**BMJ Open**

# Decomposition of the US black/white inequality in premature mortality, 2010–2015: an observational study

Mathew V Kiang ![ORCID],[1,2] Nancy Krieger,[2] Caroline O Buckee,[3,4] Jukka Pekka Onnela,[5] Jarvis T Chen[2]

[1]Center for Population Health Sciences, Stanford University, Palo Alto, California, USA
[2]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[4]Center for Communicable Disease Dynamics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

**Correspondence to**
Dr Mathew V Kiang;
mkiang@stanford.edu

## ABSTRACT

**Objective** Decompose the US black/white inequality in premature mortality into shared and group-specific risks to better inform health policy.

**Setting** All 50 US states and the District of Columbia, 2010 to 2015.

**Participants** A total of 2.85 million non-Hispanic white and 762 639 non-Hispanic black US-resident decedents.

**Primary and secondary outcome measures** The race-specific county-level relative risks for US blacks and whites, separately, and the risk ratio between groups.

**Results** There is substantial geographic variation in premature mortality for both groups and the risk ratio between groups. After adjusting for median household income, county-level relative risks ranged from 0.46 to 2.04 (median: 1.03) for whites and from 0.31 to 3.28 (median: 1.15) for blacks. County-level risk ratios (black/white) ranged from 0.33 to 4.56 (median: 1.09). Half of the geographic variation in white premature mortality was shared with blacks, while only 15% of the geographic variation in black premature mortality was shared with whites. Non-Hispanic blacks experience substantial geographic variation in premature mortality that is not shared with whites. Moreover, black-specific geographic variation was not accounted for by median household income.

**Conclusion** Understanding geographic variation in mortality is crucial to informing health policy; however, estimating mortality is difficult at small spatial scales or for small subpopulations. Bayesian joint spatial models ameliorate many of these issues and can provide a nuanced decomposition of risk. Using premature mortality as an example application, we show that Bayesian joint spatial models are a powerful tool as researchers grapple with disentangling neighbourhood contextual effects and sociodemographic compositional effects of an area when evaluating health outcomes. Further research is necessary in fully understanding when and how these models can be applied in an epidemiological setting.

## INTRODUCTION

Understanding differences in health outcomes across geographic and sociodemographic subpopulations is essential for improving population health. Research has consistently found large differences in health outcomes by geography and

### Strengths and limitations of this study

► Descriptive epidemiology is often limited in situations where areas or subpopulations may have few observations of the health outcome of interest. We present a small area estimation method used in spatial statistics that allows estimating these outcomes in a disciplined way.

► This method may be useful for understanding spatial patterns, generating new hypotheses and estimating health outcomes when data are limited.

► Like all models, this method makes assumptions that must be assessed and carefully explored and is limited to descriptive (ie, non-causal) questions.

sociodemographic characteristics including race/ethnicity and income.[1–3] This important heterogeneity is masked when using aggregated measures of health such as national life expectancy.[4 5] Measuring how health varies across geographic space and sociodemographic characteristics is crucial to informing health policy and improving overall health.

Examining geographic variation of health outcomes is often complicated due to small populations or few observations of the outcome of interest. This data sparsity results in highly variable estimates, making it difficult to understand the underlying risk. In addition, investigating geographic variation in health disparities between two subpopulations is difficult when using these volatile estimates, especially for relative measures where unstable estimates in the denominator may result in undefined or extreme over/under-estimates.[6]

In order to address these issues of small area estimation, researchers employ three common approaches—aggregation, suppression and multilevel modelling. First, researchers may choose to aggregate over geography, time, sociodemographic characteristics or a combination of these to obtain sufficiently large sample sizes for stable

estimates. Second, researchers may choose to suppress estimates from areas with small populations or remove areas with inadequate sample size from their analyses. While convenient, these two approaches result in loss of information, thus limiting their ability to inform health policy. Third, researchers may use Bayesian hierarchical models to pool information across geography, time or subpopulations. However, most Bayesian hierarchical approaches share information across geography or time while treating subpopulations independently.

One type of Bayesian hierarchical model that treats subpopulations in a more sophisticated way is the shared component model. This model jointly estimates the spatial variation in two subpopulations, which ameliorates issues of sparse data by pooling information across both geographic areas and subpopulations.[7] The simultaneous modelling of common spatial patterns of risk across different population groups can be conceptualised as providing evidence of latent risk factors that have shared spatial structure. Conversely, this model also estimates divergent spatial patterns, which may provide evidence of risk factors unique to only one group.[8] Finally, joint modelling approaches also produce estimates that are more precise than separate stratified models, especially when one subpopulation is small relative to the other.[8 9]

While originally developed for jointly modelling two related diseases, this model has also been used to estimate geographic variation in health inequalities by gender[10] and racial/ethnic disparities in epilepsy[11] and infant mortality.[6] However, the shared component model is still underused in the social epidemiology literature. In the cases where it is used, computational restrictions often require estimating only a single region[11] or using a simplified version of the model.[6]

