE Rankings



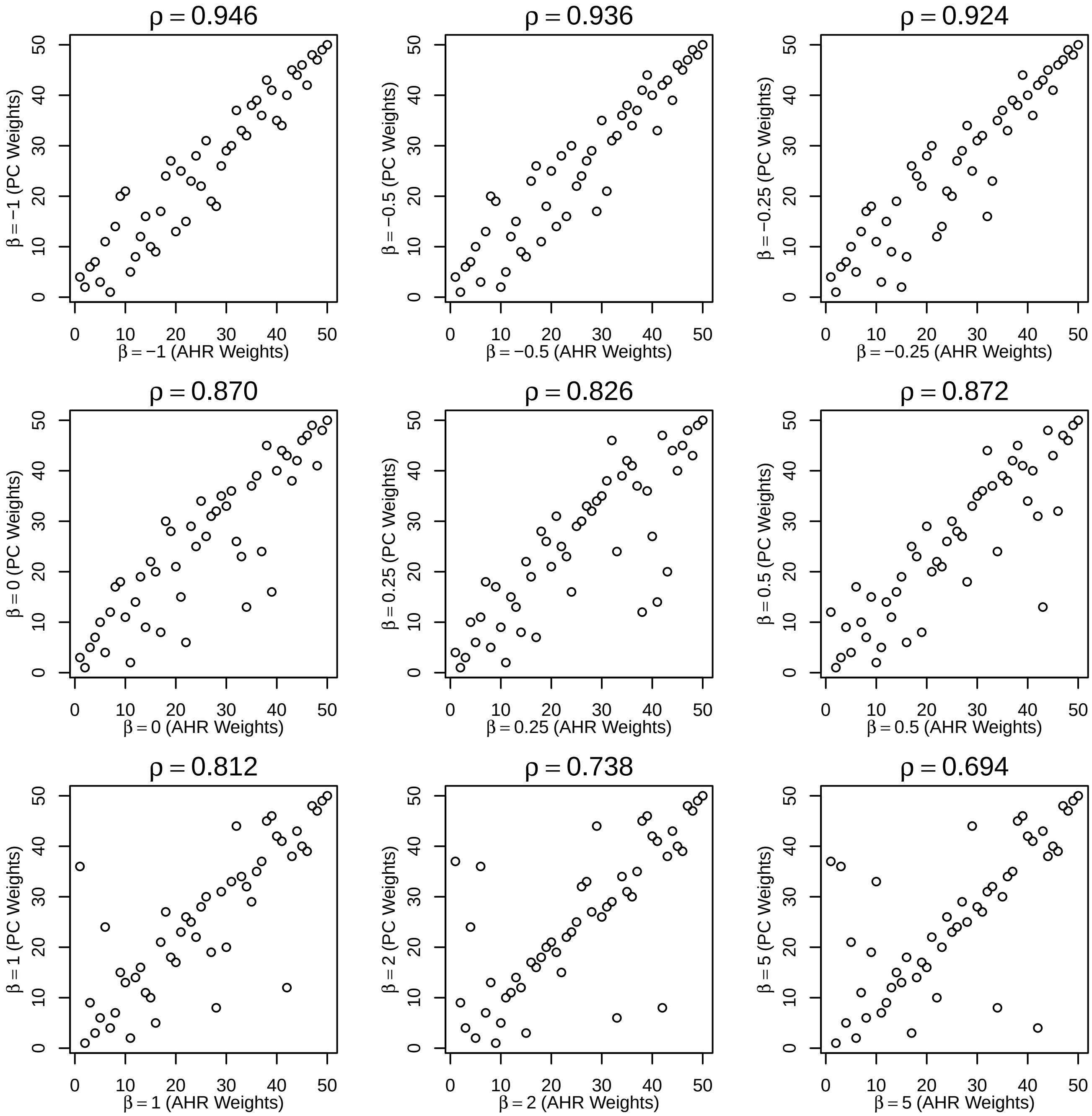
Notes: Data from the 2012 version of America's Health Rankings. The scatterplots show the AHR state health rankings compared to state health rankings using the entropic aggregation function under different assumptions regarding β, the substitution parameter. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one. All outcomes use the weights developed by AHR. The correlation coefficients correspond to Spearman's rank correlation coefficient.

