The role of changing socioeconomic conditions, healthcare, and health behaviors in US county-level mortality trends, 2000-2015

Nick Graetz, University of Pennsylvania
Irma T. Elo, University of Pennsylvania

Short Abstract

Many studies have documented significant divergence in U.S. county-level mortality trends since 2000. Recent analyses have effectively utilized spatially-explicit Bayesian hierarchical models to make robust estimates of county-level mortality over space and time. However, few studies have examined a comprehensive set of time-varying contextual covariates within such a modelling framework to illustrate how large differences in mortality trajectories by county and region are associated with shifting levels of social and economic context, health care access, behaviors and population composition. Combing vital statistics data with county-level characteristics related to healthcare, health behaviors, socioeconomic profile and population composition, this paper utilizes a spatially-explicit Bayesian hierarchical modelling framework to analyze how changing levels of mortality across age groups are associated with changes in county-levels exposures. Additionally, we employ a Shapley decomposition on the time-varying components of our models to illustrate the additive contributions of each changing characteristic to the observed mortality change in each U.S. county since 2000.

Extended Abstract

Geography has long been recognized as an important factor influencing an individual's exposures to health-related risks and economic opportunities over their life course (Krieger et al. 2005). Historically, mortality and geography have been intertwined across the United States and large geographic inequalities have been documented (Chetty et al. 2016; Dwyer-Lindgren et al. 2016; Murray et al. 2006; Wang et al. 2013). Many studies have focused on gaps in health and life expectancy across different geographic regions or groups of counties, with one 2006 study finding a gap in life expectancy of over 35 years between race-county combinations (Ezzati et al. 2008; Murray et al. 2006; Wang et al. 2013). In describing the significant heterogeneity across county-level mortality rates, several studies have importantly emphasized an initial fall and then increase in heterogeneity between 1960-2000 (Ezzati et al. 2008; Krieger et al. 2008). County-level heterogeneity in mortality levels and progress necessitates exploration of the associated county-level characteristics. Much of this research focuses on distributive justice concerns with a particular emphasis on the concentration of poverty and income inequality and racial/ethnic minorities. Public health research utilizing newly available datasets has suggested that social and environmental risk is also becoming increasingly concentrated in rural counties which may be contributing to the U.S. stagnation on various health and mortality indices (Lichter et al. 2012; United States Environmental Protection Agency 2014).

While there has been significant research on geographic mortality differentials in the United States by states and county-regions, including evidence of large gains in life expectancy across central metropolitan areas, many of these studies examine mortality trends across wide geographies which may mask mortality heterogeneity across counties. This further risks obscuring the drivers of stagnating or deteriorating life expectancy by race, in particular across the rural regions of the United States (Case and Deaton 2015). In addition, many studies examining the relationships between social exposures and mortality at the county-level ignore spatial autocorrelation in county mortality rates, which can bias the associations between exposures and health and mortality (Jackson et al. 2000; Nuru-Jeter and LaVeist 2011).

Spatial analysis has become more common in predictive modelling at the county-level. In a comprehensive analysis of county-level mortality trends, Dwyer-Lindgren et al. use a spatially explicit Bayesian small-area model to estimate mortality rates from 1980 to 2014 (Dwyer-Lindgren et al. 2017). They additionally examine associations with key hypothesized drivers of mortality change, but only do so using a cross-section of estimated mortality rates (or life expectancy) in 2009 in an ordinary-least-squares framework. While such an approach can suggest which county-level characteristics are likely important in explaining county-level variation in mortality at a point in time, it cannot describe how the changes in these characteristics relate longitudinally to shifting mortality trajectories across U.S. counties. In addition to estimated relationships, many contemporaneous indicators may be very correlated with contemporaneous mortality rates, but these indicators may be virtually unchanging since 2000 and thus cannot explain the large divergences in county-level mortality improvement.