In the USA, racial/ethnic disparities in a range of health outcomes, including premature mortality, have been noted since the earliest health records,[12] fluctuating in magnitude,[13] while remaining spatially persistent over time.[14] Premature mortality is a key indicator of population health and is often used to guide health policy worldwide.[15] Many premature deaths are considered preventable[16] and thus amenable preventive health interventions. In addition to reducing population-wide premature mortality, a central goal of the US Healthy People 2020 initiative[17] and the WHO Health for All programme[18] is to reduce health disparities between subpopulations. We used the shared component model to estimate non-Hispanic white and non-Hispanic black premature mortality risk for 3108 US counties. In addition, for each county, we calculated the racial/ethnic disparity in terms of risk ratios.

The shared component model decomposes geographic variation in premature mortality risk into three spatial 'risk surfaces'. First, the shared surface represents unobserved covariates that may explain geographic variation for both blacks and whites. In addition, the model produces two race/ethnicity-specific risk surfaces, which represent the risk in each county that is specific to the

corresponding racial/ethnic group. In this paper, we present a computationally feasible way of estimating the full shared component model on the entire contiguous USA. While this model is general enough for any pair of outcomes or groups, we present an application of the method to black/white disparities in premature mortality in the USA. We show that this nuanced decomposition of premature mortality risk at the county level could be used to inform health policy, create targeted intervention strategies and improve health while decreasing geographic and racial/ethnic disparities in US premature mortality.

## METHODS

### Data

We used the US National Center for Health Statistics' compressed mortality files and accompanying population files[19] to tabulate premature deaths by sex, race/ethnicity and 5-year age category (0–4, 5–9, …, 85+) for non-Hispanic whites and non-Hispanic blacks—hereafter referred to as 'whites' and 'blacks'—from 2010 to 2015. Following previous studies,[13 20] we conservatively defined premature mortality as death before age 65. The majority of counties observed fewer than 10 deaths among the black population (see online supplementary appendix figure 1). We calculated age/sex-standardised mortality ratios (SMRs) using the indirect method with the overall national premature mortality rate, including both non-Hispanic whites and non-Hispanic blacks, as the reference rate. Data are available by request using the National Vital Statistics System website (https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm).

We used the 5-year American Community Survey to extract 2015 median income in the previous 12 months (in 2015 dollars) of all households for each county.[21] Income was mean-centred and scaled to \$10 000 per unit. Four counties (0.1%) with missing values for median income were mean-imputed (see online supplementary appendix figure 2).

County adjacency was defined using the US Census Bureau's 2010 County Adjacency File.[22] Consistent with previous research,[6] we limited our analysis to the contiguous USA and the District of Columbia.

### Patient and public involvement

This study used de-identified data on deceased individuals and did not require patient or public involvement.

### Analysis

#### Shared component model

Our model builds on the Knorr-Held and Best[7] shared component model. Specifically, assume we have observed and expected counts $O_{dij}$ and $E_{dij}$, respectively, for each race/ethnicity $d \in \{1, 2\}$ within counties $i = 1, \ldots, I$, which are nested in states $j = 1, \ldots, J$. The county-level SMR is $SMR = O_{dij}/E_{dij}$. The county-specific relative risks $\lambda_{1ij}$ and $\lambda_{2ij}$ are modelled hierarchically as a log-linear function of a Poisson distribution. Each race/ethnicity $d$

is modelled with group-specific intercepts $\alpha_d$, vectors of coefficients $\boldsymbol{\beta}_{(d)}$, design matrices $\boldsymbol{X}_{(d)}$ and spatial parameter $\psi_{dij}$. Both racial/ethnic groups have a shared spatial parameter $\phi_{ij}$, which is scaled by $\delta$. In addition to the county-level spatial parameters, there is an independent, normally distributed state random effect $\eta_i$, which represents our belief that multiple levels of governmental characteristics and policies can affect health.

$$O_{1ij} \sim \text{Poisson}(E_{1ij}\lambda_{1ij}), O_{2ij} \sim \text{Poisson}(E_{2ij}\lambda_{2ij})$$
$$\log(\lambda_{1ij}) = \alpha_1 + \boldsymbol{\beta}_{(1)}\boldsymbol{X}_{(1)} + \phi_{ij}\delta + \psi_{1ij} + \eta_j$$
$$\log(\lambda_{2ij}) = \alpha_2 + \boldsymbol{\beta}_{(2)}\boldsymbol{X}_{(2)} + \phi_{ij}/\delta + \psi_{2ij} + \eta_j$$

The racial/ethnic-specific spatial components are assumed to be independent with $\psi_{1ij}$ and $\psi_{2ij}$ representing race/ethnicity-specific risk separately. Conversely, $\phi_{ij}$ represents a latent risk surface that is shared between both blacks and whites. All three spatial components are assigned the Besag-York-Mollié (BYM) convolution prior. In a BYM prior, the spatial effect of a particular county is dependent on the effects of all its neighbours. In addition, the BYM includes an unstructured random effect to account for independent county-specific noise.[23] The magnitude of the shared component may differ by the scaling factor $\phi_{ij}$, which is estimated from the data and allows each race/ethnicity to have a different 'risk gradient'. A simplified schematic of the shared component model is shown in online supplementary appendix figure 3.