Figure 4: Map of MGE State Population Health Rankings Under Different Values of Beta



Notes: Data from the 2012 version of America's Health Rankings. Lighter shades correspond to lower health rankings. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one.

Figure 5: Scatterplot of State Population Health Rankings
MGE Using AHR Weights vs. MGE Using Principal Component Weights



Notes: Data from the 2012 version of America's Health Rankings. The scatterplots show state health rankings using the entropic aggregation function under different assumptions regarding β, the substitution parameter, using AHR weights on the x-axis and principal component weights on the y-axis. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one. The correlation coefficients correspond to Spearman's rank correlation coefficient.

Table 1: Summary Statistics

Behavior	Category	Mean	Std.Dev	Weight (%)
Smoking	Behaviors	20.11	3.68	7.50
Binge Drinking	Behaviors	18.29	2.57	5.00
Obesity	Behaviors	27.43	2.96	5.00
Sedentary Lifestyle	Behaviors	25.43	3.95	2.50
High School Graduation	Behaviors	75.50	5.82	5.00
Violent Crime	Community and Environment	404.00	106.26	5.00
Occupational Fatalities	Community and Environment	4.10	1.30	2.50
Infectious Disease	Community and Environment	12.40	4.36	5.00
Children in Poverty	Community and Environment	21.40	4.17	5.00
Air Pollution	Community and Environment	10.50	2.28	5.00
Lack of health Insurance	Policy	16.00	4.31	5.00
Public Health Funding	Policy	92.00	29.05	2.50
Immunization Coverage	Policy	90.30	1.98	5.00
Low Birthweight	Clinical Care	8.10	1.05	5.00
Primary Care Physicians	Clinical Care	120.00	23.02	5.00
Preventable Hospitalizations	Clinical Care	67.00	12.35	5.00
Diabetes	Outcomes	9.77	1.11	2.00
Poor Mental Health Days	Outcomes	3.86	0.39	2.00
Poor Physical Health Days	Outcomes	3.95	0.42	2.00
Geographic Disparity	Outcomes	0.18	0.04	5.00
Infant Mortality	Outcomes	6.50	1.16	5.00
Cardiovascular Deaths	Outcomes	264.90	31.31	2.00
Cancer Deaths	Outcomes	182.50	14.84	2.00
Premature Death	Outcomes	7151.00	1248.68	5.00

Notes: Data from the 2012 version of America's Health Rankings.

Table 2: Examples of State Rankings Changes

Panel A: Largest Rankings Decreases

State	Rank for Value of Beta -1	-0.5	0	1	2	Change
Minnesota	7	10	22	45	45	-38
Wisconsin	16	18	34	46	46	-30
Idaho	17	19	26	40	40	-23
North Dakota	12	15	17	28	33	-21
Wyoming	23	26	37	43	43	-20

Panel B: Largest Rankings Increases

State	Rank for Value of Beta -1	-0.5	0	1	2	Change
Alabama	45	45	41	33	26	19
South Carolina	46	42	40	31	27	19
Florida	34	33	28	21	17	17
Georgia	36	34	29	22	19	17
Michigan	37	36	30	23	20	17

Panel C: Smallest Rankings Changes

State	Rank for Value of Beta -1	-0.5	0	1	2	Change
Mississippi	50	50	50	50	50	0
Nebraska	15	14	14	16	15	0
Oregon	14	13	12	14	14	0
Vermont	1	1	1	1	1	0
Hawaii	2	2	2	2	3	-1

