HSR Health Services Research

DOI: 10.1111/1475-6773.12352
METHODS ARTICLE

Modeling Area-Level Health Rankings

Charles Courtemanche, Samir Soneji, and Rusty Tchernis

Objective. Rank county health using a Bayesian factor analysis model.

Data Sources. Secondary county data from the National Center for Health Statistics (through 2007) and Behavioral Risk Factor Surveillance System (through 2009).

Study Design. Our model builds on the existing county health rankings (CHRs) by using data-derived weights to compute ranks from mortality and morbidity variables, and by quantifying uncertainty based on population, spatial correlation, and missing data. We apply our model to Wisconsin, which has comprehensive data, and Texas, which has substantial missing information.

Data Collection Methods. The data were downloaded from www.countyhealthrankings.org.

Principal Findings. Our estimated rankings are more similar to the CHRs for Wisconsin than Texas, as the data-derived factor weights are closer to the assigned weights for Wisconsin. The correlations between the CHRs and our ranks are 0.89 for Wisconsin and 0.65 for Texas. Uncertainty is especially severe for Texas given the state's substantial missing data.

Conclusions. The reliability of comprehensive CHRs varies from state to state. We advise focusing on the counties that remain among the least healthy after incorporating alternate weighting methods and accounting for uncertainty. Our results also highlight the need for broader geographic coverage in health data.

Key Words. County, rank, health, factor analysis, Bayesian

Researchers consider a broad range of determinants when assessing population health and identifying areas of greatest need (Kindig, Asada, and Booske 2008; Institute of Medicine, 2011). Population health assessments are often presented to policy makers and communities as ranks, given their ubiquity and ease of interpretation (Erwin et al. 2011; Kanarek, Tsai, and Stanley 2011). Local area rankings can motivate stakeholders in lagging communities to design and promote local public health interventions. Such rankings can also assist policy makers with resource allocation decisions, which can be especially critical in times of declining local, state, and federal revenues. Understanding one's community's relative health status can help local officials assess the importance of public health initiatives relative to other priorities. At

the state and federal levels, knowledge of the least healthy communities can assist with funding decisions regarding local interventions and demonstration projects (Remington and Booske 2011).

Despite the potential usefulness of local area-level health rankings, two important difficulties arise in credibly assessing them. The first is the lack of a single comprehensive observable measure of health. This necessitates the use of some weighting procedure that combines available health-related variables into a summary measure. The second is the need to account for uncertainty arising from sources such as sampling error and missing data. Observable attributes of health are often not available at the population level and must be estimated using samples that can become small as the geographic area narrows. The amount of uncertainty can therefore be considerable at local levels such as the county. Moreover, data are often missing entirely for all or some components of health in certain localities, so inherently noisy procedures for imputing missing data are necessary to produce comprehensive rankings.

The most prominent local area-level health rankings are the county health rankings (CHRs), produced by the University of Wisconsin Population Health Institute (UWPHI) and begun in 2010. The CHRs address the difficulties inherent in local area-level health rankings by making strong assumptions. These rankings acknowledge the multifaceted nature of health, but fix subjectively assessed deterministic weights of each component in contributing to the overall health measure. The CHRs also do not account for uncertainty, despite their use of sample data for some components and an imputation process for missing data. It is therefore not possible to assess whether differences in counties' rankings are statistically meaningful. Ultimately, a primary purpose of CHRs is to concisely convey to policy makers how population health varies across a state. Yet rankings that do not include uncertainty instill a false sense of confidence among local and state officials when they identify the least and most healthy counties. Recently, Athens et al. (2013) attempted to account for uncertainty in CHRs coming from sampling error but did not address deterministic factor weights or missing data.

Address correspondence to Charles Courtemanche, Ph.D., Department of Economics, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302; e-mail: ccourtemanche@gsu.edu. Samir Soneji, Ph.D., is with the Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth College, Norris Cotton Cancer Center, Lebanon, NH. Rusty Tchernis, Ph.D., is with the Department of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA. Rusty Tchemis, Ph.D., and Charles Courtemanche, Ph.D., are also with the National Bureau of Economic Research, Institute for the Study of Labor (IZA), Atlanta, GA.

This paper develops an alternative method for ranking county health that addresses the issues of factor weighting and uncertainty through the use of a Bayesian hierarchal model for factor analysis. We treat health as a latent variable that depends on observable factors related to mortality and morbidity. The model empirically derives factor weights and measures uncertainty, which is inversely related to population size and accounts for spatial covariance. Incorporating information from neighboring area improves precision, which can be especially helpful in areas with small populations that would otherwise have high levels of uncertainty. These features of the model follow Hogan and Tchernis (2004), who ranked census tracts' levels of material deprivation in Rhode Island. We build on the Hogan and Tchernis (2004) framework by also including an iterative procedure to impute missing data, which is important in our context because a full array of health information is often not available for small localities. We incorporate uncertainty from the imputation process into the overall measure of uncertainty.