## Priors and hyperpriors

We assigned both intercepts a flat prior of $\alpha \sim \text{Uniform}(-\infty, +\infty)$. For the income-adjusted model, the coefficient of county median household income was assigned a weakly informative prior of $\beta \sim \text{Normal}(0, 100)$. The inverse square of the scaling factor, $1/\delta^2$, can be interpreted as the risk ratio of the shared component. Accordingly, we assigned a prior of $\delta \sim \text{logNormal}(0, 0.412)$, which corresponds to our belief that the risk ratio of the shared component has a 95% probability being in the interval $\left[\frac{1}{5}, 5\right]$ and is consistent with previous research.[6 7] We used weakly informative hyperpriors for the precision terms in the BYM components. Following Gelman,[24] we reparameterised our precisions as SD with priors of $\sigma = \frac{1}{\sqrt{\tau}} \sim \text{halfNormal}(0, 5)$.

## Risk ratio

For each county, we calculated the within-county black/white disparity in premature mortality on the relative scale, which we refer to as the 'risk ratio' to differentiate it from the area-specific relative risk $\lambda_{ij}$. Specifically, the log risk ratio was defined from the earlier equations as:

**Table 1** Distribution of counts of deaths and population for non-Hispanic blacks and non-Hispanic whites under 65, and county-level standardised mortality ratios (SMRs) by race/ethnicity for the USA, 2010–2015. Modelled SMRs are county-level posterior median values with and without adjusting for county median household income. All SMRs are in reference to the total US population after sex/age-standarisation

| | Total | Min | 5th percentile | Median | 95th percentile | Max |
|---|---|---|---|---|---|---|
| Deaths (#) | | | | | | |
| Non-Hispanic white | 2 832 444 | 1 | 28 | 346 | 3631 | 34 336 |
| Non-Hispanic black | 762 639 | 1 | 1 | 28 | 1154 | 29 206 |
| Population (#) | | | | | | |
| Non-Hispanic white | 995 524 222 | 144 | 8595 | 98 872 | 1 344 399 | 13 811 154 |
| Non-Hispanic black | 220 875 881 | 1 | 117 | 4942 | 278 157 | 6 689 888 |
| Raw county-specific SMR | | | | | | |
| Non-Hispanic white | | 0.09 | 0.63 | 1.04 | 1.60 | 2.48 |
| Non-Hispanic black | | 0.00 | 0.00 | 1.15 | 2.30 | 16.62 |
| Modelled county-specific relative risk (no income adjustment) | | | | | | |
| Non-Hispanic white | | 0.36 | 0.66 | 1.03 | 1.57 | 2.46 |
| Non-Hispanic black | | 0.31 | 0.69 | 1.13 | 1.87 | 2.60 |
| Modelled county-specific relative risk (income-adjusted) | | | | | | |
| Non-Hispanic white | | 0.46 | 0.76 | 1.03 | 1.39 | 2.04 |
| Non-Hispanic black | | 0.31 | 0.73 | 1.15 | 1.76 | 3.28 |
| Within-county black/white risk ratio of premature mortality | | | | | | |
| No income adjustment | | 0.33 | 0.75 | 1.09 | 1.67 | 5.07 |
| Income-adjusted | | 0.33 | 0.75 | 1.09 | 1.65 | 4.56 |

SMR, standardised mortality ratio.

$$\log\left(RR_{ij}\right) = \left(\alpha_2 - \alpha_1\right) + \phi_{ij}\left(\frac{1}{\delta} - \delta\right) + \left(\psi_{2ij} - \psi_{1ij}\right)$$

$$log(RR_{ij}) = log\left(\lambda_{2ij}\right) - log\left(\lambda_{1ij}\right)$$

$$log(RR_{ij}) = \left(\alpha_2 - \alpha_1\right) + \phi_{ij}(\frac{1}{\delta} - \delta) + \left(\psi_{2ij} - \psi_{1ij}\right)$$

## Computation

All analyses were conducted using the statistical programming language R V.3.3.2[25] with the RStan V.2.14.1 package.[26] The split-chain Gelman-Rubin $R$ diagnostic[27] was used to check convergence. All parameters had $R<1.10$, indicating convergence with adequate effective sample size. Due to computational limitations, we did not conduct formal sensitivity analyses of model parameters and hyperparameters.

