Modeling Area-Level Health Rankings

Samir Soneji
Dartmouth Medical School

Rusty Tchernis
Georgia State University

This paper can be downloaded at: http://aysps.gsu.edu/working-papers.html

The Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract=2043214
Modeling Area-Level Health Rankings

Samir Soneji
Dartmouth Medical School

Rusty Tchernis
Georgia State University

This paper can be downloaded at: http://aysps.gsu.edu/uwrg-research.html
Modeling Area-Level Health Rankings

Samir Soneji, PhD
Rusty Tchernis, PhD

March 23, 2012

Number of Words: 3187
Number of Figures: 3
Number of Tables: 1

Acknowledgements: We thank Spencer Banzhaf, Rudy Benarjee, Heather Carlos, Joseph Hogan, John Kline, Valerie Lewis, Norris Cotton Cancer Center Shared GeoSpatial Resource, Richard Rothenberg, Angela Russell, Margaret Weden and Xiaoxi Zhao. This work was supported by funding from the National Cancer Institute (RC2CA148259). Authors are listed alphabetically and share authorship equally.

Affiliations: From the Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School and the Norris Cotton Cancer Center (SS). From the Department of Economics, Andrew Young School of Policy Studies, Georgia State University (RT), National Bureau of Economic Research (NBER), and Institute for the Study of Labor (IZA).

Address correspondence to: Samir Soneji, Norris Cotton Cancer Center, One Medical Center Drive, Lebanon NH, 03766; tel. 603-653-3908; fax 603-653-0896; email samir.soneji@dartmouth.edu.
Abstract

Objectives: To propose a model for estimation of area level health rankings and to compare it to existing measures.

Methods: Area-level measures of health are estimated using factor analysis with spatial correlation. The contribution of each variable to the resulting rank is empirically derived and population sizes are incorporated into a measure of uncertainty. The model is applied to county level data from Texas and Wisconsin; results are compared with ranks from the County Health Rankings methodology.

Results: In the case of Wisconsin we find few disagreements between the two ranking methods, but our model results in a coherent measure of uncertainty in rank. In the case of Texas we report much greater differences in the resulting rankings. We also observe a strong suburban-urban ranking gradient within metropolitan statistical areas.

Conclusions: Area-level health is not directly measurable, but can be summarized by applying statistical methods to observed health, social, and environmental variables to produce rankings. Different methods may produce similar or dissimilar rankings. These differences should be reconciled before policy decisions are made.
1 Introduction

In the field of population health, a central interest is the evaluation of health outcomes of communities and assessment of how these health outcomes vary within and among communities.\textsuperscript{1} Community-level assessments that consider multiple determinants of health enable local, state, and federal governments to identify disadvantaged and under-served populations and refocus priorities.\textsuperscript{2} Assessments that rank population health at the local or state level are an especially effective tool for communication, evaluation, and prioritization.\textsuperscript{3,4} Two widely used rankings are the state-level \textit{America’s Health Rankings} and the county-level \textit{County Health Rankings}. Both collect geographically specific information on health outcomes and health determinants. The initiatives help to increase awareness of differences and mobilize action within communities to improve population health.

In both health rankings, geographic areas (i.e., states and counties) are ranked on health outcomes and health determinants. Geographic areas represent more than political boundaries; they are also manifestations of the physical, social, and economic environments that shape the health of populations residing in these areas.\textsuperscript{5} With this framework in mind, both health rankings consider numerous variables that affect health outcomes and determinants including access to health care, health behaviors, socioeconomic status, and the physical environment.\textsuperscript{6} The rankings draw from a rich set of data sources including health and population surveys, environmental assessments, vital statistics, death certificates, crime reports, and business patterns. While comprehensive, both rankings share key methodological and substantive limitations. First, neither health rankings determine whether there is a meaningful difference in the ranking of two counties or the ranking of two states. Second, both health rankings fix subjectively assessed deterministic weights of each variable rather than allow the weights to be empirically derived. Intuition-based
approaches, such as these, regularly underperform formal statistical procedures even when conducted by the most well-trained experts. Finally, neither health ranking accounts for the population of the geographic area and uncertainty in the estimation of model parameters.

In this paper, we utilize an alternative Bayesian factor analysis model for estimating county-level health ranks, which may be an improvement over current methods. As an example, we apply our method to the County Health Rankings, developed by the University of Wisconsin Population Health Institute (UWPHI), and rank health outcomes in Texas (TX: 254 counties) and Wisconsin (WI: 72 counties). We directly incorporate county population measures, estimation uncertainty, and spatial autocorrelation to produce a probability interval for each county’s health ranking, rather than a single rank. We also offer guidance about the use of our approach that will enable scholars to produce additional rankings utilizing other health, behavioral, social, and demographic data.