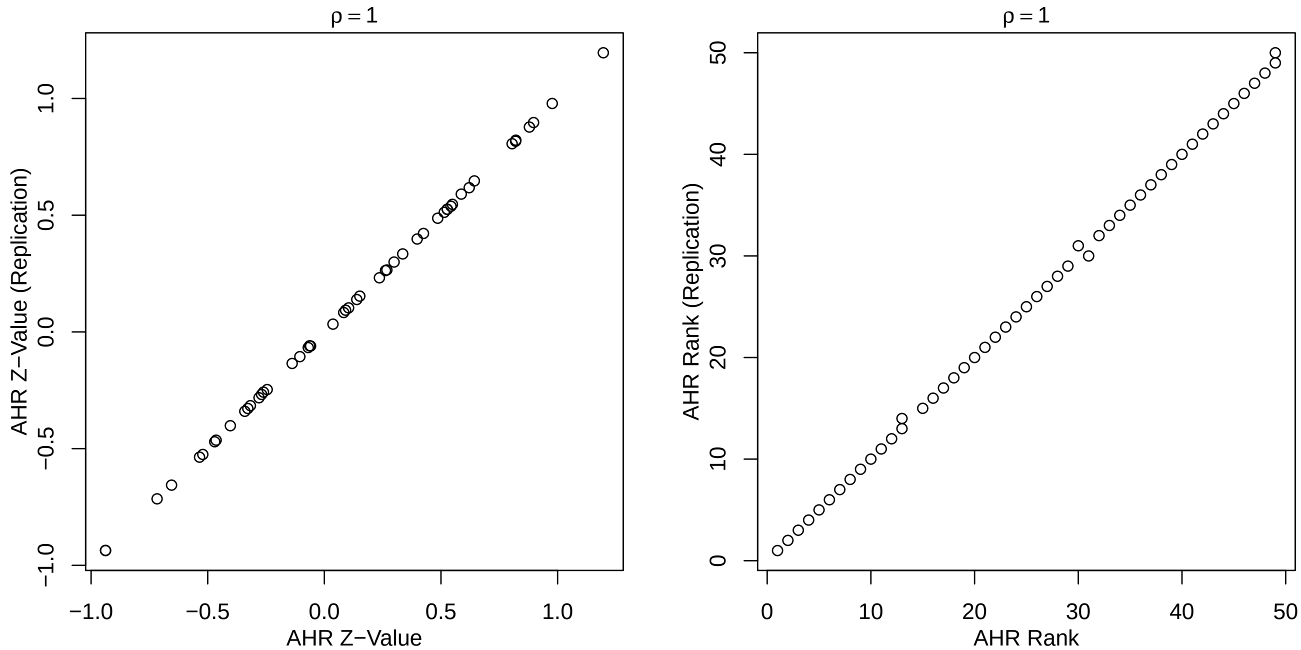
Notes: Data from the 2012 version of America's Health Rankings. Each row shows the rankings for a state's health under different assumptions regarding β, the substitution parameter. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one.

Table 3: AHR and Principal Component Weights

Behavior	Category	AHR Weight (%)	PC Weight (%)
Smoking	Behaviors	7.50	6.08
Binge Drinking	Behaviors	5.00	-4.58
Obesity	Behaviors	5.00	6.28
Sedentary Lifestyle	Behaviors	2.50	6.59
High School Graduation	Behaviors	5.00	4.04
Violent Crime	Community & Environment	5.00	3.25
Occupational Fatalities	Community & Environment	2.50	3.48
Infectious Disease	Community & Environment	5.00	1.40
Children in Poverty	Community & Environment	5.00	5.11
Air Pollution	Community & Environment	5.00	3.35
Lack of health Insurance	Policy	5.00	3.61
Public Health Funding	Policy	2.50	1.16
Immunization Coverage	Policy	5.00	0.64
Low Birthweight	Clinical Care	5.00	5.85
Primary Care Physicians	Clinical Care	5.00	3.40
Preventable Hospitalizations	Clinical Care	5.00	6.28
Diabetes	Outcomes	2.00	6.95
Poor Mental Health Days	Outcomes	2.00	5.74
Poor Physical Health Days	Outcomes	2.00	5.84
Geographic Disparity	Outcomes	5.00	-0.55
Infant Mortality	Outcomes	5.00	6.21
Cardiovascular Deaths	Outcomes	2.00	6.84
Cancer Deaths	Outcomes	2.00	5.70
Premature Death	Outcomes	5.00	7.29

Notes: Data from the 2012 version of America's Health Rankings. The AHR weight shows the weight given to each outcome in the America's Health Rankings Report. The PC weights are calculated as the normalized first principal component of matrix of attributes transformed by Equation 5.