This project aims to describe differential trends in age-standardized crude death rates (ASCDRs) by sex for key age groups (0-24, 25-64, 65+) across all US counties from 2000-2015 and examine the degree to which shifting profiles of socioeconomic characteristics, healthcare, health behaviors, and population composition may be driving heterogeneous mortality trends. We employ a spatially-explicit Bayesian hierarchical model to estimate the associations between mortality rates and a diverse set of contextual county-level covariates since 2000. As the levels and changes in these covariates themselves are often distributed very unequally across counties, we also employ a decomposition of mortality change at the county-level based on our estimated models to examine how the consequences of these changes in characteristics and their associations with county-level mortality vary across the United States.

Data

We use vital statistics data on deaths and population estimates from the National Center for Health Statistics (NCHS) to estimate age-specific (five-year age groups), all-cause mortality rates by sex for all counties (CDC 2018) for 1999-2001, 2009-2011, and 2014-2016 to obtain more robust estimates than if a single years of data were used. Mortality rates were age-standardized using the 2000 Census populations to ensure that differential mortality rates are not simply due to differences in changing population age structures. In order to account for important mortality differences across the urban-rural continuum within the United States and to capture omitted variables that may be structured around these dimensions, we additionally categorize counties by metropolitan-nonmetropolitan status and region according to the county-level typologies from the United States Department of Agriculture's (USDA) Economic Research Service (ERS) (USDA ESR 2015).

We assembled a database of time-varying contextual covariates at the county-year-level from 2000-2015 to assess how trends in mortality rates are associated with changes in factors related to the county's socioeconomic profile, healthcare availability, health behaviors, and population composition. Table 1 reports the source information for each indicator as well as the temporal coverage.

Methods

We first calculate a global Moran's I for each dataset of county-level, all-cause, age-standardized mortality rates. The Moran's I statistic is a common measure of spatial autocorrelation, i.e. the strength of the correlation between observations nearer in space compared to those that are further apart (Li et al. 2007). In this formula,

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N is the number of spatial units indexed by *i* and *j*, where *i* is each specific county, *j* is every other county, *x* is the observed county-level mortality rate, $\bar{x}$ is the mean mortality rate across all counties, *w* is a matrix of spatial weights, and *W* is the sum of all spatial weights. This statistic is dependent on assumptions about the structure of the spatial weight matrix. For this analysis, we use a *Queens nearest neighbor* matrix which defines "neighbors" as those counties sharing a boundary. Additionally, we calculate a test examining local indicators of spatial autocorrelation (LISA) over the entire dataset using the same spatial weights matrix (Anselin 2010). This approach decomposes the global Moran's I statistic to local pockets of spatial autocorrelation, and tests whether they are significantly different than what could be expected given the same observed data randomly distributed across space (weighting with county populations). This exploratory test can be useful in locating clustered observations with significant leverage on the global spatial autocorrelation, as well as identifying local pockets of potential spatial nonstationarity.

To examine the associations between mortality and our county-level predictors, we fit Bayesian generalized linear models assuming a binomial likelihood and logit link-function to estimate age-specific death rates ($m_{\mathrm{i,y,a}}$) in each county (*i*), year (*y*), and age (*a*). These models were specified and compared with an increasing number of county-year covariates ($X_{i,y}$), with the final model including a latent spatial error structure (γ_i) following the Besag model. The spatial weights matrix was constructed using a nearest-neighbors approach following the Queens convention. All models were fit in a Bayesian framework with uninformative priors. The posterior distributions were fit using computationally efficient and accurate approximations in R-INLA (integrated nested Laplace approximation) (Rue et al. 2014). The spatial model specification is below, while all others (X_i = metro categories, metro + region categories, metro + region + predictors) included no latent error structure. All models contained year (2000, 2010, 2015) and five-year age groups as dummy variables. All county-year covariates ($X_{i,y}$) were interacted with three broad age groups (0-24, 25-64, 65+).