2 Methods

2.1 Statistical Analyses

Before we present the methodology suggested in the paper, it is useful to understand how the existing rankings are calculated. UWPHI calculates an overall health score based on standardized mortality and morbidity variables and their corresponding deterministic weights. UWPHI first transforms the value of each mortality and morbidity variable into its corresponding z-score based on the distribution of values within the state. Next, the z-scores are multiplied by their corresponding deterministic weight. Finally, UWPHI sums over the weighted z-scores to create a final score for each county. This score is the basis for the county health rankings. In this paper we concentrate on health outcomes rankings.
In our methodology, we incorporate a factor analysis framework that assumes measurable variables (e.g., mortality and morbidity variables) are manifestations of an underlying latent construct (e.g., health).\textsuperscript{9} The traditional factor analysis model explains the variability in observed variables $Y_{ij}$ for county $i$ in the following way:

$$Y_{ij} = \mu_j + \lambda_j \delta_i + e_{ij},$$

where $\mu_j$ is average of variable $j$ across counties, $\delta_i \sim N(0, 1)$ represent latent health level for county $i$, $\lambda_j$ is the factor loading for variable $j$ and represents the covariance between the latent health and the observed variables, and $e_{ij} \sim N(0, \sigma_j^2)$ are the idiosyncratic error terms. Thus, the model assumes that the observed variables are jointly determined by the underlying latent health construct.

We follow methodology developed in Hogan and Tchernis (2004), which extends the traditional factor analysis model and allows for the factors to be spatially correlated.\textsuperscript{10} Thus, the factor in county $i$ is correlated with a factor in other counties and is influenced not only by observed variables in county $i$, but also observed variables in other counties. In addition, the model allows the variance of the factors and the residual variance to be inversely proportional to the population size of each county. Thus, more populous counties will have lower variance. The Bayesian model is estimated using the Metropolis-Hastings algorithm within Gibbs Sampler.\textsuperscript{11,12} We replaced missing observations with ordinary least squares predictions based on observed variables. We illustrate the performance of this model using the UWPHI data applied to county health rankings in TX and WI. We selected WI and TX as our two example states because the former served as the focus of the precursor to the \textit{County Health Rankings}, the Wisconsin County Health Rankings.\textsuperscript{13} The later is a populous state, includes several of the largest metropolitan areas in the US, and has the largest number of counties – our unit of analysis.
We used Matlab Version R2009a for all statistical analyses. The Dartmouth College and Dartmouth-Hitchcock Medical Center Committee for the Protection of Human Subjects and Georgia State University Institutional Review Board determined this research met eligibility criteria for review exemption.

2.2 Data

Mortality and morbidity data were downloaded on June 1, 2011, from www.countyhealthrankings.org, the website maintained by UWPHI. The mortality variable used in the County Health Rankings is the years of potential life lost before age 75 years ("premature death"), originally estimated using 2005-2007 life table data from the National Center for Health Statistics (NCHS). The morbidity variables used are: 1) the percent of adults reporting fair or poor health ("self-reported health"), 2) the mean number of physically unhealthy days per month for adults ("physical unhealthy days"), 3) the mean number of mentally unhealthy days per month for adults ("mental unhealthy days"), and 4) the percent of live births with birthweight < 2500 grams ("low birthweight"). The first three morbidity variables were originally estimated using 2003-2009 data from the Behavioral Risk Factor Surveillance Survey and the fourth morbidity variable was estimated using 2001-2007 birth certificate data from NCHS.

UWPHI did not rank 31 out of 254 counties in TX because of insufficient data. In addition, 51 of the remaining 223 counties had missing data on at least one variable. UWPHI ranked all 72 counties in WI and 2 of the counties had missing data on at least one variable. To assign ranks to these counties with partially missing data, UWPHI replaced the missing variable by the state average of that variable. We estimate health ranks for the counties also ranked by UWPHI.
3 Results

We estimate the model using data for TX and WI separately. For each county, we compute the posterior distribution of its health rank, including its mean and 95% probability interval.