Figure A1: Replication of AHR Population Health Levels and Ranks



Notes: Data from the 2012 version of America's Health Rankings. The scatter plots show state health rankings and health levels from the AHR website and our replication of the AHR methodology. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one. The correlation coefficient corresponds to Pearson's correlation coefficient for the z-values (health levels) and Spearman's rank correlation coefficient for the ranks.

Table A1: AHR Attribute Descriptions

Behavior	Category	Data Source	Description
Smoking	Behaviors	BRFSS	Percentage of population over age 18 that smokes on a regular basis (smoked at least 100 cigarettes in their lifetime and currently smoke every day or some days).
Binge Drinking	Behaviors	BRFSS	Percentage of population over age 18 that drank excessively in the last 30 days (five or more drinks for males and four or more drinks for females on one occasion).
Obesity	Behaviors	BRFSS	Percentage of population over 18 estimated to be obese, with a body mass index (BMI) of 30.0 or higher.
Sedentary Lifestyle	Behaviors	BRFSS	Percentage of the population over age 18 who report doing no physical activity or exercise (such as running calisthenics, golf, gardening, or walking) other than their regular job in the last 30 days.
High School Graduation	Behaviors	NCES	Percentage of incoming ninth graders who graduate in four years from a high school with a regular degree.
Violent Crime	Community & Environment	FBI	Number of murders, rapes, robberies, and aggravated assaults per 100,000 population.
Occupational Fatalities	Community & Environment	CFOI, BLS	Number of fatalities from occupational injuries per 100,000 workers.
Infectious Disease	Community & Environment	CDC, MMWR	Number of reported cases of measles, pertussis, syphilis and Hepatitis A per 100,000 population.
Children in Poverty	Community & Environment	CPS, Census Bureau	Percentage of persons under age 18 who live in households at or below the poverty threshold.
Air Pollution	Community & Environment	EPA, Census Bureau	Average exposure of the general public to particulate matter of 2.5 microns or less in size.
Lack of Health Insurance	Policy	CPS, Census Bureau	Percentage of the population that does not have health insurance privately, through their employer, or the government.
Public Health Funding	Policy	TFAH	State funding dedicated to public health as well as federal funding directed to states by the Centers for Disease Control and Prevention and the Health Resources and Services Administration.
Immunization Coverage	Policy	CDC, NIP	Average percentage of children ages 19 to 35 months who have received these individual vaccinations: four or more doses of DTP, three or more doses of poliovirus vaccine, one or more doses of any measles-containing vaccine, and three or more doses of HepB vaccine.

(Continued on next page)

Table A1 (continued)

Behavior	Category	Data Source	Description
Low Birthweight	Clinical Care	CDC, NCHS	Percentage of babies weighing less than 2,500 grams (5 pounds, 8 ounces) at birth.
Primary Care Physicians	Clinical Care	AMA	Number of primary care physicians (including general practice, family practice, OB-GYN, pediatrics, and internal medicine) per 100,000 population.
Preventable Hospitalizations	Clinical Care	Dartmouth Atlas	Discharge rate among the Medicare population for diagnoses that are amenable to non-hospital based care.
Diabetes	Outcomes	BRFSS	Percentage of population over age 18 who have been told by a health professional that they have diabetes (does not include pre-diabetes or diabetes during pregnancy).
Poor Mental Health Days	Outcomes	BRFSS	Number of days in the previous 30 days when a person indicates their activities were limited due to mental health difficulties.
Poor Physical Health Days	Outcomes	BRFSS	Number of days in the previous 30 days when a person indicates their activities were limited due to physical health difficulties.
Geographic Disparity	Outcomes	CDC, NCHS	Variation in overall mortality rates among the counties within a state.
Infant Mortality	Outcomes	CDC, NCHS	Number of infant deaths (before age 1) per 1,000 live births.
Cardiovascular Deaths	Outcomes	CDC, NCHS	Number of deaths due to all cardiovascular diseases, including heart disease and strokes, per 100,000 population.
Cancer Deaths	Outcomes	CDC, NCHS	Number of deaths due to all causes of cancer per 100,000 population.
Premature Death	Outcomes	CDC, NCHS	Number of years of potential life lost prior to age 75 per 100,000 population.

Notes: Data from the 2012 version of America's Health Rankings.

