

NBER WORKING PAPER SERIES

MEASURING THE SPATIAL DISTRIBUTION OF HEALTH RANKINGS IN THE
UNITED STATES

Will Davis
Alexander D. Gordan
Rusty Tchernis

Working Paper 27259
<http://www.nber.org/papers/w27259>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020

We would like to thank Charles Courtemanche, Ian McDonough, Erik Nesson, workshop participants at IUPUI, and participants at the 2019 Western Economics Association International annual conference. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Will Davis, Alexander D. Gordan, and Rusty Tchernis. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Measuring the Spatial Distribution of Health Rankings in the United States
Will Davis, Alexander D. Gordan, and Rusty Tchernis
NBER Working Paper No. 27259
May 2020
JEL No. I1,I14

ABSTRACT

We rank counties in the United States of America with respect to population health. We utilize the five observable county health variables used to construct the University of Wisconsin Population Health Institute's County Health Rankings (CHRs). Our method relies on a factor analysis model to directly compute weights for our rankings, incorporate county population sizes into the variances, and allow for spillovers of health stock across county lines. We find that demographic and economic variation explain a large portion of the variation in health rankings. We address the importance of uncertainty caused by imputation of missing data and show that the use of rankings leads to inherently greater uncertainty. Analyzing the health of counties both within and across state lines shows substantial degrees of disparity. We find some disagreement between our ranks and the CHRs, but we show that much can be learned by combining results from both methods.

Will Davis
Georgia State University
55 Park Place NE
Atlanta, GA 30303
gdavis29@student.gsu.edu

Alexander D. Gordan
Department of Economics
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992
zander.gordan@gmail.com

Rusty Tchernis
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992
and NBER
rtchernis@gsu.edu

1. Introduction:

Over time, both the prevalence and importance of rankings have grown considerably. Whether designed to compare universities, products, cities, or nations, rankings produced using a wide array of inputs regularly influence the decisions and opinions of individuals, firms, and governments (Marginson and van der Wende 2007, Giffinger and Gudrun 2010, Wolff et al. 2011, Qui et al. 2018). Rankings are generally well suited for combining complicated outcomes into a single easily understood measure that is comparable across units. This approach is especially important in cases where a quality of interest is difficult or impossible to observe using a single variable.

A prime candidate for the use of rankings is the measurement of county-level health. Given health's multidimensional nature, researchers and policymakers often rely on rankings to combine information from multiple sources into a tractable outcome. For health at the county-level, rankings may compel individuals within counties to take more active roles in improving the health of their communities. Policymakers can also use these rankings directly or as a framework for understanding which interventions prove effective in improving the health of a county's residents.

While county health rankings are important to policymakers, researchers, and other interested parties, creating them is challenging for two primary reasons. First, no single observable outcome captures county health. To overcome this restriction, a consistent method for combining information from different inputs is needed. The second difficulty is the need to account for uncertainty in the rankings. There may be considerable levels of uncertainty for smaller counties which often have less reliable data. Furthermore, since many observed health variables have

large amounts of missing data, uncertainty in the final rankings should incorporate error caused by the noisy imputation procedures used to replace missing values.

The most commonly used rankings of county health come from the University of Wisconsin Population Health Institute's (UWPHI's) County Health Rankings (CHRs). The CHRs produce rankings of county health for each state which have been used both directly and indirectly by local policymakers concerned with crafting interventions to improve the health of their communities.¹ While the CHRs are easy to calculate and interpret, there are several potential concerns with their design. First, the CHR of each county is calculated using a weighted average of five normalized manifest variables, the weights for which are subjectively determined rather than estimated using data. These deterministic weights rely on the assumption that each variable's relative contribution to the measure of county health is known. The second issue is that the CHRs do not incorporate uncertainty into their final rankings even though a significant number of missing observations are imputed. Without a measure of uncertainty, knowing whether differences in ranks between counties are statistically significant or not is impossible. Finally, the CHRs only rank counties within the same state, making it impossible to examine the relative health of counties across state lines. While certain questions of interest involve within-state comparisons, there are many cases where a set of national rankings is required.

In this paper, we use a Bayesian spatial factor analysis model to estimate rankings of county health. Our approach has several advantages. First, the variable weights used in our model are estimated directly. By estimating factor weights, we allow the data to determine the relative contribution of each variable to our health measure rather than relying on expert opinion.

¹ For examples of how the CHRs have been used in policy, see <http://www.countyhealthrankings.org/explore-health-rankings/use-data/communities-using-rankings-data>.

Second, we summarize the posterior distribution of health rankings for each county instead of providing a single point estimate. With a distribution of rankings, we can evaluate uncertainty regarding the rank and relative health of each county. Third, we allow for spatial correlation in the variance of our rankings to capture potential health spillover effects across county borders. The variance of our estimates also changes inversely with county population, capturing the increased uncertainty of data sampled from less populated counties. Finally, our model directly incorporates the estimation error produced by imputing missing data into the overall uncertainty of our estimates.

We estimate our model with the data used to calculate the CHR_s in 2015. Rather than restricting our results to within-state rankings, we estimate county health ranks jointly for the entire nation. Our study builds on the work of Hogan and Tchernis (2004) who use the spatial factor analysis model to produce census-tract-level rankings of material deprivation in Rhode Island, and Courtemanche et al. (2015) who use the CHR variables to estimate within-state health rankings for counties in Texas and Wisconsin. In addition to addressing the technical limitations of other rankings, our results allow for the direct comparison of county health across state lines. Comparing the health of all counties to one another allows us to discern essential characteristics of county health throughout the country.

We find significant differences between the manifest variable factor weights estimated by our model and the deterministic weightings used by the CHR_s. Our model assigns less weight to early mortality and low birthweight and more weight to outcomes of physical and mental health among adults. Our mean posterior health rankings generally agree with those of the CHR_s, but there is more disagreement between the two measures towards the middle of the rank distribution. The level of agreement between both rankings suggests a higher level of certainty

regarding the relative health of the most and least healthy counties versus those of average health.

Looking at our rankings of county health across the nation, we find a large degree of health disparity both within and across states. While some states have relatively tight distributions of county health, many contain a disparate mix of healthy and unhealthy counties. We also find that the healthiest counties in some of the unhealthy states fall well below the average health level for counties in high performing states. Using data from Chetty and Hendren (2018), we show that a county's level of adult economic mobility among children born to low-income households has a significant impact on its health ranking. Furthermore, we find that the effect of economic mobility and race covary with one another, suggesting a complex relationship between a county's racial makeup, economic mobility, and population health.

We also show that there is a substantial degree of uncertainty in our model's rankings, implying that any single set of rankings represents only one of many plausible orderings. The level of uncertainty varies considerably across counties, however. For example, we are more certain of relative health among the set of counties in the bottom 10% of health according our model and the CHRs. We find that most of these "bottom 10%" counties are in one of five states, further illustrating the underperformance of certain areas when compared to the national distribution.

Finally, we examine causes of the uncertainty in our rankings as well as methods which can reduce the overall level of uncertainty. We show that the amount of uncertainty can be decreased for most counties using alternative imputation procedures that ignore imputation error in the final estimation. Furthermore, we find that the use of rankings rather than directly using estimates of

latent county health drives a substantial portion of the uncertainty in our results. Given that rankings inflate the level of uncertainty in a county's health, it is even more concerning that traditional measures like the CHRs do not include a measure of uncertainty.

The remainder of the paper is structured as follows. Section 2 discusses methodology underlying the CHRs and our model. Section 3 provides an overview of our approach to imputing missing data. In Section 4, we discuss the data used in the study. Section 5 presents our results and Section 6 concludes.

2. *Methods:*

We begin with an overview of the University of Wisconsin Population Health Institute's (UWPHI's) County Health Rankings (CHR). Rather than jointly ranking all counties in the nation, the UWPHI produces separate sets of CHRs for each state. Creating these state-level rankings begins by choosing which counties to rank as the UWPHI does not produce CHRs for certain counties due to data concerns. Among the ranked counties, any missing observations are imputed using the state mean. Each outcome variable is then normalized as z-scores. A weighted average of these z-scores is used as the health score for each county. Finally, within-state health rankings are produced by ordering the health scores of each county.²

Our model builds on the assumption that the CHR's five observed variables are manifestations of a single latent construct - county health. In contrast to the CHR's deterministic weights, we use data to estimate the strength of the relationship between each manifest variable

² For additional information regarding the CHR's methodology, see <https://www.countyhealthrankings.org/explore-health-rankings/our-methods/rankings-overview>

and latent health. We assume a linear relationship between the latent health variable and observable outcome variables with a Gaussian error structure such that $Y_i = \mu + \lambda\delta_i + e_i$, where Y_i is a vector of the observed outcome variables for county i , μ is a vector of outcome variable specific means, λ is a vector of outcome specific factor loadings, δ_i is the scalar-valued latent health of county i , and e_i is a vector of Gaussian distributed idiosyncratic error terms. We further assume that the error terms are uncorrelated both within- and across-counties, implying that $Var(e_i) \equiv \Sigma = diag(\sigma_j^2)$, where j indexes the set of observed manifest variables. Stacking over all n counties, the model can be written in hierarchical form as:

$$Y|\delta \sim N(\iota_n \otimes \mu + \Lambda\delta, I_n \otimes \Sigma), \quad \delta \sim N(0, I_n \otimes \mu)$$

where $\Lambda = I_n \otimes \lambda$, ι_n is an $n \times 1$ vector of 1's, and I_n is an $n \times n$ identity matrix.