3.1 Disagreement and Uncertainty

In Figure 1, we illustrate the differences between UWPHI and our methods. We plot the middle 95% of the posterior distribution of ranks (horizontal line) and the mean of the posterior distribution (closed circle) relative to each county’s UWPHI rank. Comparing the panels for TX and WI we can see that there are more disagreements in TX than in WI – the intervals in WI panel are much closer to the 45 degree line. The correlation between UWPHI ranks and our ranks was 0.65 for TX (95% CI, 0.60 to 0.71) and 0.89 for WI (95% CI, 0.81 to 0.94).

To investigate possible reasons for this disparity, we compare the relationship between the observed variables and the underlying health measure. Table 1 compares UWPHI derived weights and normalized square correlations for the mortality and morbidity variables. Squared correlation represents the proportion of the variance in the variable that is explained by the factors and thus has similar information to the weights used by UWPHI.\(^1\) Our normalized square correlations differ between TX and WI. Additionally, the squared correlations differ from the UWPHI weights, which are applied uniformly across all states.

For example, we estimate the squared correlation of the mean number of physically un-

\(^1\)In addition, we estimated regression coefficients by regressing the mortality and morbidity measures on the posterior means of factors. These regression coefficients are similar to the UWPHI weights. The information in these regression coefficients was very similar to squared correlations and they are available upon request from the authors.
Figure 1: UWPHI Rank, Mean Posterior Ranks, and 95% Probability Intervals

The left (right) panel shows the the posterior rank and UWPHI rank for each county in TX (WI). The 95% probability interval of the posterior is denoted by a horizontal line and mean posterior rank is denoted by a solid circle. The gray horizontal and vertical lines represent the 80th percentile of ranks and the 45° line represents equality between the UWPHI and posterior ranks.
healthy days per month to be 0.41 for TX (95% CI, 0.34 to 0.48) and 0.21 for WI (95%, 0.11 to 0.31). UWPHI sets the weight of this variable to 0.10 for all states. The difference between squared correlations and UWPHI weights can explain why our ranks are more similar to UWPHI ranks in WI compared to TX.

Table 1: UWPHI Deterministic Weights and Normalized Square Correlations

<table>
<thead>
<tr>
<th>Category</th>
<th>Health Outcomes</th>
<th>UWPHI</th>
<th>Texas 95% CI</th>
<th>Wisconsin 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>Premature Death</td>
<td>0.50</td>
<td>0.14 (0.09,0.19)</td>
<td>0.27 (0.17,0.38)</td>
</tr>
<tr>
<td>Morbidity</td>
<td>Self-Reported Health Status</td>
<td>0.10</td>
<td>0.24 (0.20,0.29)</td>
<td>0.21 (0.12,0.30)</td>
</tr>
<tr>
<td></td>
<td>Physically Unhealthy Days</td>
<td>0.10</td>
<td>0.41 (0.34,0.48)</td>
<td>0.21 (0.11,0.31)</td>
</tr>
<tr>
<td></td>
<td>Mentally Unhealthy Days</td>
<td>0.10</td>
<td>0.15 (0.10,0.20)</td>
<td>0.17 (0.08,0.25)</td>
</tr>
<tr>
<td></td>
<td>Low Birthweight Births</td>
<td>0.20</td>
<td>0.06 (0.02,0.10)</td>
<td>0.15 (0.05,0.24)</td>
</tr>
</tbody>
</table>

Note: \(w\)=weight; \(\rho^2\)=squared correlation; CI=confidence interval.

An additional important question that we address is how far apart the ranks of two counties be to give a researcher reasonable confidence to conclude that they are different, for example 90% and 95%. To answer this question we calculate the percentage overlap in the posterior distribution of ranks between two counties that are \(k\) units apart in their mean rank. Consider Harris County, TX with mean rank equal to 66. The mean rank of Blanco County, TX equals \(k = 5\) ranks higher at 71; 2.2% of the posterior distribution of Harris County overlaps with the posterior distribution of Blanco County. As the difference between mean ranks, \(k\), increases, the median percentage overlap decreases. For example, as \(k\) increases from 1 to 5 to 10, the median overlap in the posterior distributions equals 40.8%, 12.8%, and 1.3%. When \(k\) exceeds 25, 90% of county pairs have 1% or less overlap. Similarly, when \(k\) exceeds 30, 95% of county pairs have 1% or less overlap. In WI, as \(k\) increases from 1 to 5 to 10, the median overlap in the posterior distributions equals 41.7%, 13.5%, and 0.5%. When \(k\) exceeds 21, 90% of county pairs have 1% or less overlap and when \(k\) exceeds 23, 95% of county pairs have 1% or less overlap. Thus, for a researcher...
to be reasonably confident, say 90%, that two counties are difference with respect to their
health ranking, the distance between the mean of their health rank distribution should be
approximately 25 counties apart in TX (10% of the number of counties) and 21 counties
apart in WI (29% of the number of counties).