Table A2: Replication of AHR Rankings and Robustness Checks

State	Replication Number (1)	(2)	(3)	(4)	(5)
Alabama	-0.52 (45)	-0.525 (45)	-0.541 (46)	-0.613 (45)	0.395 (45)
Alaska	0.08 (28)	0.083 (28)	0.103 (26)	0.063 (27)	0.523 (28)
Arizona	0.14 (25)	0.139 (25)	0.139 (25)	0.108 (25)	0.541 (25)
Arkansas	-0.72 (48)	-0.715 (48)	-0.736 (48)	-0.843 (48)	0.349 (48)
California	0.26 (22)	0.263 (22)	0.253 (22)	0.277 (22)	0.570 (22)
Colorado	0.55 (11)	0.547 (11)	0.573 (11)	0.575 (12)	0.636 (11)
Connecticut	0.82 (6)	0.817 (6)	0.829 (6)	0.900 (6)	0.695 (6)
Delaware	-0.06 (31)	-0.059 (30)	-0.051 (30)	-0.108 (31)	0.497 (31)
Florida	-0.14 (34)	-0.135 (34)	-0.135 (34)	-0.223 (34)	0.480 (34)
Georgia	-0.26 (36)	-0.258 (36)	-0.258 (36)	-0.295 (36)	0.460 (36)
Hawaii	0.98 (2)	0.979 (2)	1.065 (2)	1.189 (2)	0.739 (2)
Idaho	0.42 (17)	0.422 (17)	0.422 (17)	0.447 (17)	0.601 (17)
Illinois	-0.06 (30)	-0.060 (31)	-0.060 (31)	-0.096 (30)	0.502 (30)
Indiana	-0.34 (41)	-0.340 (41)	-0.340 (41)	-0.383 (39)	0.442 (39)
Iowa	0.30 (20)	0.299 (20)	0.299 (20)	0.326 (20)	0.578 (20)
Kansas	0.15 (24)	0.153 (24)	0.153 (24)	0.151 (24)	0.544 (24)
Kentucky	-0.47 (44)	-0.470 (44)	-0.507 (44)	-0.572 (44)	0.407 (43)
Louisiana	-0.94 (49)	-0.937 (50)	-0.955 (49)	-1.099 (49)	0.303 (49)
Maine	0.62 (9)	0.618 (9)	0.618 (9)	0.677 (9)	0.649 (9)
Maryland	0.34 (19)	0.334 (19)	0.338 (19)	0.366 (19)	0.584 (19)
Massachusetts	0.88 (4)	0.878 (4)	0.971 (3)	1.038 (3)	0.723 (3)
Michigan	-0.27 (37)	-0.268 (37)	-0.268 (37)	-0.350 (37)	0.453 (37)
Minnesota	0.82 (5)	0.821 (5)	0.786 (7)	0.895 (7)	0.688 (7)
Mississippi	-0.94 (49)	-0.936 (49)	-1.111 (50)	-1.279 (50)	0.273 (50)
Missouri	-0.40 (42)	-0.402 (42)	-0.402 (42)	-0.487 (42)	0.424 (42)
Montana	0.04 (29)	0.033 (29)	0.025 (29)	-0.032 (29)	0.513 (29)
Nebraska	0.51 (15)	0.512 (15)	0.512 (15)	0.569 (13)	0.625 (15)
Nevada	-0.28 (38)	-0.282 (38)	-0.321 (39)	-0.444 (41)	0.439 (41)
New Hampshire	0.90 (3)	0.897 (3)	0.925 (4)	1.011 (4)	0.716 (4)
New Jersey	0.64 (8)	0.647 (8)	0.647 (8)	0.689 (8)	0.655 (8)
New Mexico	-0.07 (32)	-0.067 (32)	-0.075 (32)	-0.166 (33)	0.486 (33)
New York	0.40 (18)	0.398 (18)	0.398 (18)	0.388 (18)	0.599 (18)
North Carolina	-0.10 (33)	-0.106 (33)	-0.106 (33)	-0.143 (32)	0.490 (32)
North Dakota	0.54 (12)	0.539 (12)	0.561 (12)	0.603 (11)	0.631 (12)
Ohio	-0.24 (35)	-0.247 (35)	-0.247 (35)	-0.264 (35)	0.465 (35)
Oklahoma	-0.46 (43)	-0.464 (43)	-0.470 (43)	-0.552 (43)	0.405 (44)
Oregon	0.53 (13)	0.526 (13)	0.526 (13)	0.538 (16)	0.625 (14)
Pennsylvania	0.10 (26)	0.103 (26)	0.103 (27)	0.104 (26)	0.537 (26)
Rhode Island	0.59 (10)	0.590 (10)	0.608 (10)	0.657 (10)	0.648 (10)