We next alter our model to incorporate uncertainty from population size. We assume that the variances of our latent factor and idiosyncratic errors are both decreasing in county population. Defining $M = diag(m_i)$, where m_i is the population of county i , the model is written as:

$$Y|\delta \sim N(\iota_n \otimes \mu + \Lambda\delta, M^{-1} \otimes \Sigma) \quad \delta \sim N(0, M^{-1})$$

The intuition behind this specification is that larger counties have less sampling error in their manifest variables and are less likely to have populations on the extreme ends of the health distribution relative to smaller counties.

We next add spatial correlation into the model to produce our final specification. The presence of spatial spillovers implies that the health of each county is partially determined by the health of its neighboring counties, a feature we incorporate with spatial dependence at the factor level. With the addition of a spatial correlation matrix Ψ , the model is specified as:

$$Y|\delta \sim N(\iota_n \otimes \mu + \Lambda \delta, M^{-1} \otimes \Sigma) \quad \delta \sim N(0, M^{-1/2} \Psi M^{-1/2})$$

We model spatial dependence in the latent health factor as a conditionally autoregressive process in which a county's health is determined by the mean health of its neighbors. We define county i 's neighbors as the set of counties R_i which are geographically adjacent to it. Latent health for county i conditional on the health of its neighbors is then modeled such that:

$$\delta_i | \{\delta_j : j \in R_i\} \sim N\left(\sum_{j \in R_i} \omega \delta_j, 1\right)$$

This conditional autoregressive model yields the spatial correlation matrix $\Psi = (I_n - \omega R)^{-1}$, where ω is a spatial correlation parameter to be estimated and R is the neighborhood adjacency matrix with elements $r_{ij} = r_{ji} = 1$ if counties i and j are adjacent, and 0 otherwise (including that $r_{ii} = 0$).

Using a Bayesian approach to estimate our model, we also specify prior distributions for our parameters. We use a set of conjugate non-informative priors which simplify the derivation of the conditional posterior distributions without contributing significant information to our estimates. This specification implies that the posterior distributions of our model are informed almost entirely by the data rather than the prior distributions. The posterior distribution of model parameters and county ranks are then estimated using Markov Chain Monte Carlo methods (Chib and Greenberg 1996). More specifically, we use a Gibbs sampling algorithm (Gelfand and Smith 1990) with a Metropolis-Hastings step (Chib and Greenberg 1995) for the estimation of ω . The details of our prior distributions and estimation algorithm are discussed in Hogan and Tchernis (2004).

3. Missing Data:

The CHRs deal with missing data in two ways: not ranking certain counties with missing data and imputation. Given that we would like to lose the fewest possible number of counties, we choose to impute missing data for all counties with at least one non-missing outcome variable.

While there are many ways to impute data, the CHRs replace missing observations with state-level means. There are two primary concerns with this approach. First, using state-level means does not incorporate the inherent uncertainty of the imputation process into the final rankings. The second issue is that rankings for counties with more missing data are likely to be biased if they are systematically different from counties with little to no missing data.

As an alternative to state-level means, we use a posterior imputation procedure. Our posterior imputation method consists of adding one step to our Gibbs sampling routine wherein we sample missing manifest variable observations using the first level of our model conditional on the current iteration's draws of latent health and parameter estimates. This approach allows us to holistically incorporate uncertainty from missing data into the posterior distributions of our final rankings. We evaluate the impact of different imputation procedures on our estimates in Section 6 using state averages and Predictive Mean Matching (PMM).

4. Data:

We use the set of CHR manifest variables in 2015 which includes one mortality measure and four morbidity measures. The mortality measure used is years of potential life lost before age 75 (YPLL) which is compiled by the National Center for Health Statistics (NCHS) using the 2010-2012 mortality files. Three morbidity measures come from the Behavioral Risk Factor Surveillance System Survey (BRFSS) using 2006-2012 data: percentage of adults reporting fair

or poor health (FP%), average number of reported physically unhealthy days (PUD), and average number of reported mentally unhealthy days (MUD). The final morbidity variable is the percentage of live births with low birth weight (LBW) which comes from the 2006-2012 NCHS natality files.

Among the set of all 3,142 U.S. counties and county equivalents, 79 are not officially ranked by the CHRs. The three states with the most unranked counties are Texas with 17, Nebraska with 15, and Montana with 10. Of the officially ranked counties, 142 are missing YPLL rate, 400 counties are missing FP%, 334 are missing PUD, 552 are missing MUD, and 99 are missing LBW. Given the considerable amount of missing data, it is especially important to consider how noisy imputation procedures influence the overall uncertainty of relative health and its subsequent ranking.

5. Results:

We now present our results. First, we discuss the factor loadings and normalized squared correlation coefficients of the five manifest variables. Second, we examine the levels of across- and within-state variation in our health rankings. Third, we estimate regressions of our posterior mean health rankings and each of the five manifest variables on a set of socioeconomic outcomes to see which observable factors best explain county health. Fourth, we compare our data-driven posterior mean rankings to a set of rankings produced using the CHR's methodology and discuss how well both measures agree with one another. Fifth, we evaluate the level of uncertainty present in our primary results. Sixth, we examine the set of counties in the bottom 10% of health according to the ranks produced by our model and the CHRs. Seventh, we analyze how the amount of uncertainty changes under different imputation strategies. Finally, we discuss the

increase in uncertainty inherent to the use of rankings rather than directly using estimates of underlying health.

Factor Loadings and Normalized Squared Correlation Coefficients:

Table 1 shows the posterior mean and 95% confidence interval of each manifest variable's factor loading, residual standard deviation, and normalized squared correlation coefficient. Following Hogan and Tchernis (2004), the normalized squared correlation coefficients signify the proportion of variation in latent county health that is explained by each manifest variable, giving us a comparable measure to standard index variable weights.

Looking at the mean posterior factor loadings of our five manifest variables, we find that the rate of fair/poor health and years of potential life lost rate have the largest and second largest loadings, respectively, while the factor loadings for rate of low birth weight, physically unhealthy days, and mentally unhealthy days are substantially smaller. When considering the posterior standard deviations of each residual, however, we see that years of potential life lost rate, number of mentally unhealthy days, and rate of low birth weight have posterior mean residual standard deviations larger than their mean factor loadings. The relative magnitudes of both the mean factor loading and residual standard deviation features directly into our normalized squared correlation coefficients. We find that rate of self-reported fair/poor health and number of physically unhealthy days explain the largest shares of variation in latent county health at 0.28 and 0.27, respectively. Both years of potential life lost rate and number of mentally unhealthy days have mean normalized squared correlation coefficients of 0.18 while rate of low birth weight explains the smallest portion of variation in latent health at 0.08.

Comparing our normalized squared correlation coefficients in Table 1 to the weights used by the CHR's, we find a considerable amount of disagreement regarding the relative contribution of each manifest variable. While years of potential life lost rate is assigned the weight of 50% in the CHR's, we find that it captures only 18% of the variation in latent county health according to our model. Furthermore, self-reported rates of fair/poor health and average number of physically unhealthy days receive a weight of 10% under the CHR's while we find that they explain the most variation in our latent factor at 28% and 27%, respectively. The rate of babies born low birth weight explains the least amount of variation in county health according to our estimation but receives the second highest weight in the CHR's at 20%. We find the most agreement between our squared correlation coefficients and the CHR's weights for number of mentally unhealthy days, but the two measures still differ by roughly 80%. The CHR weights of all five manifest variables do not fall within the posterior 95% credible intervals of the squared correlation coefficients estimated using our model.