We apply our method to the CHR's county-level data for Texas (TX) and Wisconsin (WI). Though our model could be utilized for any state, we choose these two because TX is the state with the most counties while WI's rankings served as the UWPHI's template for the CHRs (Peppard et al. 2003, 2008). Our applications assess the importance of factor weighting and uncertainty. We first implement the model with uncertainty depending only on population and spatial covariance. For both states, we compare our data-derived factor weights to the CHRs' deterministic weights and compare the rankings obtained using our method to the CHRs. Uncertainty is evaluated by computing probability intervals (PIs) for the county ranks. We then show how the results change if uncertainty from imputing missing data is also considered, hypothesizing that this will increase the amount of uncertainty considerably in TX as the state has extensive missing data. We close by discussing implications for efforts to rank locality health.

METHODS

The Model

We incorporate a latent variable framework in which mortality and morbidity variables are observed manifestations of the latent construct—health. The underlying assumption of this framework is that health is not directly observed, but rather is manifested through a number of measurable variables. Latent variable frameworks have been widely adopted to assess quality of life

(McAuley et al. 2006), investigate geographic patterns of disease and their relationship with behavioral and social risk factors (Best and Hansell 2009), and measure health inequality across population subgroups (Murray, Gakidou, and Frenk 1999; van Doorslaer and Jones 2003).[1] We consider the same mortality and morbidity variables as the CHRs to facilitate comparison. In contrast to our latent variable framework, the CHRs utilize a deterministic framework in which the construct health outcome explicitly consists of a weighted combination of mortality and morbidity variables.

UWPHI determines the CHRs by calculating an overall health outcomes score based on standardized mortality and morbidity variables and their corresponding deterministic weights (Booske et al., 2010). 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 final scores for each county, which are then ranked.

We utilize a factor analysis model with spatially correlated factors to estimate the distribution of ranks for counties within a state. A factor analysis model given by Bartholomew and Knott (1999) explains the variability in observed variables Y_{ij} for county i as follows:

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

where μ_j is variable j's average across counties, the factor $\delta_i \sim N(0,1)$ represents latent health for county $i = 1, \ldots, n$, λ_j is the factor loading for variable $j = 1, \ldots, J$ (covariance between latent health and the observed variables), and $e_{ij} \sim N(0, \sigma_j^2)$ are the idiosyncratic error terms. The model assumes that the observed variables are influenced by the underlying latent health factor δ_i. The model is identified by decomposing the covariance matrix of the variables within the county into the correlation represented by the factors as the error terms, e_{ij}, are assumed to be uncorrelated, $\sigma_{ij} = 0 \forall \mathrm{j} \neq k$.

Stacking over the variables within the county, we can rewrite the model in vector notation as follows:

$$Y_i = \mu + \lambda \delta_i + e_i,$$

where λ is a vector of stacked λ_j and $\mathrm{Var}(e_i) = \sum = \mathrm{diag}\{\sigma_j^2\}$. Finally, stacking over the counties, we can write the model in hierarchical form:

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

$$\delta \sim N(0, I_n)$$

where $\Lambda = I_n \otimes \lambda$.

The next step is to introduce the population sizes in the variance of both the error terms and the factors. The assumption is that error terms and the factors in more populous counties have smaller variance. We define $M = \text{diag}\{m_i\}$, where m_i is the population of county i and specify the new model as:

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

In this specification, the variances are inversely proportional to the county population sizes.[2] In sum, our model accounts for stochastic uncertainty as well as uncertainty from sampling error and the factor loadings being estimated rather than known with certainty.

Spatial Correlation

The last step in our model's development introduces spatial dependence of the factors, as spatial spillovers may affect area health measures. Adding the spatial correlation matrix, ψ, the model can be rewritten as:

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

We use the conditional autoregressive specification (Besag, 1974; Sun et al., 1999), which produces a tractable relationship between the conditional and the marginal specifications. Hogan and Tchernis (2004) show that this specification performs well relative to several alternatives. We start from specifying the conditional relationship between the factor for county i and other counties in the neighborhood of i, R_i, and define the neighborhood as the set of counties adjacent to i:

$$\delta_{ji}|\{\delta_j : j \in R_i\} \sim N\left(\sum_{j \in R_i} \beta_{ij}\delta_j, v/\alpha_i\right)$$

As discussed in Hogan and Tchernis (2004), as part of the factor analysis model, we can only identify one parameter and thus restrict $\beta_{ij} = \omega$ and $v/\alpha_i = 1$. This specification models the conditional mean of the distribution of the factors as a weighted average of the factors from neigh-

boring counties, with higher values of ω representing stronger spatial dependence. This conditional specification results in the marginal distribution of $\delta \sim N(0, (I - \omega R)^{-1})$, where $R_{ij} = 1$ if a county j is adjacent to county i and $R_{ij} = 0$. Thus, $\Psi = (I - \omega R)^{-1}$ is a full matrix inducing the correlation between variables between counties. This specification induces a restriction on the support of ω to be between the reciprocals of the smallest and the largest eigenvalues of R.