To quantify the between-county variation, we calculated the empirical variance for each risk surface. Similarly, we calculated the empirical variance of the overall (log) county-specific relative risks and the fraction of total spatial empirical variation shared by both racial/ethnic groups.

We estimated the posterior probability of excess relative risk greater than 1 or reduced relative risk less than 1 for each region, each surface and for the spatially varying disparity. Consistent with previous literature,[28] we used a cutpoint of 80%, which simulation studies have shown to have adequate sensitivity and specificity,[29] and only plot counties that meet this cutpoint. Estimates for all counties and their posterior probability are shown in the online supplementary appendix figures 4 and 5.

## RESULTS

### Standardised mortality ratios and county-specific relative risks

From 2010 to 2015, 2.85 million non-Hispanic white premature deaths and 762 639 non-Hispanic black premature deaths were recorded from observed populations of 995.5 million and 220.9 million, respectively (table 1). The spread of county-level premature mortality risk for blacks was greater than that of whites in both the raw data (SMRs) and modelled relative risk estimates (table 1). High white SMRs are observed in Appalachia, Oklahoma and other southern states, as well as parts of Nevada and Arizona (figure 1). In the income-adjusted model, the 95th percentile county had three times higher white relative risk than the 5th percentile county (table 1). High black relative risks are observed in the South, Mid-Atlantic region and parts of California and Nevada (figure 1). In the income-adjusted model, the 95th percentile county had 5.6 times higher black relative risk than the 5th percentile county (table 1). Model parameters are shown in online supplementary appendix table 1.

County median household income was associated with lower premature mortality rates for both whites and blacks. Specifically, a $10 000 increase in median household income corresponded to within-race rate ratio of 0.88 (95% uncertainty interval (UI): 0.87 to 0.88) and
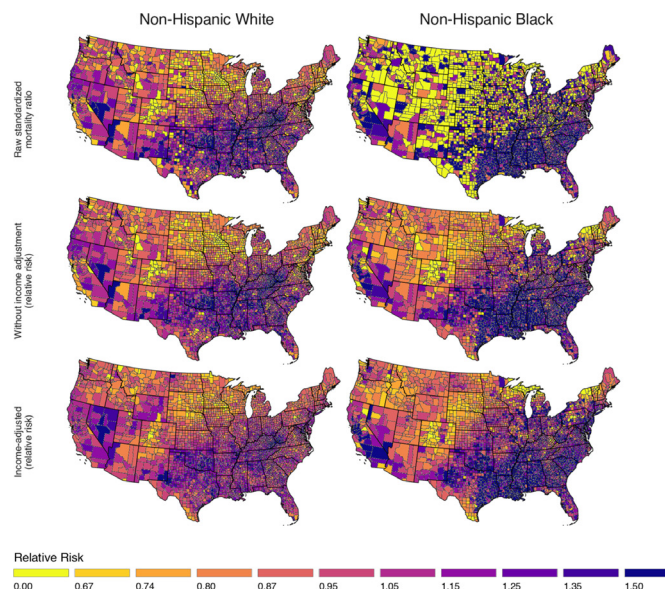


**Figure 1** County-level premature mortality risk by race/ethnicity after sex- and age-standardisation. The top row is the unsmoothed (raw) standardised mortality ratio. The middle row is smoothed county-specific relative risk with no income adjustment. The bottom row is smoothed county-specific relative risk and adjusted for county-level median household income.

0.91 (95% UI: 0.90 to 0.92) for whites and blacks, respectively. After adjusting for county median household income, the between-county empirical variance in (log) relative risk was reduced. Specifically, the between-county
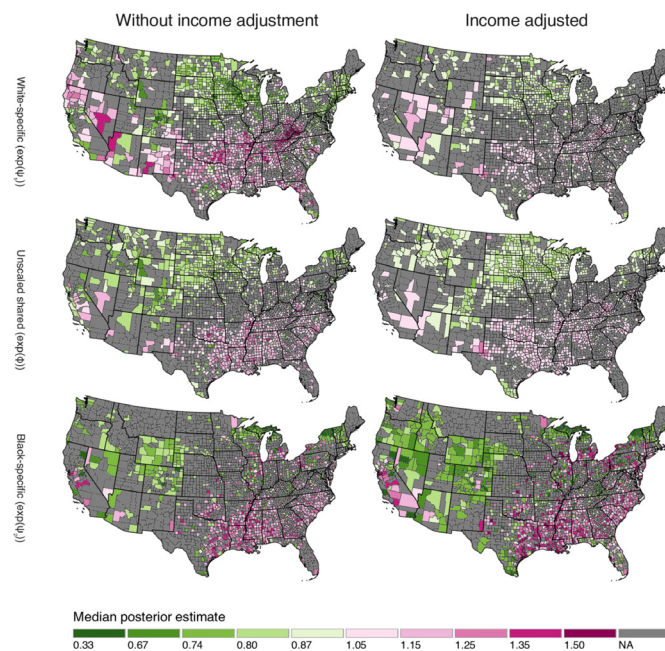