3.2 Least Healthy Counties

An important purpose of county health rankings is to identify the least healthy counties to
more effectively mobilize action to improve the health of these populations. The vertical
and horizontal lines in Figure 1 represent the 80th percentile (178th for TX and 62nd for
WI) separating the least healthy quintile of counties. According to our model, the least
healthy counties will be those counties with a mean of the rank distribution to the right of
the vertical line. According to the UWPHI model, the least healthy counties will be those
with ranks above the horizontal line. In TX, 26 counties are classified as least healthy by
both models, 19 are classified as least healthy by our model and not UWPHI, and 19 are
classified as least healthy by UWPHI and not our model. In WI, 13 counties are classified
as least healthy by both models, 3 are classified as least healthy by our model and not
UWPHI, and 2 are classified as least healthy by UWPHI and not our model. Considering
only the counties classified as least healthy either by our model or UWPHI, the correlation
between UWPHI ranks and our ranks was -0.06 for TX (95% CI, -0.17 to 0.14) and 0.68
for WI (95% CI, 0.34 to 0.85).

Figure 2 presents county-level maps of TX and WI and indicates the probability of a
county being in the least healthy quintile. According to the UWPHI model, counties either
are or are not in this quintile. In contrast, each county in our model has a probability of
being in the least healthy quintile, which is reflected in increasing shades of gray. In TX, we
observe a large concentration of unhealthy counties in East Texas. In TX, the probability
of being in the least healthy quintile is 1.00 for Austin County\textsuperscript{2}, 0.63 for Colorado County, and 0.59 for Polk County. In other words, the entire posterior distribution of rankings for Austin County lies above the 80th percentile. In WI, the probability is 1.00 for Milwaukee County and also high for several of the northernmost counties.

Figure 2: Probability of Being in Least Health Quintile, TX and WI

![Map of Texas and Wisconsin showing probability distribution]

The left (right) panel shows the probability of being in the least healthy quintile under our model for Texas (Wisconsin). Counties not ranked are dotted.

3.3 Metropolitan Statistical Areas

We also consider differences in health rankings within metropolitan statistical areas (MSAs).\textsuperscript{14} In Figure 3, we plot the posterior rank of counties in the largest MSAs in WI and TX. For each county, we create a boxplot to reflect the posterior distribution’s minimum and maximum (horizontal line), middle 50% (grey box), and median (vertical line). The corresponding UWPHI rank is indicated as an open circle. We observe a suburban-urban gradient in health rankings in several MSAs. In the Milwaukee MSA, counties adjacent

\textsuperscript{2}Austin County is located southeast of Houston, TX. Austin, TX is located in Travis County.
to Milwaukee County (Waukesha, Ozaukee, and Washington) rank healthier and they are among the healthiest in the state. A similar conclusion may be drawn in the Milwaukee MSA with UWPHI ranks, although the ranks lack uncertainty. In the Dallas-Ft. Worth-Arlington MSA, several counties north of Dallas County rank healthier and Denton, Collin, and Rockwall are among the healthiest counties in the state. In contrast, counties south of Dallas County rank less healthy (Johnson, Ellis, and Kaufman). We observe more agreement among the healthiest counties between our ranks and UWPHI ranks and more disagreement among the least healthy counties. For example, we rank Johnson County at the 85th percentile (95% CI, 82nd to 86th percentile) compared to the 36th percentile by UWPHI.

4 Discussion

This study has three main findings. First, we propose a new approach to county health rankings using a factor analysis model, which has a number of appealing properties including data-driven weights and incorporation of population sizes, spatial correlation, and statistical uncertainty. Second, we report health ranking results for TX and WI and find that the weights are different between states. Consequently, our approach and UWPHI had a lot of disagreement in TX, but not in WI. However, for TX and WI, we also estimate the statistical uncertainty in each county’s health ranking, which is important when assessing whether there is a meaningful difference in the health ranks of two counties. For example, counties within 25 ranks of each other in TX (21 ranks in WI) are not statistically different at the 90% confidence level. Third, we provide a number of useful ways to summarize our findings — computing the probability of being in the least healthy quintile and comparing urban and suburban areas within MSAs.
Each panel shows a boxplot of the distribution of posterior ranks of counties in select metropolitan statistical areas. UWPHI rank is shown as an open circle (◦). For each metropolitan statistical area, the central urban county is written in bold.
The core motivation of our framework is a belief that health is a latent construct and observed variables are its manifestations. In contrast, the framework of *County Health Rankings* creates a construct called health that explicitly consists of 50% mortality and 50% morbidity (split among four variables). One important consequence of our framework is that the correlation of observed variables with one another may be different among states. We observe this pattern when comparing empirically derived squared correlations for WI and TX and compare them to the UWPHI subjectively assessed weights. Our rankings and UWPHI rankings of WI counties shares more agreement than of TX counties because the squared correlations for WI are more similar to UWPHI weights than for TX. Additionally, the large number of small TX counties (26% less than 10,000 residents) introduces greater uncertainty.