Estimation

The model is estimated using Markov Chain Monte Carlo methods (Chib and Greenberg 1996). We use Gibbs Sampler (Gelfand and Smith 1990) with one Metropolis–Hastings (Chib and Greenberg 1995) step to obtain draws from ω. At each iteration of the sampler, we rank the posterior means of factors, resulting in one sample from the posterior distribution of ranks. The exact conditional distributions are summarized in Hogan and Tchernis (2004).

Missing Data

In our application, some of the values of the manifest variables Y_{ij} are unobserved and thus need to be imputed. The CHRs replace missing covariates with their corresponding state-level means, which is potentially problematic for two reasons. First, it ignores the uncertainty inherent in the imputation process. Second, imputing missing data with state averages may lead to biased rankings if counties with missing data are systematically more or less healthy than average. For example, the rank of a county missing data on all but one factor will approach the state average.

We consider multiple approaches to missing data. Our baseline model simply replaces the missing factors(s) with ordinary least squares (OLS) predictions based on the other factors. This solves the problem of counties with missing data automatically being drawn toward the middle, but it does not address uncertainty. We also replicated CHR's state averages approach, and the results were very similar to those obtained using OLS imputation. This suggests the biggest problem with the state averages imputation is its neglect of uncertainty, not its introduction of systematic bias. Our second imputation approach is therefore to sample from the distribution of missing values conditional on the parameters of the model using (1) at each iteration of the sampler (Rubin 1976; Little and Rubin 1987; Hogan and Tchernis 2004). This

incorporates the uncertainty of predicting missing values as part of the estimation, similarly to multiple imputations procedure.

Data

We implement the model using UWPHI data applied to the year 2011 county health outcome rankings in TX and WI. Mortality and morbidity data for the year 2011 were downloaded on June 1, 2012 (www.countyhealthrankings.org). The mortality variable is the years of potential life lost before age 75 years ("premature death"), estimated by UWPHI using 2005–2007 life table data from the National Center for Health Statistics (NCHS). The morbidity variables 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 <2,500 g ("low birthweight"). The first three morbidity variables were estimated by UWPHI using 2003–2009 data from the Behavioral Risk Factor Surveillance System (BRFSS) and the fourth using 2001–2007 birth certificate data from NCHS. The mortality variable is therefore based on the entire population, whereas the first three morbidity variables are from a representative survey and low birthweight reflects the universe of births.

UWPHI did not rank 31 of 254 counties in TX because they were missing at least four of the five variables. Of the remaining 223, 116 counties had missing data on at least one variable, and UWPHI applied their aforementioned imputation procedure to these counties. UWPHI ranked all 72 counties in WI, only two of which had missing data on any variables. To facilitate comparability, we focus only on the counties ranked by UWPHI. Our sample, therefore, consists of 223 TX counties and 72 WI counties.

RESULTS

Factor Weights

We estimate the model using data for TX and WI separately. The means, standard errors, and 95 percent PIs of the posterior distributions of all of the model's parameters are available in Appendix Table A1. We focus our discussion on the results of most interest, beginning with the estimated factor weights.

Table 1: County Health Ranking (CHR) Deterministic Weights and Normalized Square Correlations

Health Outcomes	*UWPHI* ω	*Texas* ρ^2 *(95% CI)*	*Wisconsin* ρ^2 *(95% CI)*
Premature deaths	0.50	0.14 (0.09, 0.19)	0.27 (0.17, 0.38)
Self-reported health status	0.10	0.24 (0.20, 0.29)	0.21 (0.12, 0.30)
Physically unhealthy days	0.10	0.41 (0.34, 0.48)	0.21 (0.11, 0.31)
Mentally unhealthy days	0.10	0.15 (0.10, 0.20)	0.17 (0.08, 0.25)
Low birthweight births	0.20	0.06 (0.02, 0.10)	0.15 (0.05, 0.24)

UWPHI, University of Wisconsin Population Health Institute; ω, weight; ρ^2, squared correlation; CI, credible interval.