(Continued on next page)

Table A2 (continued)

State	Replication Number (1)	(2)	(3)	(4)	(5)
South Carolina	-0.54 (46)	-0.536 (46)	-0.537 (45)	-0.644 (46)	0.392 (46)
South Dakota	0.09 (27)	0.093 (27)	0.081 (28)	0.035 (28)	0.527 (27)
Tennessee	-0.32 (39)	-0.316 (39)	-0.288 (38)	-0.364 (38)	0.444 (38)
Texas	-0.33 (40)	-0.328 (40)	-0.331 (40)	-0.389 (40)	0.442 (40)
Utah	0.80 (7)	0.806 (7)	0.884 (5)	0.962 (5)	0.704 (5)
Vermont	1.20 (1)	1.196 (1)	1.225 (1)	1.366 (1)	0.778 (1)
Virginia	0.27 (21)	0.265 (21)	0.265 (21)	0.286 (21)	0.573 (21)
Washington	0.53 (13)	0.525 (14)	0.525 (14)	0.553 (14)	0.626 (13)
West Virginia	-0.66 (47)	-0.656 (47)	-0.676 (47)	-0.789 (47)	0.363 (47)
Wisconsin	0.49 (16)	0.487 (16)	0.497 (16)	0.545 (15)	0.622 (16)
Wyoming	0.24 (23)	0.232 (23)	0.222 (23)	0.183 (23)	0.557 (23)

Notes: Data from the 2012 version of America's Health Rankings and the U.S. Census. All rankings use the AHR weights. The numbers are the values of the summary function, or health levels, state health rankings are in parentheses. Each row shows the health levels and rankings for a state's health under different methodological assumptions:

(1): Original AHR Values from `http://americashealthrankings.org`.

(2): Replication of AHR values.

(3): Not truncating z-values to [-2,2].

(4): Using weighted means in place of unweighted means and median values and weighted standard deviations, where weights are based on 2012 state population, in calculation of z-values.

(5): Using the World Bank's HDI transformation instead of Z-Value transformation.

Table A3: Comparison of State Ranks with AHR Weights vs. Principal Component Weights