Table 1: Posterior Means and 95% Confidence Intervals of Model Factor Loadings, Residual Variances, and Normalized Squared Correlation Coefficients

	Factor Loadings (95% CI)	Residual Standard Deviations (95% CI)	Normalized Squared Correlation Coefficients (95% CI)
YPLL Rate	19.25 (18.45, 20.01)	22.20 (21.53, 22.92)	0.18 (0.17,0.19)
Self-Reported Fair/Poor Health	57.30 (55.23, 59.32)	39.30 (37.56, 40.99)	0.28 (0.27,0.29)
Physically Unhealthy Days	9.74 (9.37, 10.10)	7.52 (7.23, 7.81)	0.27 (0.26,0.29)
Mentally Unhealthy Days	7.43 (7.06, 7.79)	9.48 (9.19, 9.78)	0.18 (0.15,0.20)
Low Birth Weight Rate	10.90 (10.16, 11.67)	23.53 (22.91, 24.17)	0.08 (0.07,0.09)

Comparing County Health Rankings Within and Across States:

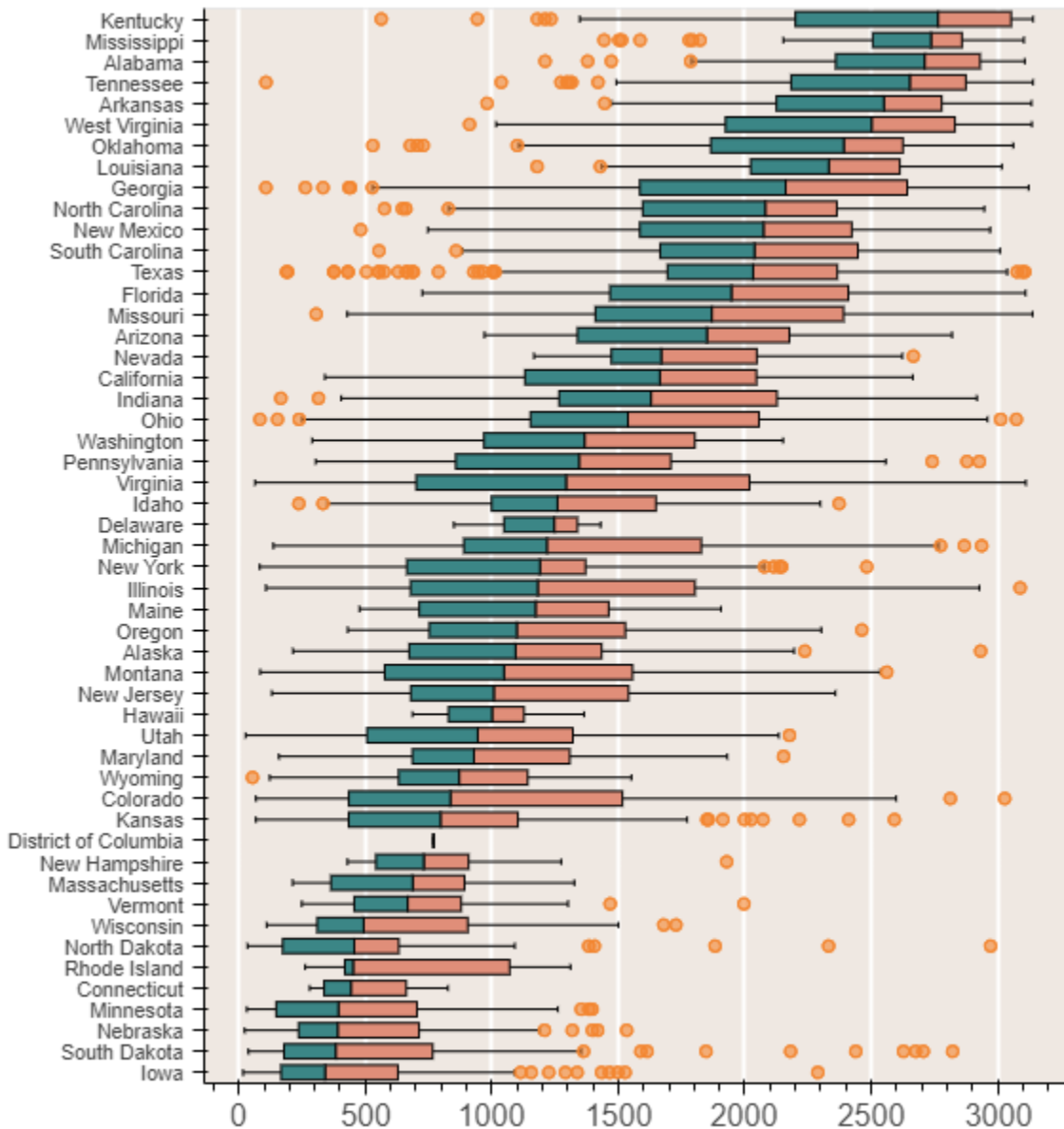
While the official CHRs separately rank county health by state, there are often cases that require comparisons across state lines. Figure 1 shows box plots of the mean posterior health rankings produced by our model separated by state, with lower rankings corresponding to better population health levels and higher rankings representing worse population health. The figure illustrates a considerable degree of disparity in the rankings of county health both across and within states. For example, the healthiest county in Mississippi, DeSoto, has a mean posterior rank roughly 1,000 ranks away from the state's median, making it an extreme within-state

outlier. However, while DeSoto is the best performing county within Mississippi it falls below the median county rank of 31 healthier states. Therefore, even though DeSoto is found to be the healthiest county in Mississippi, it performs worse than average in most states and would even be an unhealthy outlier in some of the nation's healthiest states. A system that only ranks counties within states would make this comparison impossible as DeSoto would be indistinguishable from the top ranked counties of the other 49 states.

We can make a similar comparison for the least healthy counties located in otherwise healthy states. Counties in Iowa are the nation's healthiest at the median, but the state also has a few unhealthy outliers. The least healthy county in Iowa, Appanoose, has a mean posterior rank of 2,289, placing it nearly 2,000 ranks away from the state's median rank. While Appanoose is considered an extremely unhealthy county within Iowa, it still outperforms the median county in eight of the nation's least healthy states. If we were to simply rank the counties in Iowa, Appanoose would remain the worst ranked county but the *degree* to which the county underperforms relative to its peers would be masked.

The within-state county health disparities shown in Figure 1 are also unobservable if counties are ranked separately on a state-by-state basis. For example, Michigan, New York, and Illinois have counties with mean posterior ranks at nearly every point of the national distribution, implying that the three states include a mix of the country's most and least healthy counties. Alternatively, while the median county health rank in Delaware is worse than Michigan, New York, and Illinois, its distribution of health is much tighter. This indicates lower levels of within-state disparity in Delaware. Again, these comparisons cannot be made when rankings are calculated separately for each state.

Figure 1: Box Plots of Posterior Mean Rankings of County Health by State



Explaining County Health Ranks:

We now turn to the factors associated with a county's health ranking. To investigate which variables significantly predict ranks, we first estimate regressions of our model's mean posterior ranks on a set of county-level socioeconomic outcomes including total population, the

percentage of residents with less than a high school education, average income, income Gini, percentage of black residents, and the percentage of Hispanic residents.³ We then include measures of a county's adult economic mobility among children born to low-income families. For ease of interpretation, all independent variables are normalized as z-scores.

Table 2 shows results from three separate regressions of our mean health rankings on the set of county outcome z-scores. Beginning with Column 1 of Table 2, we see that a county's percentage of residents without a high school education, average income, income Gini, and percentage of Hispanic residents all have statistically significant effects on mean posterior health ranks while total population and percentage of black residents do not. Having a higher proportion of residents with less than a high school education and greater levels of income inequality predict worse health rankings and counties with higher average incomes are expected to be better ranked than lower-income counties. The percentage of Hispanic residents is also positively correlated with health, implying that there may be a beneficial health return to specific features of a county's racial and ethnic diversity.

Column 2 of Table 2 shows results from a regression of our posterior mean ranks on each of the variables in Column 1 along with the addition of a variable capturing the effect of living in a county during childhood on adult income mobility among children born to low-income parents. This variable comes from the work of Chetty and Hendren (2018) which we refer to from here on as CH. The CH variable shows an estimated percent change in future household income at age 26 relative to the national average caused by spending one additional year of childhood in the county for a child born to parents with incomes at the 25th percentile of the national

³ On average, black, Hispanic, and white residents make up roughly 95% of a county's population. We therefore exclude percent white as a reference group in our regressions in order to avoid high degrees of multicollinearity.

distribution. Positive (negative) values of the CH variable correspond to a positive (negative) effect from spending a year of childhood in the county on relative levels of adult income for children born to lower-income families. While the CH variable likely captures many things, we refer to it as county economic mobility throughout the remainder of this study. Like other socioeconomic variables used in our regressions, we normalize our CH variable as a within-sample z-score.

With the addition of county economic mobility, we find the estimated effects of a county's percentage of residents without a high school education, average income, income Gini, and percentage of Hispanic residents remain relatively unchanged compared to our results in Column 1. Alternatively, total population and the percentage of black residents become statistically significant at the 1% and 0.1% levels, respectively. Both effects are negative, implying that counties with larger populations and higher percentages of black residents are predicted to have better health all else equal after controlling for economic mobility. We also find that higher levels of economic mobility among children born to low-income parents has an economically and statistically significant effect on a county's mean posterior health ranking with greater income mobility predicting better health. This finding suggests that providing disadvantaged children with greater opportunity for economic advancement may prove beneficial to a county's health level.