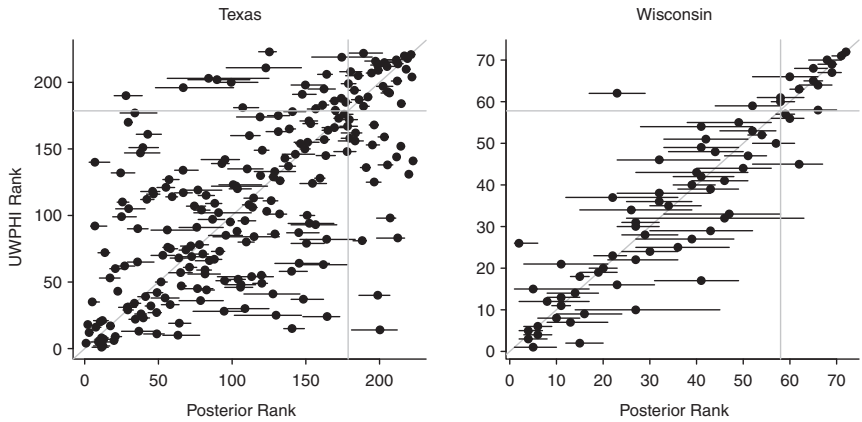
Table 1 compares the deterministic CHR weights to our normalized square correlations for the mortality and morbidity variables. Normalized square correlation represents the proportion of the variance in the variable that is explained by the factors and, therefore, is comparable to the CHR weights. Our square correlations differ between TX and WI and also differ from the CHR weights. For example, we estimate the squared correlation of the mean number of physically unhealthy days per month to be 0.41 for TX (95 percent CI, 0.34–0.48) and 0.21 for WI (95 percent CI, 0.11–0.31), whereas the CHRs set the weight of this variable to 0.10 for all states. The fact that the difference between our squared correlations and the CHR weights is greater for TX than WI suggests that our rankings will be less similar to the CHRs for TX.

Mean Rankings

For each county, we compute the posterior distribution of its health outcomes rank, including its mean and 95 percent PI. Tables showing our full county rankings for TX and WI are available upon request. Figure 1 illustrates the differences between our rankings and the CHRs, and the extent of uncertainty in the rankings. Specifically, we plot the middle 95 percent of the posterior distribution of ranks (horizontal line) and the mean of the posterior distribution (solid circle) relative to each county's CHR rank. These initial estimations use the OLS imputations for missing data.

We begin by discussing the differences between our (mean) rankings and the CHRs. Greater distances between solid circles and the 45-degree line indicate larger differences between our rankings and the CHRs. A solid circle to the right of the 45-degree line indicates that our method ranks the county

Figure 1: County Health Ranking (CHR) Ranks, Mean Posterior Ranks, and 95 percent Probability Intervals



Notes: The left (right) panel shows the posterior rank and CHR rank for each county in Texas (Wisconsin). The 95 percent probability interval of the posterior distribution 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-degree line represents equality between the CHRs and posterior ranks.

worse than the CHRs, and vice versa. Comparing the panels for TX and WI, there are more disagreements in TX than in WI; the intervals in the WI panel are much closer to the 45-degree line. The correlations between the CHR ranks and our ranks are 0.65 for TX (95 percent CI, 0.60–0.71) and 0.89 for WI (95 percent CI, 0.81–0.94).

Uncertainty

We next consider uncertainty in the rankings. The horizontal lines from Figure 1 suggest considerable uncertainty in both states. The 95 percent PIs range from 0 to 69 ranks wide in TX, with a mean width of 14.5 ranks and median width of 11 ranks. In WI, the 95 percent PIs range from 1 to 38 ranks wide, while the mean and median widths are 12 and 12 ranks, respectively.

We can use the measures of uncertainty to ask how far apart the ranks of two counties should be to give a researcher reasonable confidence that they are different (e.g., 90 or 95 percent confident). 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 percent of the posterior distribution of Harris County overlaps with the posterior distribution of Blanco County. In TX, there are 216 unique pairs of counties with mean rank $k = 5$ units apart and we calculate the percentage overlap for each pair. The median percentage overlap among the 216 pairs of counties is 12.8 percent.

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 in TX equals 40.8, 12.8, and 1.3 percent, respectively. Fifty percent of county pairs have 1 percent or less overlap when k exceeds 11, as do 75 percent of pairs when k exceeds 17 and 95 percent of pairs when k exceeds 30. 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 percent, respectively. Ten percent of county pairs have 1 percent or less overlap when k exceeds 11, as do 75 percent of pairs when k exceeds 16 and 95 percent of pairs when k exceeds 23. Thus, to be reasonably confident, say 90 percent, that two counties are different with respect to their health outcomes ranking, the distance between the mean of their health rank distribution should be approximately 25 counties apart in TX and 21 counties apart in WI.