State	Rank for Value of Beta								
	-1	-0.5	-0.25	0	0.25	0.5	1	2	5
Alabama	45/46	45/46	44/45	41/44	35/42	35/39	33/34	26/32	21/22
Alaska	28/18	31/21	33/23	33/23	33/24	34/24	35/29	36/30	35/30
Arizona	25/22	25/22	24/21	20/21	20/21	21/20	27/19	31/28	32/31
Arkansas	48/47	47/47	46/46	45/46	42/47	38/45	38/45	38/45	38/45
California	22/15	23/16	25/20	32/26	39/36	41/40	41/41	41/41	41/41
Colorado	11/5	11/5	11/3	11/2	11/2	10/2	11/2	9/1	6/2
Connecticut	6/11	5/10	5/10	5/10	4/10	4/9	3/9	4/24	10/33
Delaware	31/30	32/31	30/31	27/31	26/30	27/27	25/28	25/25	24/26
Florida	34/32	33/32	31/32	28/32	25/29	24/26	21/23	17/16	13/12
Georgia	36/39	34/36	34/35	29/35	27/33	25/30	22/26	19/20	18/14
Hawaii	2/2	2/1	2/1	2/1	2/1	2/1	2/1	3/4	9/19
Idaho	17/17	19/18	19/22	26/27	37/37	39/41	40/42	40/42	40/42
Illinois	30/29	28/29	27/29	23/29	22/25	22/22	20/17	22/15	22/10
Indiana	39/41	38/41	37/39	36/39	31/38	31/36	34/32	35/31	36/34
Iowa	20/13	21/14	23/14	21/15	24/16	28/18	30/20	30/26	28/25
Kansas	24/28	22/28	20/28	19/28	19/26	18/23	24/22	32/29	33/32
Kentucky	43/45	46/45	47/47	49/48	49/49	49/49	49/49	49/49	49/49
Louisiana	49/49	49/48	49/48	46/47	44/44	37/42	37/37	34/34	30/28
Maine	9/20	9/19	9/18	9/18	9/17	9/15	9/15	8/13	7/11
Maryland	19/27	17/26	17/26	15/22	15/22	15/19	13/16	12/11	8/6
Massachusetts	3/6	3/6	3/6	3/5	3/3	3/3	4/3	2/9	5/21
Michigan	37/36	36/34	36/33	30/33	28/32	26/28	23/25	20/21	20/16
Minnesota	7/1	10/2	15/2	22/6	38/12	40/34	45/40	45/40	45/40
Mississippi	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50
Missouri	42/40	40/40	38/38	35/37	30/35	30/35	29/31	28/27	31/27
Montana	29/26	27/27	26/27	24/25	23/23	23/21	19/18	18/18	14/15
Nebraska	15/10	14/9	13/9	14/9	14/8	16/6	16/5	15/3	17/3
Nevada	41/34	44/39	45/41	48/41	48/43	48/46	48/47	48/47	48/47
New Hampshire	4/7	4/7	4/7	4/7	5/6	5/4	5/6	13/14	16/18
New Jersey	8/14	7/13	7/13	7/12	6/11	7/10	8/7	10/5	12/9
New Mexico	33/33	37/37	40/40	44/42	46/45	47/47	47/48	47/48	47/48
New York	18/24	16/23	14/19	13/19	12/15	12/14	10/13	7/7	4/5
North Carolina	32/37	30/35	28/34	25/34	21/31	20/29	18/27	23/22	25/23
North Dakota	12/8	15/8	16/8	17/8	17/7	19/8	28/8	33/6	34/8
Ohio	35/38	35/38	35/37	31/36	29/34	29/33	26/30	24/23	26/24
Oklahoma	44/44	43/43	43/43	42/43	36/41	36/38	36/35	37/35	37/35
Oregon	14/16	13/15	12/15	12/14	13/13	13/11	14/11	14/12	15/13
Pennsylvania	26/31	24/30	21/30	18/30	18/28	17/25	17/21	21/19	23/20
Rhode Island	10/21	8/20	8/17	8/17	7/18	6/17	6/24	6/36	3/36

(Continued on next page)

Table A3 (continued)

State	Rank for Value of Beta								
	-1	-0.5	-0.25	0	0.25	0.5	1	2	5
South Carolina	46/42	42/42	42/42	40/40	34/39	33/37	31/33	27/33	27/29
South Dakota	27/19	29/17	32/16	39/16	41/14	43/13	42/12	42/8	42/4
Tennessee	38/43	39/44	39/44	38/45	32/46	32/44	32/44	29/44	29/44
Texas	40/35	41/33	41/36	43/38	45/40	45/43	44/43	44/43	43/43
Utah	5/3	6/3	6/5	6/4	8/5	11/5	15/10	16/17	19/17
Vermont	1/4	1/4	1/4	1/3	1/4	1/12	1/36	1/37	1/37
Virginia	21/25	20/25	18/24	16/20	16/19	14/16	12/14	11/10	11/7
Washington	13/12	12/12	10/11	10/11	10/9	8/7	7/4	5/2	2/1
West Virginia	47/48	48/49	48/49	47/49	47/48	44/48	39/46	39/46	39/46
Wisconsin	16/9	18/11	22/12	34/13	43/20	46/32	46/39	46/39	46/39
Wyoming	23/23	26/24	29/25	37/24	40/27	42/31	43/38	43/38	44/38

Notes: Data from the 2012 version of America's Health Rankings. Each row shows the rankings for a state's health under different assumptions regarding β, the substitution parameter. Lower ranking numbers indicate better health, so the state with the best health receives a ranking of one. For each column, the first number shows the state's rank using AHR weights and the second number shows the state's rank using principal component weights.