Finally, to test if economic mobility has a differential effect on county health rankings for counties with different racial and ethnic makeups, we include interaction terms of the CH economic mobility variable with the percentage of black residents and the percentage of Hispanic residents in Column 3 of Table 2. After including our interaction effects, we find that the effects of total population, percentage of residents without a high school education, average

income, and income Gini change only trivially. Alternatively, we find that the percentage of black residents becomes statistically insignificant while its interaction effect is significant at the 0.1% level. Our results suggest that counties with higher percentages of black residents see lower returns to increases in economic mobility while the percentage of black residents alone does not have a statistically significant effect on a county's health ranking keeping economic mobility fixed. Alternatively, the coefficient on percentage of Hispanic residents remains statistically significant at the 0.1% level while the interaction of percent Hispanic and economic mobility is much smaller in magnitude and only marginally statistically significant. Like the percentage of black residents, our findings suggest that counties with higher percentages of Hispanics see smaller returns to economic opportunity. Unlike percent black, we find that increases in the percentage of Hispanic residents living in a county has a beneficial effect on county health that is independent of its relationship to economic mobility.

We now estimate regressions of our five manifest variables on the same set of county-level socioeconomic variables. These estimates allow us to examine the degree to which each manifest variable covaries with our socioeconomic outcomes. This approach provides information regarding the relative contribution of each manifest variable to the overall change in our mean ranks.

Table 3 shows regressions of mean rank and our manifest variables on the set of county-level outcomes. All dependent and independent variables are normalized as z-scores in order to provide coefficients which are comparable across regressions. Looking at the results of Table 3, we see that the overall effect of total population on our mean ranks is largely driven by changes in years of potential life lost rate and rate of low birth weight. The relationship between total population and YPLL rate is in line with recent work showing that urban areas often have better

longevity at age 65 due to greater health care quality, availability, and utilization (Finkelstein et al. 2019). These differences in the health care environment between urban and rural areas may also explain the negative relationship between low birth weight and total population. The percent of individuals living in a county with less than a high school education corresponds to significantly worsened outcomes for all five manifest variables with the largest effects for the rate of fair/poor health and average number of physically unhealthy days. The same is true for county-level income Gini, but the effect is largest for YPLL rate. Alternatively, average county income level corresponds to significant improvements in each of the five manifest variables with the largest change being YPLL rate.

Looking to our measures of race and race interacted with economic mobility, we find several interesting relationships. Starting with the percentage of black residents living in a county, we find that increases in percent black lead to various effects on PUD and MUD depending on a county's level of economic mobility. For a county with an economic mobility score of 0, implying that spending an additional year in the county has neither a positive nor negative effect on adult earnings for children born to low income parents, increases in percent black lead to expected improvements in number of positive physical and mental health days. Interestingly, the beneficial effect of percent black on PUD and MUD is larger in counties with lower levels of economic mobility and smaller in counties with higher economic mobility. For the percentage of babies born low birth weight, increases in percent black lead to substantial increases in LBW%. Furthermore, the effect of percent black on birth weight is independent of economic mobility. This finding is supported by the literature on differential probabilities of low birth weight by race which finds that black mothers are more likely to give birth to low birth weight babies due to

factors such as maternal age at time of birth and ingrained institutions related to systemic disadvantage among African Americans (CDC 2002, Love et al. 2010, Ncube et al. 2016).

Finally, we evaluate the effects of county Hispanic percent and Hispanic percent interacted with economic mobility on our manifest variables. Like percent black, increases in percent Hispanic lead to decreases in YPLL rate, rate of fair/poor health, PUD, and MUD. Unlike percent black, however, these effects are largely independent of economic mobility with the interaction only being marginally statistically significant for the regression of YPLL rate. Furthermore, the effect of percent Hispanic on LBW% is negative while the coefficient on percent Hispanic interacted with economic mobility is positive. This relationship implies that increases in a county's share of Hispanic residents produces higher levels of LBW% among counties with higher economic mobility and lower LBW% in counties with worse economic mobility.

Table 2: Regression of Mean Ranks on Socio-Economic Outcome Z-Scores

	(1)	(2)	(3)
	Mean Posterior Rank	Mean Posterior Rank	Mean Posterior Rank
Total Population	7.160 (6.936)	-25.69** (8.006)	-20.54* (9.028)
No Highschool%	373.5*** (24.30)	361.7*** (21.73)	333.7*** (22.53)
Mean Income	-345.1*** (17.98)	-313.6*** (16.78)	-332.2*** (17.85)
Gini	176.7*** (13.26)	123.9*** (13.25)	130.2*** (13.34)
Black%	-12.76 (11.62)	-119.3*** (12.68)	-35.74 (19.96)
Hispanic%	-167.3*** (15.52)	-151.7*** (14.89)	-167.8*** (14.67)
CH		-247.6*** (15.79)	-290.5*** (18.47)
CH * Black%			115.8*** (18.69)
CH * Hispanic%			47.66* (18.87)
<i>N</i>	2832	2798	2798
adj. <i>R</i> ²	0.573	0.616	0.621

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regressions of Mean Rank and CHR Manifest Variable Z-Scores on Socio-Economic Outcome Z-Scores with CH Interactions

	(1) Mean Posterior Rank Z-Score	(2) Years of Potential Life Lost Rate Z- Score	(3) Rate of Fair/Poor Health Z-Score	(4) Avg Number of Physically Unhealthy Days Z-Score	(5) Avg Number of Mentally Unhealthy Days Z-Score	(6) Rate of Low Birth Weight Z- Score
Total Population	-0.0233* (0.0102)	-0.0500*** (0.0123)	-0.0209 (0.0135)	0.00137 (0.00842)	0.0245 (0.0127)	-0.0332** (0.0104)
No Highschool%	0.378*** (0.0255)	0.270*** (0.0293)	0.493*** (0.0298)	0.427*** (0.0311)	0.312*** (0.0303)	0.151*** (0.0257)
Mean Income	-0.376*** (0.0202)	-0.380*** (0.0213)	-0.295*** (0.0224)	-0.266*** (0.0215)	-0.205*** (0.0218)	-0.0966*** (0.0177)
Gini	0.147*** (0.0151)	0.214*** (0.0186)	0.133*** (0.0171)	0.105*** (0.0189)	0.0656*** (0.0186)	0.121*** (0.0157)
Black%	-0.0405 (0.0226)	-0.0561* (0.0281)	-0.0408 (0.0298)	-0.248*** (0.0304)	-0.221*** (0.0374)	0.490*** (0.0256)
Hispanic%	-0.190*** (0.0166)	-0.306*** (0.0219)	-0.136*** (0.0197)	-0.259*** (0.0203)	-0.233*** (0.0196)	-0.0763*** (0.0216)
CH	-0.329*** (0.0209)	-0.168*** (0.0285)	-0.142*** (0.0224)	-0.283*** (0.0253)	-0.386*** (0.0290)	-0.178*** (0.0210)
CH*Black%	0.131*** (0.0212)	-0.00951 (0.0262)	0.0860** (0.0282)	0.103*** (0.0289)	0.117*** (0.0345)	0.00538 (0.0257)
CH*Hispanic%	0.0540* (0.0214)	0.0595* (0.0235)	0.00649 (0.0210)	0.0240 (0.0200)	0.0400 (0.0213)	0.0930*** (0.0235)
<i>N</i>	2798	2798	2798	2798	2798	2798
adj. <i>R</i> ²	0.621	0.537	0.519	0.401	0.295	0.612

Standard errors in parentheses

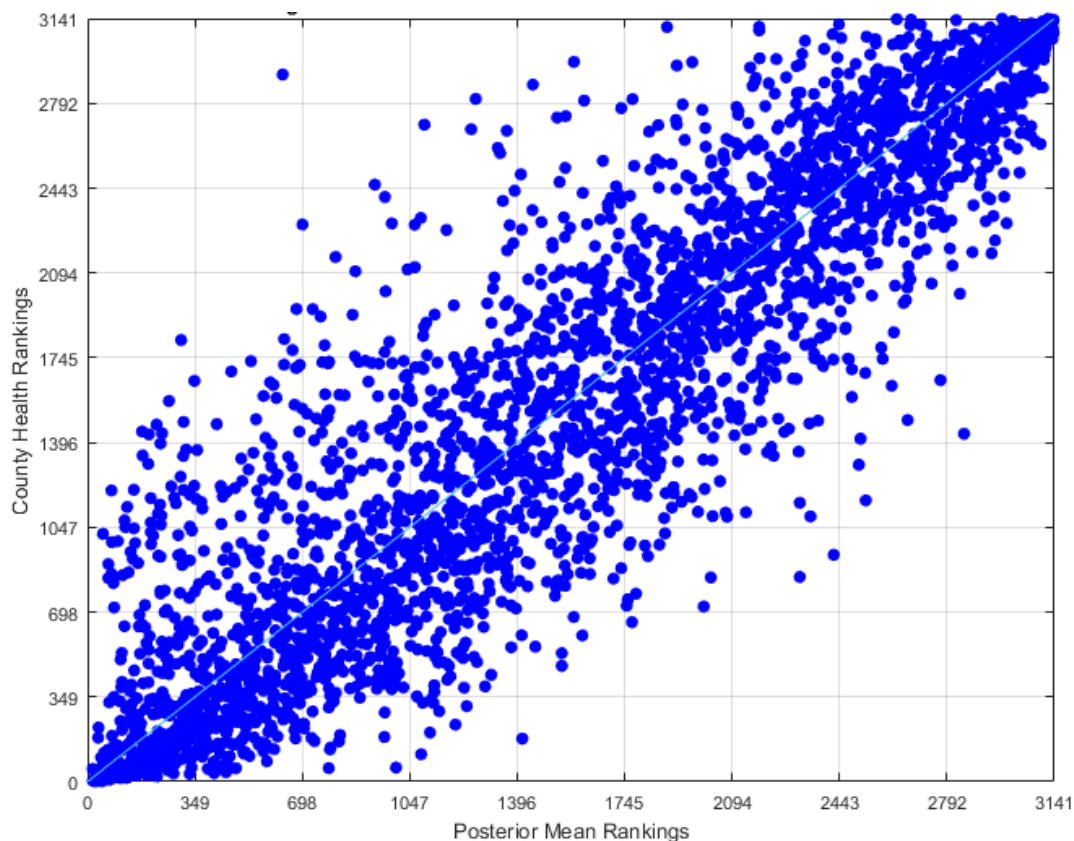
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comparing CHRs to Posterior Mean Ranks:

We now compare the similarity of our model's mean posterior health rankings to a set of joint rankings for all U.S. counties created using the official CHR methodology. Figure 2 shows the posterior mean ranks produced by our model on the x-axis along with the corresponding CHR of each county on the y-axis. While the two rankings generally agree with one another, the posterior mean rankings of our model most closely match those of the CHRs for counties at the top and bottom of the rank distribution. Alternatively, there is less agreement between both sets of rankings for counties towards the middle of the distribution. This comparison implies that

particularly healthy and unhealthy counties are relatively easy to identify and rank compared to counties towards the middle of the national distribution.

Figure 2: CHR Ranks vs. Posterior Mean Ranks



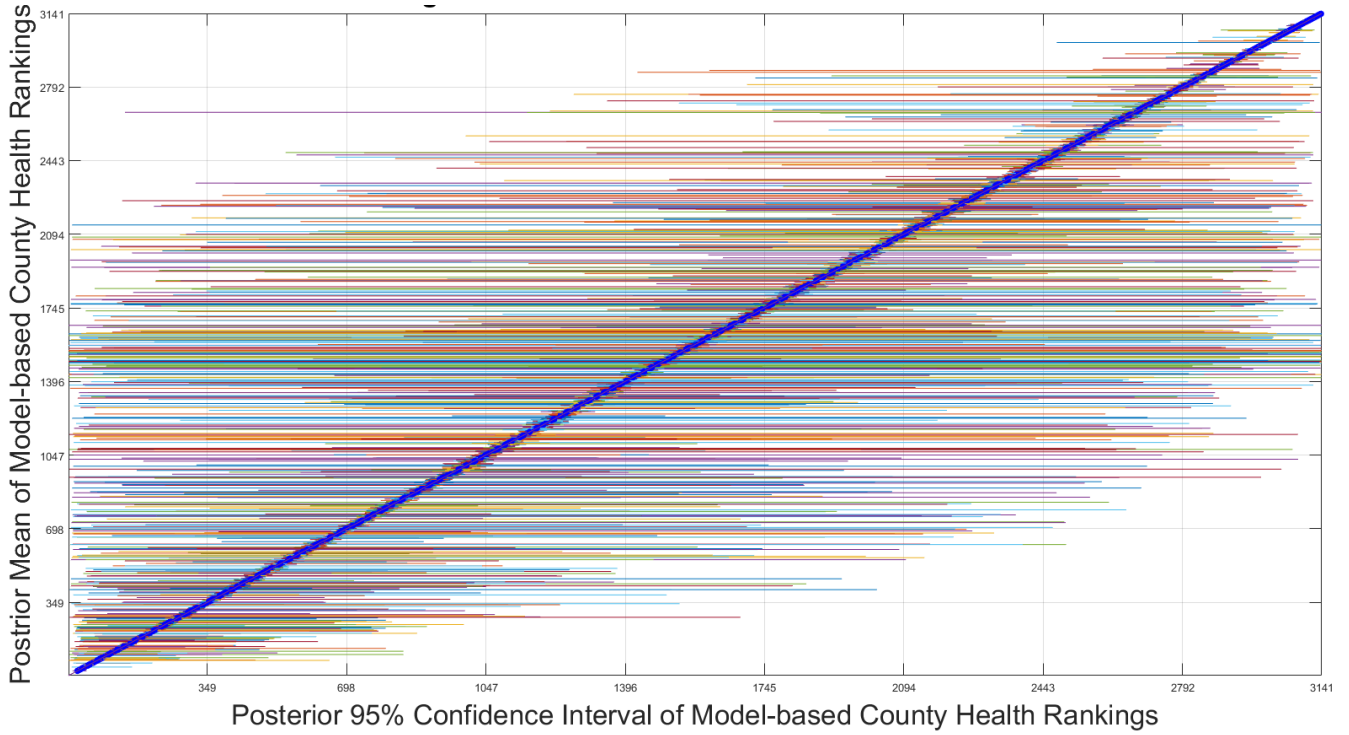
Mean Rankings and Uncertainty:

Since our Bayesian methodology produces draws from a posterior distribution of health rankings rather than a single point estimate, we can utilize both the mean and 95% confidence interval of each county's rank distribution when presenting our results. Figure 3 shows the ordered posterior mean ranks of each county along the 45-degree line with their corresponding 95% confidence

intervals represented by the solid horizontal lines. Graphically, we see a significant amount of uncertainty in the county health rankings produced by our model with the 95% rank confidence intervals of some counties reaching across the entire distribution of rankings. We also find less uncertainty among counties with posterior mean ranks towards the distribution's left and right tails, further supporting that we have greater certainty about the relative health rankings of very healthy and unhealthy counties. There is, however, a significant amount of uncertainty in health rankings for counties at all points of the distribution.

The degree of uncertainty found in our rankings suggest that single point estimates provide very little information regarding the relative health ranking of many counties. For policymakers concerned with the health of poorly performing counties, lower levels of uncertainty allow for greater accuracy when selecting at-risk counties for policy interventions or evaluation. We discuss these counties in more detail below. Additionally, with less uncertainty policymakers and researchers can more easily identify the nation's healthiest counties in order to study what conditions produce better health. For counties with rankings towards the distribution's middle, however, considerable uncertainty implies that we can only tentatively determine their health relative to similar counties.

Figure 3: Mean Posterior Ranks and 95% Confidence Intervals



Identification of the Least Healthy Counties:

While uncertainty in our model's rankings can make it difficult to measure relative health, we have shown that this is not the case for all counties. Some counties have smaller mean posterior rank 95% confidence interval (95% CI) ranges which provides greater assurance that our mean estimates accurately reflect underlying relative health. Additionally, certain counties see more agreement between the mean posterior health rank produced by our model and the CHR's. If both methods capture aspects of underlying health, we may be more certain regarding the true health of a county when the two approaches produce similar rankings.

To highlight cases where relative health is easier to identify, we examine the set of counties which are most likely to be ranked in the bottom 10%. Counties in this bottom 10% group must

satisfy two to three criteria. First, the county's mean posterior ranking must fall within the bottom 10% of rankings according to our model. Second, the county must also be in the bottom 10% according to the CHRs. These two criteria allow us to include counties which are found to be among the least healthy regardless of which model is used. Finally, among the set of counties with mean posterior ranks and CHRs in the bottom 10%, we evaluate the subset of counties with 95% CIs that do not cross the 90th percentile threshold. This further restriction provides additional assurance that the counties in question truly belong in the bottom 10%.