Least Healthy Counties

A key purpose of the CHRs is identifying the least healthy counties (Kindig and Stoddart 2003). We therefore next examine how closely the least healthy counties identified by our rankings correlate to those from the CHRs. The vertical and horizontal lines in Figure 1 represent the 80th percentile (178th of 223 ranked counties in TX and 62nd of 72 in WI) separating the least healthy quintile of counties. According to our model, the least healthy counties will be those in which the mean of their rank distribution lies to the right of the vertical gray line. The least healthy counties according to the CHRs will be those with ranks above the horizontal gray line. Again, the differences are much more pronounced for TX than WI. In TX, 26 counties are classified as least healthy by both models, 19 are classified as least healthy by our model and not the CHRs, and 19 are classified as least healthy by the CHRs 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 the CHRs, and 2 are classified as least healthy by the CHRs and not our model. The correlation between the least healthy counties identified by our rankings and those from the CHRs

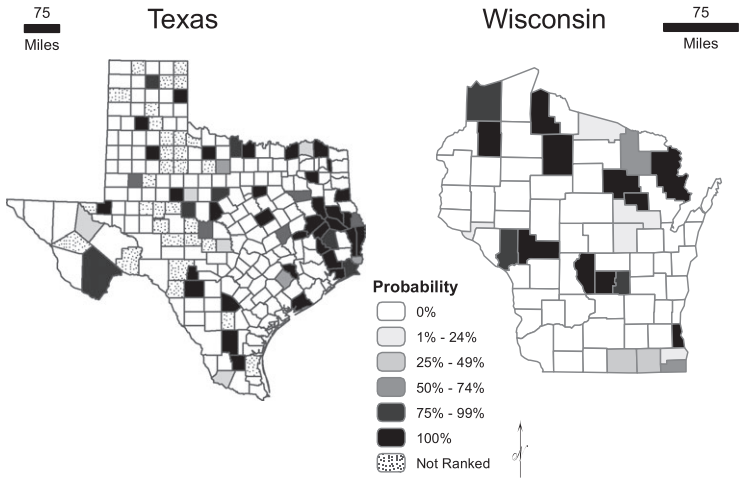
equals −0.20 for TX (95 percent CI, −0.26 to −0.07) and 0.32 for WI (95 percent CI, 0.16–0.54).

It is also useful to examine the degree of uncertainty in our identification of the least healthy counties, which we do by presenting the probability of each county being in the least healthy quintile in state maps in Figure 2. In TX, we observe a large concentration of unhealthy counties in East Texas. The probability of being in the least healthy quintile is 1.00 for Austin County, 0.63 for Colorado County, and 0.59 for Polk County. In other words, the entire posterior distribution of rankings for Austin County and 59 percent of the posterior distribution of rankings for Polk County lie to the right of 80th percentile. In WI, the probability is high for several of the northernmost counties (e.g., 0.96 for Douglas County) and several counties in the Milwaukee area (e.g., 1.00 for Milwaukee County and 0.54 for Kenosha County).

Missing Data

All our rankings thus far use OLS imputations for missing data. This approach does not account for uncertainty about the imputed values, perhaps leading to PIs that are too narrow and estimates of minimum ranking differences necessary for statistical significance that are too small. This is likely especially prob-

Figure 2: Probability of Being in Least Healthy Quintile, Texas and Wisconsin

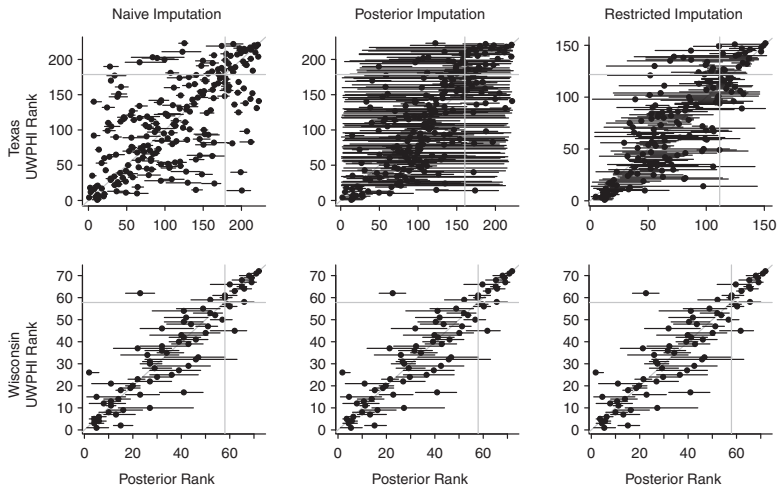


Notes: The left (right) panel shows the probability of being in the least healthy quintile under our model for Texas (Wisconsin). Unranked counties dotted.

lematic in TX, which has much more missing data than WI. Figure 3 therefore explores the sensitivity of our results to the use of the iterative imputation procedure. The left column ("Naïve Imputation") reproduces the results using the OLS imputations from Figure 2, while the middle column ("Posterior Imputation") uses the iterative procedure. Recall that our main sample includes all counties with valid data for at least one of the five factors. Another useful comparison is therefore to assess how the results change using a more restrictive set of counties. Thus, the right column ("Restricted Imputation") drops counties with more than one missing factor. This reduces the number of counties from 223 to 152 in TX, but it does not affect the number of counties in WI.