$$D_{i,y,a} | m_{\mathrm{i,y,a}}, N_{\mathrm{i,y,a}} \sim Binomial(m_{\mathrm{i,y,a}}, N_{\mathrm{i,y,a}})$$

$$logit(m_{\mathrm{i,y,a}}) = \propto\ + \beta_1 X_{i,y} + \beta_2 Age + \beta_3 Year + \gamma_i + \varepsilon_{\mathrm{i,y,a}}$$

$$\gamma_i \sim Besag(0, \tau)$$

$$\beta \sim Normal(0{,}1000)$$

$$\tau \sim Gamma(1{,}10)$$

All models were estimated separately by sex. We calculated several statistics to compare model performance, examining both in-sample fit and the degree of spatial autocorrelation in the county-level residuals. A high degree of spatial autocorrelation in the residuals would indicate that our model does not fully capture the spatial autocorrelation in our observed data conditional on our predictors, which can result in biased estimates for the coefficients given the assumption of conditional *iid* errors does not hold (Anselin 2010). We calculate the Moran's I for all model residuals using the same nearest-neighbors spatial weight matrices used in fitting the spatial model. We also calculate two fit statistics to compare model performance. The deviance information criterion (DIC) is a hierarchical modeling generalization of the more commonly used Bayesian information criterion (BIC). As with BIC, a smaller DIC is preferable in comparison and penalizes models for fit as well as the number of effective parameters. General model in-sample error is compared using the root-mean-squared-error (RMSE), weighting errors by county populations.

We use Shapley decomposition to quantify the contributions of each time-varying contextual covariate, the secular trend, and unobserved variation to changes in observed mortality rates within counties (Fortin et al. 2010; Madden 2012; Wang et al. 2014). Shapley decomposition is a method originating within game theory that allows for decomposition of changes in a variable attributable to changes in contributory factors. Specifically, to assess the effect of each time-varying factor in our model on mortality rates between 2000-2015, we constructed the universe of scenarios in which all factors took on values from either 2000 or 2015 in each specific scenario. To compute the effect of any one factor, we assessed each pair of scenarios in which that factor changed but all other factors maintained the same values. The average of the changes across all pairs of scenarios was the contribution of that factor to change in observed mortality. We repeated the same process for all contextual covariates, the secular trend (year), and the residual between predicted and observed mortality rates.

Preliminary findings

For the purpose of this abstract, all Figures are reported only for the female age-standardized death rates (ages 25-64). Table 2 presents the estimated coefficients from a series of models for females, as well as the fit statistics and Global Moran's I calculated on the model residuals. This table includes coefficients from the county predictors interacted with the 25-64 age group to explore correlates of middle-age mortality trends. The combination of metropolitan category and region explain some, but not all, of the spatial autocorrelation in the observed county-level mortality rates. There is a significant mortality advantage for the large central and fringe metro counties compared to the medium/small metro and non-metro counties. However, conditional on several contextual covariates this advantage disappears. In the full spatial model, poverty and reliance on transfers are significantly associated with higher mortality rates, while the proportion of the county population that is foreign-born and the proportion college-educated are associated with lower mortality.

However, the inference we can make regarding change based solely on these coefficients is limited. They are useful in illustrating the conditional relationships of change between covariates and mortality across our entire dataset, but the levels of these covariates and how they are changing over time is very heterogeneous across counties within the United States. Figures 2 and 3 illustrate the additive contribution at the county-level of the proportion foreign-born and the proportion of personal income as transfers, respectively, calculated from the Shapley decomposition of observed mortality change between 2000-2015. Figure 2 demonstrates how despite the significant negative relationship between the proportion foreign-born and mortality across all counties, suggesting that higher the share of the foreign-born the lower the morality at ages 25-64. It is also clear that the magnitude of this association in explaining observed mortality change is very geographically confined to large metropolitan areas on the coasts and several cities. These counties, such as those containing Chicago, Atlanta, and New York City, are where dramatic increases in the proportion foreign-born have been observed since 2000. In terms of geography, this aspect of population composition has remained virtually unchanged across most of the United States. Figure 3 shows how the rising proportion of personal income as government transfers has contributed positively to mortality change, i.e., increasing mortality with increasing proportion of personal income coming from government transfers, almost everywhere, although this contribution is particularly dramatic across the South and Appalachia.