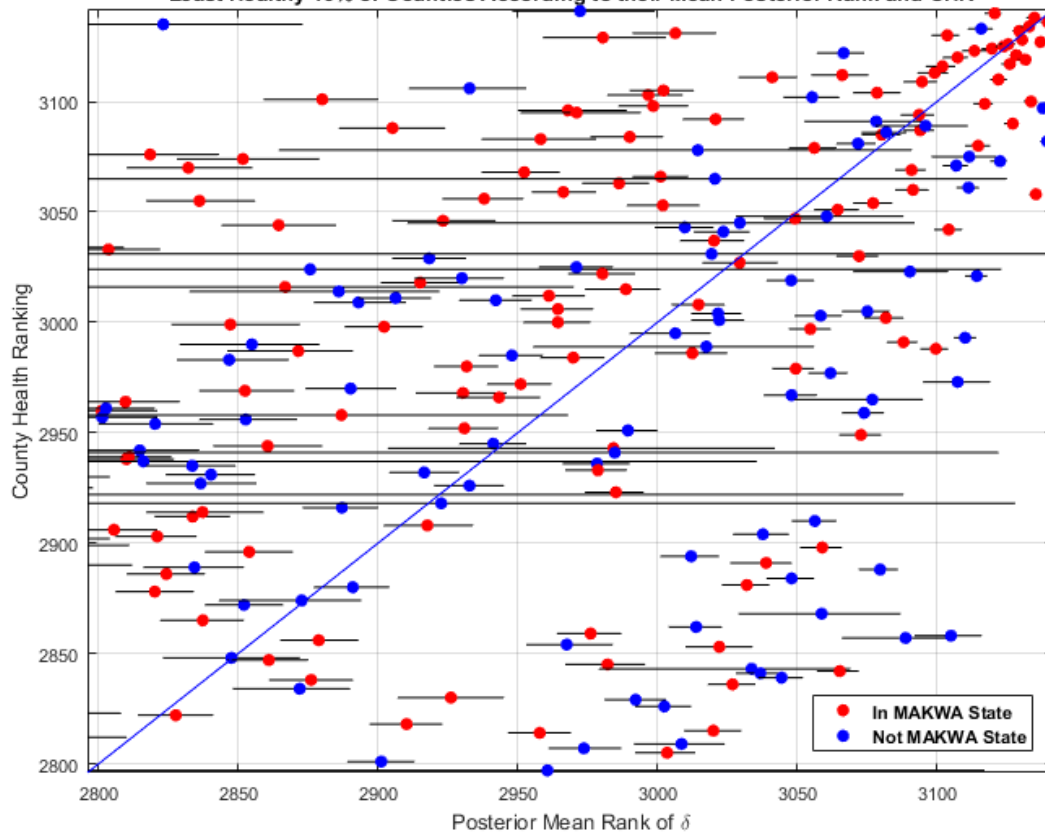
Figure 4 shows the mean posterior rank, 95% CI, and CHR for counties in our bottom 10% group according to their mean posterior rank and CHR. We show the same measures for the subset of counties whose 95% CI also falls within the bottom 10% range in Figure 5. We find 209 counties that are in the bottom 10% according to their mean posterior rank and CHR. Among these 209 counties, 189 have 95% mean posterior rank confidence intervals that fall within the bottom 10% range. Unsurprisingly, some states have no counties in our bottom 10% group while other states have several. Of the 209 counties shown in Figure 4, 121 are in one of five states: Mississippi, Alabama, Kentucky, West Virginia, and Arkansas (MAKWA). The mean posterior rank of counties in MAKWA and non-MAKWA states are identified in our two figures as red and blue, respectively.

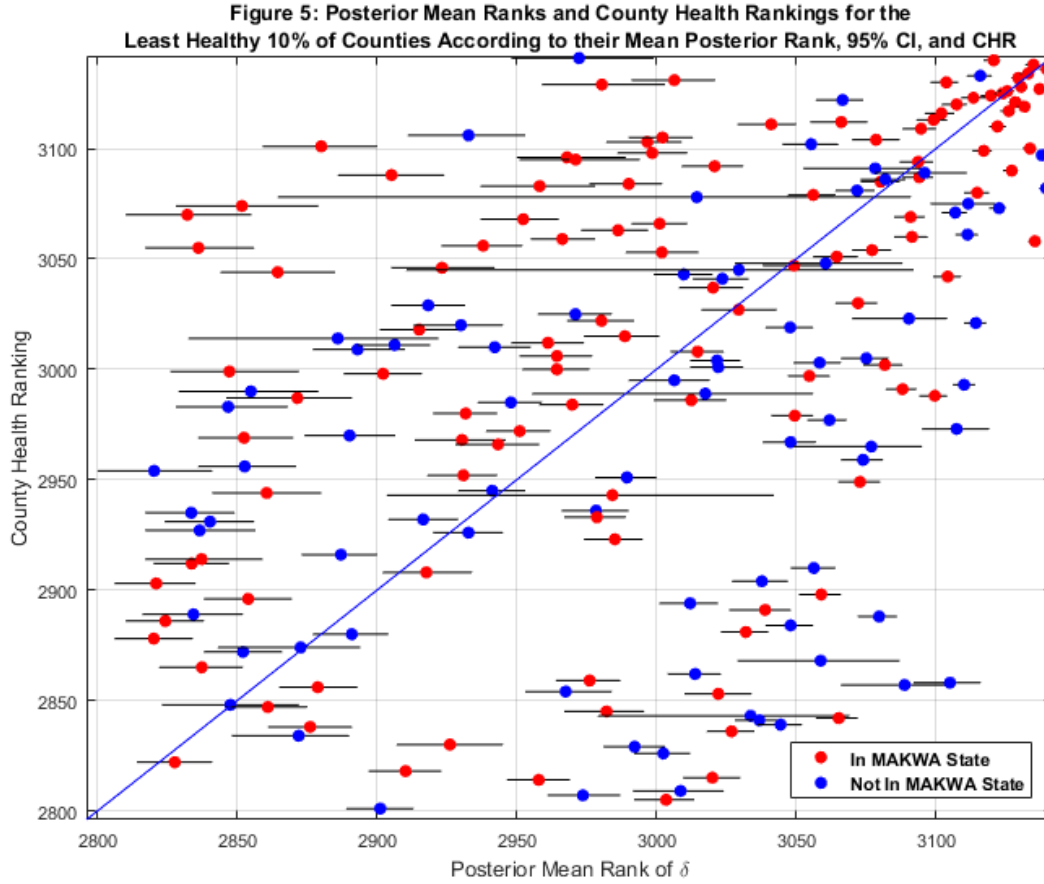
Figures 4 and 5 provide a few key takeaways. First, while our approach differs substantially from the CHRs in certain respects, these differences do not necessarily carry over into subsequent rankings. By ranking 3,141 total counties, 314 will necessarily be in the bottom 10% under each model. We find 209 counties that are in the bottom 10% according to both, meaning that the two methods generally agree for nearly two-thirds of all potential counties. Furthermore, all but 20 of these counties have 95% rank confidence intervals that fall entirely within the

bottom 10% range. While Figure 3 shows that many counties have high levels of uncertainty regarding their mean posterior rank, relative health among the nation's least healthy counties is far more certain.

Finally, we show that a small number of states contain most counties in our bottom 10%. Like our state-level examination of mean posterior rank distributions, we find that counties in the south are overrepresented among the bottom 10%. In addition to making up more than half of all bottom 10% counties, most of the bottom 10%'s least healthy counties are also in one of five MAKWA states. Only 3 of the 20 worst performing counties are in a non-MAKWA state and all 20 are located in the south. The share of MAKWA counties in the bottom 10% according to our model and the CHRs further solidifies the poor health of southern counties relative to the national average.

Figure 4: Posterior Mean Ranks and County Health Rankings for the Least Healthy 10% of Counties According to their Mean Posterior Rank and CHR





Uncertainty Under Different Imputation Methods:

As discussed previously, there are a considerable number of counties which are missing observations for one or more of the five manifest variables. While our primary results utilize posterior imputation to predict and replace missing observations, we can estimate our rankings under various imputation procedures. A significant difference between posterior imputation and alternative methods is that posterior imputation incorporates the uncertainty from missing data directly into the posterior distribution of each county's health rankings while other imputation procedures do not. We therefore expect a significant increase in the degree of uncertainty for counties with missing data using posterior imputation versus traditional imputation approaches.

Figure 6 shows density plots for the distribution of 95% rank confidence interval magnitudes produced by our model under different approaches to dealing with missing data. Figure 6A shows the density of posterior rank confidence intervals estimated using posterior imputation. We see from the figure that the density is heavily right-skewed with the confidence interval of some counties stretching across hundreds of ranks. Figure 6B shows the density of 95% rank confidence interval magnitudes produced by our model using Predictive Mean Matching (PMM) to impute missing data rather than posterior imputation.⁴ As opposed to similar methods like naïve imputation using linear regression, the PMM algorithm preserves many features of the observed data since all missing values are imputed using the non-missing observation of its nearest neighbor.