The bottom half of Figure 3 shows that incorporating uncertainty from imputing missing data has essentially no effect on the results for WI. This is because only two of WI's 72 counties have any missing data, and those two counties are missing only one factor. As no counties are missing more than one factor, the second and third figures are exactly the same.

Figure 3: County Health Ranking (CHR) Ranks, Mean Posterior Ranks, and 95 percent Probability Intervals with Different Approaches to Missing Data



Notes: The top (bottom) panel shows the posterior and CHR ranks for each county in Texas (Wisconsin). The 95 percent probability interval of the posterior distribution is denoted by a horizontal line and mean posterior rank by a solid circle. The gray horizontal and vertical lines represent the 80th percentile of ranks and the 45-degree line represents equality between the CHR and posterior ranks. The restricted imputation excludes 71 TX counties with more than one missing variable.

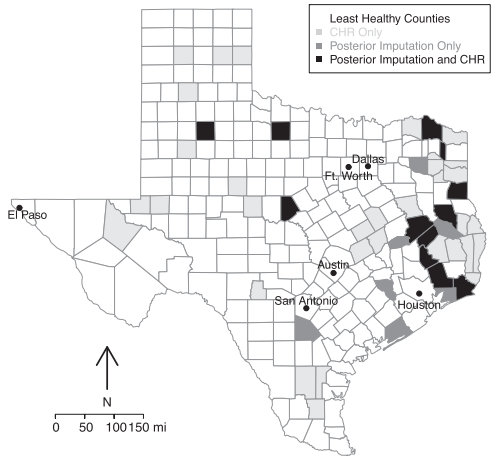
The top half of Figure 3 shows that the results are much different for TX. Keeping all counties with at least one nonmissing factor in the sample ("Posterior Imputation"), we see that incorporating uncertainty from imputing missing data drastically increases the overall amount of uncertainty. The 95 percent PIs for the 116 counties with at least one missing factor generally expand to include almost the entire range of possible ranks. Even the PIs for counties with no missing data widen considerably in most cases because of the shared component in the uncertainty measure. When we drop the counties with two or more missing factors ("Restricted Imputation"), the level of uncertainty drops relative to the previous "Posterior Imputation" but remains considerably greater than "Naïve Imputation." Of the 152 remaining counties, 71 have 95 percent PIs that reach into the "least healthy quintile" range and 47 of these extend into the "least healthy decile" range.

The TX results illustrate the broader point that when the amount of missing data is extensive, it may be difficult to obtain comprehensive county rankings after accounting for all sources of uncertainty. However, clear conclusions can still be reached for some counties. For instance, even in the full-sample TX graph with the iterative imputation process, the 95 percent PIs for 19 counties lie entirely to the right of the 80th percentile line. These 19 counties can therefore confidently be identified as among the least healthy. Figure 4 maps these counties, which are mostly in eastern TX. The figure further shows that 12 of these counties are also among the least healthy according to the CHRs. These 12 counties' classification as least healthy therefore not only holds up after accounting fully for uncertainty but is also robust to the use of quite different factor weights.

DISCUSSION

This paper implements a Bayesian hierarchal model for factor analysis to rank the health of localities, where health is a latent variable that depends on observable factors related to mortality and morbidity. Our model builds on the CHRs by using a data-driven process to determine factor weights and including a measure of uncertainty that incorporates population, spatial covariance, and missing data. Applying our method to county-level data from TX and WI reveals the importance of these innovations. We show that our data-derived factor weights differ substantially from the deterministic CHR weights in TX but less so in WI. Consequently, our rankings are much more similar to the CHRs for WI. We also document considerable uncertainty in

Figure 4: Texas (TX) Counties That Are Definitively among the Least Healthy in Posterior Imputation



Notes: The PIs of counties shaded black or dark gray lie entirely to the right of the 80th percentile using our posterior imputation. The counties shaded black are also classified as among the least healthy 20 percent by the county health rankings (CHRs), whereas those shaded dark gray are not. The counties shaded light gray are classified as among the least healthy 20 percent by the CHRs but cannot be definitively classified as such by our model.

both states that rises sharply in TX after accounting for uncertainty from imputing the state's substantial missing data. It becomes impossible to reach clear conclusions for most counties in TX, although some of the least healthy counties can still be identified.