**Figure 2** County-level white-specific (top), shared (middle) and black-specific (bottom) premature mortality risk before (left) and after (right) adjusting for county median household income. Both models use sex/age-standardised rates. Non-significant counties are grey. Significant counties are defined as counties with greater than 80% of posterior estimates above or below 1.

empirical variance in white (log) relative risk was reduced 50% from 0.08 (95% UI: 0.07 to 0.08) to 0.04 (95% UI: 0.03 to 0.04). On the other hand, adjustment for county median household income reduced the larger between-county variance in the black (log) relative risk by only 14%, from 0.14 (95% UI: 0.13 to 0.16) to 0.12 (95% UI: 0.11 to 0.13). The between-state variance remained small (0.00; 95% UI: 0.00 to 0.01) in both models.

## Spatial components
### Geographic variation of shared risk $\phi_{ij}$
Before adjusting for county median household income, patches of high shared risk exist in the South and around Appalachia (figure 2, middle row). Meanwhile, the Midwest and Pacific Northwest have areas of low shared risk; however, after accounting for income, the shared risk is attenuated (online supplementary appendix figure 6) such that only Appalachia and parts of Texas remain at an elevated shared risk. After income adjustment, the between-county variance of the shared (log) risk decreased 33% from 0.03 (95% UI: 0.02 to 0.05) to 0.02 (95% UI: 0.01 to 0.03).

### Geographic variation of white-specific risk $\psi_{1ij}$
The white-specific risk, $\psi_{1ij}$, in the unadjusted model shows significant excess risk in Appalachia and parts of the South and Nevada (figure 2, top row) as well as lower risk in the Midwest and coastal Mid-Atlantic states. However, after adjusting for county median household income, there is substantial attenuation of white-specific

risk (online supplementary appendix figure 6) with some elevated risk still present in Appalachia and the lower risk in the Mid-Atlantic no longer significant. The between-county variance of the white-specific log risk decreased 60% from 0.05 (95% UI: 0.03 to 0.06) to 0.02 (95% UI: 0.00 to 0.03). The fraction of the total geographic variation of premature mortality risk among whites that is shared with blacks increased after adjustment for county median income, from 30% (95% UI: 13% to 36%) to 52% (95% UI: 13% to 88%) (table 2).

### Geographic variation of black-specific risk $\psi_{2ij}$
In the unadjusted model, the black-specific risk, $\psi_{2ij}$, is elevated in the South, Mississippi Delta and parts of California (figure 2, bottom row). There is reduced risk in the Northeast, Colorado and northern Michigan. However, unlike the shared and white-specific risks, the geographic variation in the black-specific risk was amplified after adjusting for county median household income (online supplementary appendix figure 6), with the between-county variance increasing 29% from 0.07 (95% UI: 0.03 to 0.11) to 0.09 (95% UI: 0.05 to 0.11). Specifically, high-risk counties in the South become even higher risk while low-risk counties of the Midwest and West become even lower risk (figure 2) with more high-risk areas in Wisconsin and California. The fraction of total geographic variation of black risk that is shared with whites decreased after adjustment for county median household income from 42% (95% UI: 23% to 81%) to 15% (95% UI: 6% to 57%) (table 2).

**Table 2** Posterior median (95% uncertainty interval) for variance components. The adjusted model takes into account county median household income while the unadjusted model does not. Both models are based on sex/age SMRs

| | | Unadjusted | Adjusted |
|---|---|---|---|
| Between-state variance in (log) risk | $\sigma_\eta^2$ | 0.00 (0.00 to 0.01) | 0.00 (0.00 to 0.01) |
| Total between-county variance in (log) relative risk | | | |
| Non-Hispanic white | $Var\left(\alpha_1 + \phi_{ij}\delta + \psi_{1ij} + \eta_j\right)$ | 0.08 (0.07 to 0.08) | 0.04 (0.03 to 0.04) |
| Non-Hispanic black | $Var\left(\alpha_2 + \frac{\phi_{ij}}{\delta} + \psi_{2ij} + \eta_j\right)$ | 0.14 (0.13 to 0.16) | 0.12 (0.11 to 0.13) |
| Between-county variance of shared (log) risk | | | |
| Non-Hispanic white | $Var\left(\phi_{ij}\delta\right)$ | 0.02 (0.01 to 0.05) | 0.01 (0.00 to 0.02) |
| Non-Hispanic black | $Var\left(\phi_{ij}/\delta\right)$ | 0.05 (0.03 to 0.11) | 0.03 (0.01 to 0.06) |
| Between-county variance of race-specific (log) risk | | | |
| Non-Hispanic white | $Var\left(\psi_{1ij}\right)$ | 0.05 (0.03 to 0.06) | 0.02 (0.00 to 0.03) |
| Non-Hispanic black | $Var\left(\psi_{2ij}\right)$ | 0.07 (0.03 to 0.11) | 0.09 (0.05 to 0.11) |
| Fraction of total geographic variation shared with the other race/ethnicity | | | |
| Non-Hispanic white | $\frac{Var(\phi_{ij}\delta)}{Var(\phi_{ij}\delta) + Var(\psi_{1ij})}$ | 30% (13% to 63%) | 52% (13% to 88%) |
| Non-Hispanic black | $\frac{Var(\phi_{ij}/\delta)}{Var(\phi_{ij}/\delta) + Var(\psi_{2ij})}$ | 42% (23% to 81%) | 15% (6% to 57%) |