Consumers of rankings – policy makers, public health researchers, and community leaders – may use rankings for different purposes. The standard by which two counties are determined to be different will depend on this specific purpose. Our method, which yields a distribution of ranks for a given county, may be adopted to various standards of confidence.

Our examination of differences in health rankings within MSAs addresses disparities defined by both demographic characteristics and geographic location. Eberhardt and Pamuk (2004) investigate evidence of rural health disadvantage and find differences in mortality and morbidity among rural, suburban, and urban areas. We observe larger differences within MSAs, between the central urban county and its surrounding counties, compared to differences between metropolitan and non-metropolitan areas.

Population health rankings have been conducted annually at the county level since 2010 and at the state level since 1990. Substantively, we find concentrated areas of poor health in East Texas and Southeast and Northern Wisconsin, as does the *County Health Rankings*. 
Rankings. These areas are predominantly poor, rural, or medically underserved. Recent work has identified specific pathways that link these and other determinants to poor health outcomes.\textsuperscript{17–19} Methodologically, both \emph{America’s Health Rankings} and the \emph{County Health Rankings} consider numerous mortality, morbidity, and health variables and produce a single deterministic rank. We expand on this work by estimating statistical uncertainty of ranks and producing a probability interval, which can be applied at the county or the state level.

We acknowledge several limitations in this study. First, we only rank counties with at least one morbidity or mortality variable measured. Counties without any measured variables, which are often the smallest, may also be among the most disadvantaged and least healthy. Second, our imputation method of replacing missing variables with their least squares prediction did not account for the uncertainty of the imputation. We also ranked county health by following the UWPHI practice of replacing missing values with that variable’s state average, which also did not account for imputation uncertainty, and came to no substantively different conclusions. Third, several morbidity measures are based on self-reported physical and mental status, which may be subject to differential recall bias.

While rankings are useful to compare the health of populations, they do not convey absolute differences in health. A county may improve in health ranking even though its population became less healthy if, for example, the health of other counties declined faster over time. Depending on the research or policy goal, progress may be better measured using absolute measures of population health.

Whether we use a relative or absolute measure of population health, further research is needed to determine how medical, behavioral, social, physical, and biological determinants interact to produce health and perpetuate disparities.\textsuperscript{20} A larger nationwide ranking of
county health may reveal how the least healthy areas coincide with areas of greatest social
disadvantage and extend beyond state borders. Our ranking methodology may also be
extended to international comparisons of health system performance\textsuperscript{21,22} or human develop-
ment.\textsuperscript{23} An additional application could focus on ranking health and healthcare across
hospitals, hospital referral regions, and physician hospital networks to identify optimal
and suboptimal areas.

Equally important as variation in health among populations is the variation within
specific populations. Neither UWPHI nor our methodology explicitly incorporates the level
of variation in a health variable within a given county. Through spatial autocorrelation,
our methodology does allow each county to be correlated with one another. Future work
could subset counties and incorporate the health of specific neighborhoods within counties.
Additionally, future work could utilize our methodology and add measures of variability
for each mortality and morbidity variable by county (e.g., the standard error of the sample
mean of a health variable). These additional data could be easily incorporated into a new
model using our framework, the weights would be re-estimated analytically without re-
eliciting expert opinion, and the resulting ranks would reflect variation in health variables
among populations and within populations.

In conclusion, public health researchers consider a broad range of health, social, eco-
nomic, and environmental determinants when assessing population health and identifying
areas of greatest need.\textsuperscript{1,24,25} Population health assessments are often presented to poli-
cymakers and communities as ranks, given their ubiquity and ease of interpretation. Our
framework produces probability intervals for these ranks and allows direct probabilistic
comparison of geographic areas.
References