Our framework allows numerous variations in future research on health rankings. Our model could be estimated in other states using the same five mortality and morbidity variables. It would also be straightforward to produce alternative rankings using additional health-related variables. Indeed, numerous other measures of health outcomes are routinely measured at the population level (Kindig 2007). Our latent variable framework could incorporate these additional manifestations of health and empirically derive their relationship on health outcomes without requiring subjective expert opinion on variable weights. Moreover, our method could easily be used to produce rankings at geographic levels besides the county, such as the state or metropolitan statistical area. The existing state-level America's Health Rankings (United Health Foundation, 2010) suffer from

the same limitations as the CHRs, although sampling error and missing data are probably less problematic at larger geographic levels. Our ranking methodology could even be extended to international comparisons of health system performance or human development (United Nations Development Programme, 2011).

Our approach also allows different users to vary the necessary level of confidence. Academic researchers generally have objectives that require a high degree of confidence, such as assessing how population health varies geographically or estimating the impact of an intervention. Alternatively, state-level policy makers may need to make a timely resource allocation decision with the best available information, in which case the necessary level of confidence may be lower. To illustrate, in TX we conclude that mean rank of counties should be at least 30 ranks apart to provide 95 percent confidence of a ≤10 percent difference in health rankings, but only 11 ranks apart to provide 80 percent confidence of a ≤10 percent difference.

Additionally, our method allows for flexibility about the geographic level at which the factor weights vary. Our applications take the perspective of state-level policy makers, so we conduct separate estimations for WI and TX and allow the weights to vary between states. If the objective were instead to compute one set of CHRs for the entire United States, then a single estimation with equal factor weights across all states would be appropriate.

We acknowledge several limitations in this study. First, only using data from two states limits the generalizability of our conclusions. Second, we rank only 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. Next, while rankings are useful to compare the health of populations, they do not convey absolute differences. A county's health ranking may improve even though its population became less healthy if the health of other counties declines faster over time. Additionally, our research does not address why some counties are healthier than others. Additional research is needed to understand how medical, social, and physical determinants interact to produce health and perpetuate disparities (Stoddart 1995; Kawachi, Subramanian, and Almeida-Filho 2002). Also, ranking health at the county level masks heterogeneity within counties along dimensions such as age, geography, income, national origin, and primary language. These distinctions may lead to relatively healthy and sick subpopulations, in which case an observed indicator based largely on one subpopulation (e.g., younger adults largely influence birthweight) may not load on a single factor driven largely

by the other subpopulation (e.g., older adults). Similarly, we model health as a single latent factor, whereas health could also be modeled as consisting of multiple domains (e.g., physical and mental health). Murray et al. (2013), for example, show how to use copula latent factor models, which perform well particularly with non-Gaussian distributed data. Finally, while our emphasis is on the accuracy of county health measurement, future research should investigate other criteria such as efficiency, reliability, and replicability. One recent example is Arndt et al. (2013), who examine variability in reliability of indexes based on different variables used in CHR. They show that health measures vary in their ability to provide consistent ranks across states, with population-based measures generally performing better than survey-based measures.

In conclusion, policy makers will find CHRs most useful when they can draw meaningful and valid conclusions about the least healthy counties that require additional attention and the healthiest counties that can serve as models for public health initiatives. Yet current CHRs suffer from limitations that make local and state officials vulnerable to faulty conclusions. By allowing for uncertainty, our results reveal instances where apparent geographic differences are misleading. States with considerable missing data might be better served by focusing on the counties that can be precisely identified as being among the least healthy, rather than attempting to utilize rankings of all counties. Alternatively, our results could be interpreted as calling for broader geographic coverage in health data.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: We acknowledge helpful feedback from Guy David, Paul Ferraro, Tom Getzen, Barry Hirsch, Dann Millimet, Patrick Remington, Chris Ruhm, Harold Sox, and audiences at Clemson University, University of California at Merced, Georgia State University, Universidad del Pacifico, Eastern Economics Conference, Georgia Health Economics Research Day, and Southern Economic Association Annual Meeting. This research was supported by The Norris Cotton Cancer Center's GeoSpatial Resource, which is funded by 5P30CA023108 Cancer Center Support Grant.

Disclosures: None.

Disclaimers: None.

NOTES

1. While there is therefore precedent for our approach of modeling health as a single latent construct, we acknowledge that health could also be modeled as having multiple distinct domains.
2. Ideally, we would use sample sizes rather than populations, but this would complicate the model considerably because the factors come from different datasets with different numbers of observations per county. Using population instead of sample sizes should not meaningfully impact the results, as counties with larger populations should have larger sample sizes.

REFERENCES

Arndt, S., L. Acion, K. Caspers, and P. Blood. 2013. "How Reliable Are County and Regional Health Rankings?" *Prevention Science* 14: 497–502.