SMR, standardised mortality ratio.

Risk ratio (black/white)

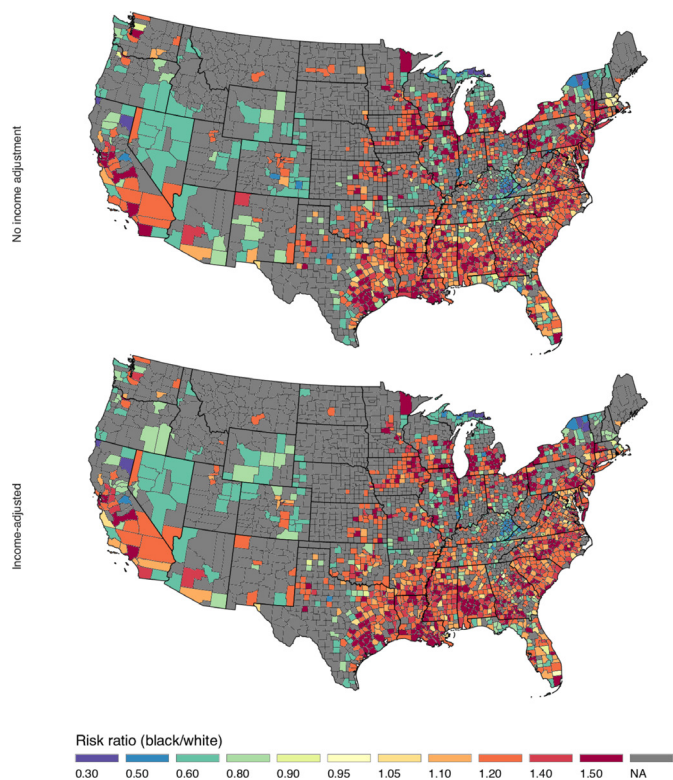0.30  0.50  0.60  0.80  0.90  0.95  1.05  1.10  1.20  1.40  1.50  NA

**Figure 3** County-level risk ratio (black/white) of premature mortality before (top) and after (bottom) adjusting for county median household income. Both models use sex/age-standardised rates. Non-significant counties are grey. Significant counties are defined as counties with greater than 80% of posterior estimates above or below 1.

## County-specific black/white risk ratio

Within-county black/white risk ratio had a minimum value of 0.33 (95% UI: 0.20 to 0.44) in both models to 5.07 (95% UI: 4.77 to 5.40) in the model without county median household income and 4.55 (95% UI: 4.24 to 4.89) in the model with income adjustment. Counties in the South, California and the Mid-Atlantic region had statistically significantly higher premature mortality risk for blacks than for whites. Conversely, counties in Appalachia, Nevada and parts of Colorado showed statistically significantly lower premature mortality risk for blacks than whites. The county-level risk ratio remained largely unchanged after adjusting for county median household income (figure 3 and online supplementary appendix figure 7), despite substantial changes in the race/ethnicity-specific and shared components. The global disparity (ie, the spatially invariant difference in risk) was only a small part of the county-level risk ratio at 1.10 (95% UI: 1.08 to 1.12) for both the income-adjusted and unadjusted models.

## DISCUSSION

Within the contiguous USA, there is considerable geographic variation in premature mortality for both whites and blacks. For reference, the fifth percentile of county premature mortality rates for both whites and

blacks is approximately 30% higher than the average premature mortality rate in high-income European counties such as Norway, the Netherlands and Luxembourg (approximately 180 vs 140 per 100,000; see online supplementary appendix table 2 for other list of premature mortality rates across various countries). Conversely, the 95th percentile county premature mortality rate for whites is similar to the average rate in Bulgaria or Latvia (approximately 330 per 100,000); for blacks, the 95th percentile rate is similar to post-Soviet nations such as Belarus, Ukraine and Lithuania (approximately 420 per 100,000; see online supplementary appendix table 2). In addition, there is large geographic variation in the county-level racial/ethnic disparity between these two groups. Both within-race geographic variation and within-county racial/ethnic disparity persist after adjusting for county median household income. This geographic variation is masked when aggregating over larger areas and cannot be directly estimated for many counties with small populations; however, the shared component model offers a valuable method of smoothing data across both geography and race/ethnicity simultaneously. This method allows for reliable estimates of premature mortality, by race/ethnicity as well as estimates of the disparities between groups. Further, the shared component model allows for premature mortality risk to be decomposed into county-level risk specific to each race/ethnicity as well as the county-level risk shared between each group.