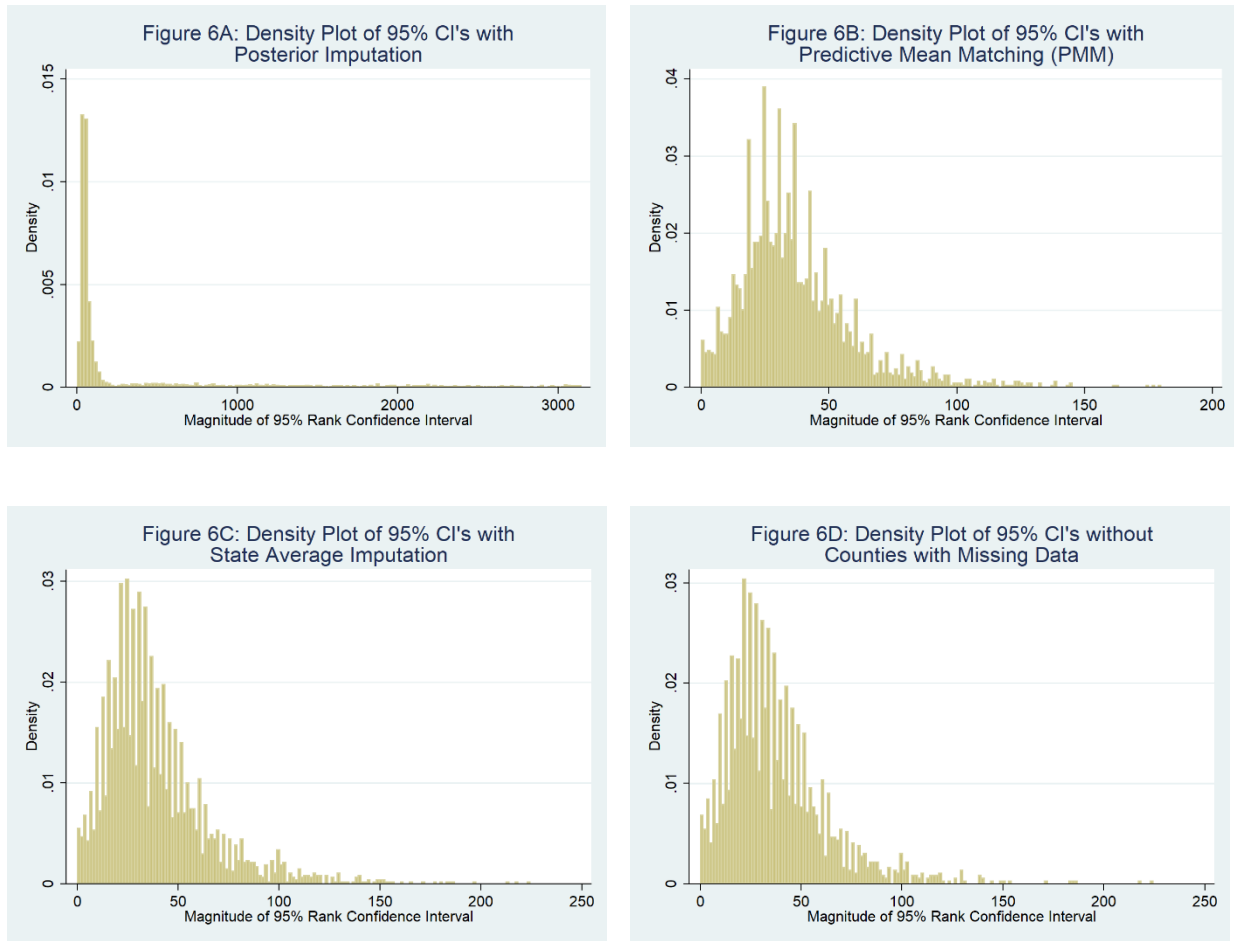
Compared to Figure 6A, the 95% rank confidence intervals produced using PMM vary in magnitude significantly less than the ranks produced under posterior imputation. Since the uncertainty of missing data is not directly incorporated into our posterior rank distributions using PMM, counties with missing data have tighter confidence intervals on average relative to our primary results. Comparing the change in 95% rank confidence intervals more formally, the average percent change in interval magnitude from PMM to posterior imputation is roughly 897% with a median of 81%. Compared to PMM, nearly 45% of counties see a percent increase of 100% or more under posterior imputation and roughly 18% of counties see an increase greater than 1,000%.

⁴ PMM imputes data in two stages. The first stage involves predicting values of each manifest variable for all counties in the data set using linear regression. In the second stage, the algorithm matches the predicted manifest variable value of each county with the most similar predicted value among the set of counties without missing data. This county is known as the missing county's "nearest neighbor." The missing value of the county is then imputed using its nearest neighbor's observed manifest variable value.

Alternatively, we can use state averages to impute missing data. We show the density of 95% rank confidence interval magnitudes produced using state-average imputation to replace missing data in Figure 6C. Like our PMM estimates, we find that the rank confidence interval distributions are much tighter on average and less heavily skewed using state-average imputation compared to our posterior imputation results. Around 47% of counties see a 100% or greater increase in their 95% rank confidence interval magnitude when moving from state-average imputation to posterior imputation and 18% of counties see a change of 1,000% or more.

While the smaller rank confidence interval magnitudes produced using alternative imputation strategies may seem attractive, they do not holistically incorporate the uncertainty of missing data in the same way as posterior imputation. Since we would expect to have less confidence in the rankings of counties with significant amounts of missing data *a priori*, ignoring this uncertainty may be undesirable. We can examine the relationship between uncertainty from missing data and our uncertainty in health ranks in another way by looking at Figure 6D which shows the density of 95% rank confidence interval magnitudes produced by our model using only the set of counties that have no missing data. As Figure 6D shows, the density of confidence interval magnitudes produced using only counties with full information is very similar to the densities produced using PMM and state-average imputation. This similarity suggests that using traditional imputation procedures which ignore the uncertainty from missing data produces deflated levels of uncertainty akin to only using data without any missing observations, leading to artificially high levels of certainty regarding the relative health ranking of counties.

Figure 6: Density Plots of 95% Confidence Interval Magnitudes Using Different Imputation Procedures



Using Mean Posterior Draws of Latent County Health vs. Posterior Mean Ranks:

As an alternative to rankings of county health, one could use the estimates of latent county health produced by our model directly. While rankings are often more convenient to interpret, they have an inherent disadvantage compared to directly using estimates of county health. Specifically, the amount of variation in each county's posterior rank distribution is driven by two factors: sampling variation in their own latent factor's posterior distribution and sampling variation in the overlapping latent factor distributions of counties with similar levels of health. To see this, assume we have a sample containing only two counties, county A and county B. Let county A

have a mean posterior latent health draw of 1 in iteration t while county B has a mean posterior health draw of 1.5, implying that county A has a rank of 1 and B has a rank of 2. Now, at iteration $t + 1$ assume that county A draws a mean posterior latent health of 1.6 and county B draws another sample at exactly 1.5. While the mean posterior latent health level of county B has not changed across iterations, the variation in county A's latent factor will cause the ranks of both counties to reverse. Since sampling variation in the latent factor of counties with overlapping distributions of health can produce movement in a county's rank independently of variation in their own mean posterior health distribution, the level of uncertainty in relative county health is almost certainly higher when using rankings rather than the estimates of latent health directly.

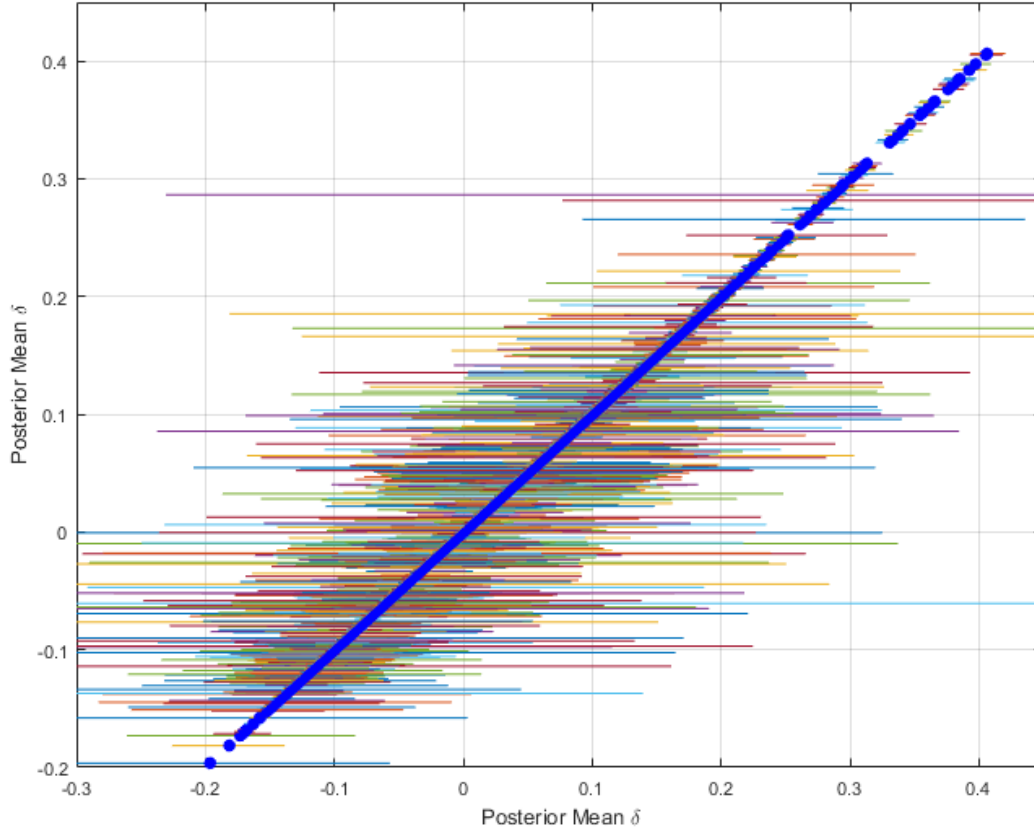
To investigate the differences in uncertainty between these two approaches more closely, Figure 7 shows ordered mean posterior latent health values on the 45-degree line along with 95% posterior mean confidence intervals shown as horizontal lines. Graphically, we see that using posterior means of latent county health directly produces significantly lower levels of overall uncertainty than we find using ranks in Figure 3.