Athens, J. K., B. B. Catlin, P. L. Remington, and R. E. Gangnon. 2013. "Using Empirical Bayes Methods to Rank Counties on Population Health Measures." *Preventing Chronic Disease* 10 [accessed on February 8, 2014]. Available at http://www.cdc.gov/pcd/issues/2013/pdf/13_0028.pdf

Bartholomew, D., and M. Knott. 1999. *Latent Variable Models and Factor Analysis.* New York: Oxford University Press.

Best, N., and A. L. Hansell. 2009. "Geographic Variations in Risk: Adjusting for Unmeasured Confounders through Joint Modeling of Multiple Diseases." *Epidemiology* 20 (3): 400–10.

Besag, J. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society* 36 (2): 192–236.

Booske, B., J. Athens, D. Kindig, H. Park, and P. Remington. 2010. "Different Perspectives for Assigning Weights to Determinants of Health." University of Wisconsin Population Health Institute [accessed July 19, 2013]. Available at http://www.countyhealthrankings.org/sites/default/files/differentPerspectivesForAssigningWeightsToDeterminantsOfHealth.pdf.

Chib, S., and E. Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician* 49 (4): 327–35.

———. 1996. "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory* 12: 409–31.

van Doorslaer, E., and A. M. Jones. 2003. "Inequalities in Self-Reported Health: Validation of a New Approach to Measurement." *Journal of Health Economics* 22 (1): 61–87.

Erwin, P. C., C. Myers, G. Myers, and L. Daugherty. 2011. "State Responses to America's Health Rankings: The Search for Meaning, Utility, and Value." *Journal of Public Health Management and Practice* 17 (5): 406–12.

Gelfand, A. E., and A. F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85: 398–409.

Hogan, J. W., and R. Tchernis. 2004. "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 (466): 314–24.

Institute of Medicine. 2011. *For the Public's Health: The Role of Measurement in Action and Accountability.* Washington, DC: The National Academies Press.

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

Kawachi, I., S. V. Subramanian, and N. Almeida-Filho. 2002. "A Glossary for Health Inequalities." *Journal of Epidemiology and Community Health* 56 (9): 647–52.

Kindig, D. 2007. "Understanding Population Health Terminology." *Milbank Quarterly* 85 (1): 139–61.

Kindig, D., and G. Stoddart. 2003. "What Is Population Health?" *American Journal of Public Health* 93 (3): 380–3.

Kindig, D., Y. Asada, and B. Booske. 2008. "A Population Health Framework for Setting National and State Health Goals." *Journal of the American Medical Association* 299 (17): 2081–3.

Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data.* New York: John Wiley and Sons.

McAuley, E., J. F. Konopack, R. W. Motl, K. S. Morris, S. E. Doerksen, and K. R. Rosengren. 2006. "Physical Activity and Quality of Life in Older Adults: Influence of Health Status and Self-Efficacy." *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 31 (1): 99–103.

Murray, C. J., E. E. Gakidou, and J. Frenk. 1999. "Health Inequalities and Social Group Differences: What Should We Measure?" *Bulletin of the World Health Organization* 77 (7): 537–43.

Murray, J. S., D. B. Dunson, L. Carin, and J. E. Lucas. 2013. "Bayesian Gaussian Copula Factor Models for Mixed Data." *Journal of the American Statistical Association* 108 (502): 656–65.

Peppard, P., D. Kindig, A. Riemer, E. Dranger, and P. Remington. 2003. *Wisconsin County Health Rankings, 2003.* Madison, WI: Wisconsin Public Health and Health Policy Institute.

Peppard, P. E., D. A. Kindig, E. Dranger, A. Jovaag, and P. L. Remington. 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): 209–12.

Remington, P. L., and B. C. Booske. 2011. "Measuring the Health of Communities —How and Why?" *Journal of Public Health Management and Practice* 17 (5): 397–400.

Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63: 581–92.

Stoddart, G. 1995. *The Challenge of Producing Health in Modern Economies.* Centre for Health Economics and Policy Analysis Working Paper Series No. 1995-15, Centre for Health Economics and Policy Analysis (CHEPA). Hamilton, Canada: McMaster University.

Sun, D., R. K. Tsutakawa, and P. L. Speckman. 1999. "Posterior Distribution of Hierarchal Models Using CAR(1) Distributions." *Biometrika* 86: 341–50.

United Health Foundation. 2010. *America's Health Rankings.* Minnetonka, MN: United Health Foundation, American Public Health Association, Partnership for Prevention.

United Nations Development Programme. 2011. *Human Development Report 2011.* New York: United Nations.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table A1. Means, Standard Deviations, and 95 percent PIs of Posterior Distributions of Model Parameters.