Before interpreting our results, it is important to note the limitations of this study. First, we refrain from making any causal interpretations due to the ecological nature of the study and the lack of temporal data. This descriptive model is designed to inform policy, generate hypotheses and predict areas of risk, but is not suited for causal inference. This is especially important in regard to attempts to interpret the impact of county median household income. Our finding that county income explains more white risk than black risk is, at least in part, likely due to whites comprising a larger proportion of the population and thus county median household income is more correlated with white income. Finally, it is possible that the composition of people within counties changed substantially over the 6-year period. However, despite these limitations, we believe this study demonstrates the value of the shared component model in health disparities research.

Consistent with the literature, our results suggest that whites $\psi_{1ij}$ experience excess premature mortality in Appalachia,[30–32] while blacks $\psi_{2ij}$ experience excess premature mortality in the South.[1 5 33] The race/ethnicity-specific components reflect spatially patterned risk factors that are not shared with the other group. Thus, for whites, this heightened risk in Appalachia likely reflects noted issues of low income and education, geographic isolation, reduced access to care and environmental factors.[34] Similarly, the black-specific component in the South is consistent with the research about the lasting impact of slavery and racism, differences in opportunity structure,

black-specific experience of county-level poverty and socioeconomic conditions, and differential access to care.[13 35–37] In addition, the shared component describes areas that are at higher or lower risk for both blacks and whites, simultaneously. Even after adjusting for county median household income, much of the South has increased premature mortality risk for both whites and blacks while the Midwest has decreased risk for both.

In addition, our results suggest that county median household income explains more of the county-level white risk than black risk. Furthermore, we find that half of the geographic variation in white risk is shared with blacks, while only 15% of the geographic variation in the black risk is shared with whites. That is, after adjusting for county median household income, the majority of geographic variation in premature mortality risk for blacks is not shared with whites. Further research should be conducted to understand sources of county-level variability in black risk. Finally, we found between-state variation to be small relative to between-county variation in both models, again reiterating the importance of estimating geographic variation at the local level.

There is wide geographic variation in the racial/ethnic disparity of premature mortality risk. Specifically, the county-level black/white risk ratio estimates ranged from 0.30 (95% UI: 0.20 to 0.44) to 4.56 (95% UI: 4.24 to 4.89) in the income-adjusted model. However, despite substantial changes in the shared and race/ethnicity-specific surfaces after adjusting for county median household income, we find almost no change in the spatial pattern of the black/white risk ratios of premature mortality before and after adjustment. Other aspects of structural racism, such as racial residential and occupational segregation, could plausibly contribute to the inequalities not accounted for solely by median household income.[12]

This study is the first to jointly model premature mortality risk in non-Hispanic blacks and non-Hispanic whites. The joint modelling approach identified counties with higher or lower risk unique to blacks or whites as well as counties with shared risk, despite small counts of the black population and deaths in many counties. This decomposition of risk, in addition to more precise estimates in small populations, suggests joint spatial modelling in general, and the shared component model specifically, may be useful tools for researchers to measure the impact of interventions, inform policy and generate new hypotheses in studying in health disparities across geography and sociodemographic characteristics. This nuanced decomposition of risk may be a powerful tool as researchers grapple with disentangling neighbourhood contextual effects and sociodemographic compositional effects of an area when evaluating health outcomes.

**ORCID iD**
Mathew V Kiang http://orcid.org/0000-0001-9198-150X

## REFERENCES

1. Murray CJ, Kulkarni SC, Michaud C, *et al*. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med* 2006;3:e260.
2. Krieger N, Chen JT, Waterman PD, *et al*. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the public health disparities geocoding project. *Am J Public Health* 2005;95:312–23.
3. Subramanian SV. Income inequality and health: what have we learned so far? *Epidemiol Rev* 2004;26:78–91.
4. Kochanek KD, Murphy S, Xu J, *et al*. Mortality in the United States, 2016. *NCHS Data Brief* 2017:1–8.
5. Cullen MR, Cummins C, Fuchs VR. Geographic and racial variation in premature mortality in the U.S.: analyzing the disparities. *PLoS One* 2012;7:e32930.
6. Rossen LM, Khan D, Schoendorf KC. Mapping geographic variation in infant mortality and related Black-White disparities in the US. *Epidemiology* 2016;27:690–6.
7. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A* 2001;164:73–85.
8. Dabney AR, Wakefield JC. Issues in the mapping of two diseases. *Stat Methods Med Res* 2005;14:83–112.
9. Earnest A, Beard JR, Morgan G, *et al*. Small area estimation of sparse disease counts using shared component models-application to birth defect registry data in New South Wales, Australia. *Health Place* 2010;16:684–93.
10. Ibáñez-Beroiz B, Librero-López J, Peiró-Moreno S, *et al*. Shared component modelling as an alternative to assess geographical variations in medical practice: gender inequalities in hospital admissions for chronic diseases. *BMC Med Res Methodol* 2011;11:172.
11. Wheeler DC, Waller LA, Elliott JO. Modeling epilepsy disparities among ethnic groups in Philadelphia, PA. *Stat Med* 2008;27:4069–85.