Examining the degree to which changes in the uncertainty of our results under different imputation strategies is driven by the choice to use ranks, we consider percent changes in the posterior 95% confidence interval magnitudes for both posterior distributions of health rankings and posterior distributions of mean latent health caused by switching from PMM to posterior imputation and state average imputation to posterior imputation. While posterior imputation still produces significantly higher levels of uncertainty relative to the alternative imputation strategies, the average percent change in uncertainty is substantially smaller when using mean posterior latent health directly instead of rankings. More formally, the average increase in 95%

confidence interval magnitude caused by switching from PMM to posterior imputation is 319% when using posterior means of latent county health compared to an average increase of 998% when using ranks. Similarly, switching from state average imputation to posterior imputation leads to an average increase of 661% using posterior means of latent county health compared to an average increase of 1,031% when using ranks.

Given that many deterministic measures like the CHRs only provide rankings, the inflation in uncertainty inherent to ranking a continuous underlying factor is especially important when interpreting estimates. Our methodology gives practitioners the choice to directly utilize estimates of the latent factor for each county rather than having to rely on rankings. While directly using the latent factor produces a measure with less uncertainty, it comes at a cost since point estimates of county health have no direct scale or interpretation aside from their direction and magnitude relative to other counties in the sample.

Figure 7: Model-driven Posterior Means of δ with 95% Confidence Intervals



6. Conclusion:

In this paper, we produce rankings of health for all counties in the United States using a Bayesian spatial factor analysis model. We utilize the five observable variables used to construct the County Health Rankings (CHRs) produced by the University of Wisconsin Population Health Institute. Our estimation approach has several advantages over the CHR's methodology. First, our model estimates factor loadings for each manifest variable using information present in the data rather than deterministic weightings that are subject to expert opinion. Second, we produce a measure of uncertainty in our rankings which is directly influenced by a county's population size

and amount of missing data. Finally, we rank all counties in the nation simultaneously rather than separately by state, allowing us to compare the health of counties both within and across state lines.

Compared to the CHRs, we find that the normalized squared correlation coefficients produced by our model attribute significantly less weight to Years of Potential Life Lost (YPLL) and the rate of Low Birth Weight (LBW) while Self-Reported Fair/Poor Health (FP%) rate, average number of Physically Unhealthy Days (PUD), and average number of Mentally Unhealthy Days (MUD) receive significantly higher weightings. These differences suggest that the CHRs may attribute too much weight to outcomes of early mortality and birth weight and too little weight to adult physical and mental health status.

Our study also contributes to the literature on health disparity measurement (Bleichrodt and van Doorslaer 2006, Nesson and Robinson 2017). Comparing the health of counties across state lines, we find a significant amount of disparity in the rankings of counties relative to different states as well as among counties within the same state. While county health in some states has a relatively tight distribution, most have a mix of healthy and unhealthy counties. Additionally, the healthiest counties in some of the unhealthiest states often fall below the average of healthier states. Since the CHRs only rank county health separately for each state, it cannot be used to make these types of comparisons.

We investigate which county-level variables predict the health rankings produced by our model. In addition to total population, the percentage of residents with less than a high school education, average income, income Gini, percentage of black residents, and percentage of Hispanic residents, we include a county economic mobility variable created by Chetty and

Hendren (2018) (CH) which capture the effect of spending an additional year of childhood living in a county on adult income for children born to low-income parents. While the effect of increasing a county's percentage of black residents is found to be statistically insignificant in a regression excluding economic opportunity, the effect of having higher percentages of both black and Hispanic residents becomes beneficial to health after accounting for adult economic mobility. Furthermore, we find that the effect of racial diversity on health varies with a county's level of economic mobility. Increased economic mobility predicts better health rankings, but the effect is smaller for counties with higher percentages of black residents and Hispanic residents. Furthermore, the effect of percent black on health only operates through its interaction with economic mobility while increasing the percentage of Hispanic residents has a beneficial effect that is independent of economic mobility.

We also find that our rankings harbor a considerable degree of uncertainty. While the relative rankings of counties towards the middle of the distribution harbor the most uncertainty, there is significant ambiguity regarding the relative health of counties at every points of the distribution. This is not the case for all counties, however, which we illustrate by examining the set of counties in the bottom 10% of health according to their mean posterior ranking, 95% mean posterior ranking confidence interval, and CHR.

One source of uncertainty in our rankings comes from incorporating the inherent uncertainty of missing data directly into our model using posterior imputation. We show that using alternative imputation procedures like Predictive Mean Matching (PMM) and imputation using state averages produces far tighter rank confidence intervals. While this may seem like an attractive feature, these alternative imputation methods abstract from the uncertainty of estimating ranks for counties that are missing data for one or more manifest variables.

Furthermore, we highlight how the use of ranks rather than direct measures of latent health inflates the uncertainty of our estimates. This finding further illustrates the importance of recognizing uncertainty in the rankings of deterministic frameworks like the CHR.

The primary limitation of our study is data quality. More specifically, roughly 23% of all counties in the data set are missing observations for one or more manifest variable. While one of our study's primary aims is to illustrate a serious treatment of these missing data, the use of more complete data would result in better precision.

In addition to the use of variables with less missing data, future research would benefit from the use of longitudinal data on observable county health outcomes rather than cross-sectional data taken from a single year. While the manifest variables used in this paper are available in earlier periods, the three variables collected from the Behavioral Risk Factor Surveillance System (BRFSS) become unreliable for smaller counties. With longitudinal data covering an adequate amount of time, researchers can not only measure the health of counties at a given point but also how the health of counties evolves across time. Being able to measure this time-variant aspect of county health is crucial to understanding the potential policies and factors which are effective at improving or worsening a county's health in the long term.

References:

1. Bleichrodt, Han, and Eddy Van Doorslaer. "A welfare economics foundation for health inequality measurement." *Journal of Health Economics* 25, no. 5 (2006): 945-957.
2. Centers for Disease Control and Prevention (CDC). "Infant mortality and low birth weight among black and white infants--United States, 1980-2000." *MMWR. Morbidity and mortality weekly report* 51, no. 27 (2002): 589.
3. Chetty, Raj, and Nathaniel Hendren. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics* 133, no. 3 (2018): 1107-1162.
4. Chib, Siddhartha, and Edward Greenberg. "Understanding the metropolis-hastings algorithm." *The american statistician* 49, no. 4 (1995): 327-335.
5. Chib, Siddhartha, and Edward Greenberg. "Markov chain Monte Carlo simulation methods in econometrics." *Econometric theory* 12, no. 3 (1996): 409-431.
6. Courtemanche, Charles, Samir Soneji, and Rusty Tchernis. "Modeling Area-Level Health Rankings." *Health services research* 50, no. 5 (2015): 1413-1431.
7. Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Place-Based Drivers of Mortality: Evidence from Migration." *NBER Working Paper*, no. 25975 (2019).
8. Gelfand, Alan E., Susan E. Hills, Amy Racine-Poon, and Adrian FM Smith. "Illustration of Bayesian inference in normal data models using Gibbs sampling." *Journal of the American Statistical Association* 85, no. 412 (1990): 972-985.
9. Giffinger, Rudolf, and Haindlmaier Gudrun. "Smart cities ranking: an effective instrument for the positioning of the cities?." *ACE: architecture, city and environment* 4, no. 12 (2010): 7-26.
10. Hogan, Joseph W., and Rusty Tchernis. "Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data." *Journal of the American Statistical Association* 99, no. 466 (2004): 314-324.
11. Love, Catherine, Richard J. David, Kristin M. Rankin, and James W. Collins Jr. "Exploring weathering: effects of lifelong economic environment and maternal age on low birth weight, small for gestational age, and preterm birth in African-American and white women." *American journal of epidemiology* 172, no. 2 (2010): 127-134.
12. Marginson, Simon, and Marijk Van der Wende. "To rank or to be ranked: The impact of global rankings in higher education." *Journal of studies in international education* 11, no. 3-4 (2007): 306-329.
13. Ncube, Collette N., Daniel A. Enquobahrie, Steven M. Albert, Amy L. Herrick, and Jessica G. Burke. "Association of neighborhood context with offspring risk of preterm birth and low birthweight: A

systematic review and meta-analysis of population-based studies." *Social science & medicine* 153 (2016): 156-164.

14. Nesson, Erik, and Joshua J. Robinson. "The Measurement of Health and the Connection between Health Inequality and Income." (2017).
15. Qiu, Qihua, et al. "Using spatial factor analysis to measure human development." *Journal of Development Economics* 132 (2018): 130-149.
16. Wolff, Heinz-Peter, Leif Wolf, Ali Subah, Joseph Guttman, Abdelraman Tamimi, Aiman Jarrar, Amer Salman, and E. Karablieh. "Methodological challenges in evaluating performance, impact and ranking of IWRM strategies in the Jordan Valley." *Water Science and Technology* 66, no. 7 (2012): 1407-1415.