12  Williams DR, Sternthal M. Understanding racial-ethnic disparities in health: sociological contributions. *J Health Soc Behav* 2010;51 Suppl:S15–27.

13  Krieger N, Rehkopf DH, Chen JT, *et al*. The fall and rise of US inequities in premature mortality: 1960-2002. *PLoS Med* 2008;5:e46–41.

14  Cossman RE, Cossman JS, Jackson R, *et al*. Mapping high or low mortality places across time in the United States: a research note on a health visualization and analysis project. *Health Place* 2003;9:361–9.

15  Norheim OF, Jha P, Admasu K, *et al*. Avoiding 40% of the premature deaths in each country, 2010-30: review of national mortality trends to help quantify the UN sustainable development goal for health. *Lancet* 2015;385:239–52.

16  García MC, Bastian B, Rossen LM, *et al*. Potentially preventable deaths among the five leading causes of death - United States, 2010 and 2014. *MMWR Morb Mortal Wkly Rep* 2016;65:1245–55.

17  US Department of Health and Human Services. Foundation health measures: disparities. Available: https://www.healthypeople.gov/2020/about/foundation-health-measures/Disparities [Accessed 1 Jan 2017].

18  World Health Organization. The world health report 1998: life in the 21st century a vision for all. world heal organ, 1998. Available: http://www.who.int/whr/1998/media_centre/50facts/en/#%5Cnhttp://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=PAHO&lang=p&nextAction=lnk&exprSearch=25006&indexSearch=ID

19  National Center for Health Statistics. *Compressed mortality file, 1995-2015 (machine readable data file and documentation, CD-ROM series 20, no. 2u) as compiled from data provided by the 57 vital statistics jurisdictions through the*. Hyattsville, Maryland: Vital Statistics Cooperative Program, 2016. https://www.cdc.gov/nchs/data_access/cmf.htm

20  Krieger N, Kosheleva A, Waterman PD, *et al*. 50-year trends in US socioeconomic inequalities in health: US-born black and white Americans, 1959-2008. *Int J Epidemiol* 2014;43:1294–313.

21  United States Census Bureau. *B19013: median household income in the past 12 months (in 2015 inflation-adjusted dollars)*, 2016.

22  United States Census Bureau. County adjacency file, 2010. Available: https://www.census.gov/geo/reference/county-adjacency.html

23  Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 1991;43:1–20.

24  Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and draper). *Bayesian Anal* 2006;1:515–34.

25  R Core Team. *R: a language and environment for statistical computing*, 2017.

26  Stan Development Team. RStan: R package version 2.14.1. Available: http://mc-stan.org/ [Accessed 1 Jan 2016].

27  Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statist Sci* 1992;7:457–72.

28  Best N, Hansell AL. Geographic variations in risk: adjusting for unmeasured confounders through joint modeling of multiple diseases. *Epidemiology* 2009;20:400–10.

29  Richardson S, Thomson A, Best N, *et al*. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 2004;112:1016–25.

30  Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, *et al*. US county-level trends in mortality rates for major causes of death, 1980-2014. *JAMA* 2016;316.

31  Case A, Deaton A. Mortality and morbidity in the 21st century. *Brookings Pap Econ Act* 2017;2017:397–476.

32  Singh GK, Kogan MD, Slifkin RT. Widening disparities in infant mortality and life expectancy between Appalachia and the rest of the United States, 1990–2013. *Health Aff* 2017;36:1423–32.

33  Geronimus AT, Bound J, Waidmann TA, *et al*. Excess mortality among blacks and whites in the United States. *N Engl J Med* 1996;335:1552–8.

34  Behringer B, Friedell GH. Appalachia: where place matters in health. *Prev Chronic Dis* 2006;3:A113.

35  Galster G, Sharkey P. Spatial foundations of inequality: a conceptual model and empirical overview. *RSF Russell Sage Found J Soc Sci* 2017;3:1–33.

36  O'Connell HA. The impact of slavery on racial inequality in poverty in the contemporary U.S. South. *Social Forces* 2012;90:713–34.

37  Link BG, Phelan JC, Miech R, *et al*. The resources that matter: fundamental social causes of health disparities and the challenge of intelligence. *J Health Soc Behav* 2008;49:72–91.