Figure 4 summarizes the geographic heterogeneity in these additive contributions by aggregating to observed mortality change for each region-metro-nonmetro group of counties. Here we can observe geographic variation in the level and direction of total observed mortality change, as well as how our model decomposes those additive contributions. As observed in prior work, female mortality rates in the 25-64 age group have generally improved in almost all large metro areas and across the coastal and Mountain regions. However, these rates have increased dramatically across counties in Appalachia and the South Central region. An increasing reliance on personal transfers is a large positive contributor to this change, as well as increased absolute poverty.

Next Steps

We have illustrated our analytic strategy using only one age-group and gender (25-64, female). We will estimate similar models for both men and women across all age groups. In addition, we will examine the richer set of county-level characteristics shown in Table 2. Their inclusion will permit the analyses of the role of changing socioeconomic characteristics, health care resources, health behaviors and population composition to the diverging mortality trends that have been observed in the last decades in the United States. They will shed light on which contextual-level factors have been associated with widening mortality disparities by gender, metro-nonmetro status and region of the country. The paper will end with a discussion of the results and their implications for future analyses and public policies.

Table 1. County-level characteristics by source and year coverage. Several indicators have to be interpolated to our modelling range (2000-2015) by using linear projection with an annualized growth rate forward to 2015 or backward to 2000.

	Source	Years	Interpolated
Socioeconomic			
Population with a college degree, %	Census/ACS	2000, 2010, 2015	
Population below poverty line, %	SAIPE	2000, 2010, 2015	
Median household income, log dollars	SAIPE	2000, 2010, 2015	
Transfers, as % of total personal income	BEA	2000, 2010, 2015	
Unemployment, %	BEA	2000, 2010, 2015	
Employment in manufacturing sector, %	Census/ACS	2000, 2010, 2015	
Eviction rate, %	Princeton Eviction Lab	2000-2016	
Healthcare			
Active MDs per 1000	Area Health Resource File	2000, 2010, 2015	
Medicare expansion (dummy)	Advisory Board	–	
Population insured, % (aged <65)	SAIPE	2000, 2010, 2015	
Healthcare quality index	Dartmouth Atlas Project	2006-2014	X
Diabetic monitoring (% of Medicare enrollees, 65-75)	Dartmouth Atlas Project	2006-2014	X
Mammography screening (% of female Medicare enrollees, 67-69)	Dartmouth Atlas Project	2006-2014	X
Health behaviors and risk factors			
Smoking (cigarette smoking prevalence in adults aged 20 and older)	IHME time series (BRFSS)	1996-2012	X
Obesity (% of adults aged 20 and older that report a BMI of 30 or more)	Wisconsin CHR (BRFSS)	2004-2013	X
Alcohol use (% any drinking, heavy drinking, binge drinking)	IHME time series (BRFSS)	2002-2012	X
Physical activity (% of adults aged 20 and older reporting no lesiure-time physical activity)	Wisconsin CHR (BRFSS)	2004-2013	X
Hypertension prevalence (adults age 30 and older)	IHME time series (NHANES, BRFSS)	2001-2009	X
Diabetes prevalence (Adults aged 20 and older)	IHME time series (NHANES, BRFSS)	1999-2012	X
Population composition			
Net-migration (indirect estimate using variable-r)	Based on vital statistics	2000, 2010, 2015	
Foreign-born population, %	Census/ACS	2000, 2010, 2015	
Support ratio	Census/ACS	2000, 2010, 2015	
Black population, %	Census/ACS	2000, 2010, 2015	
American Indian, Native Alaskan, and Native Hawaiian population, %	Census/ACS	2000, 2010, 2015	
Hispanic population, %	Census/ACS	2000, 2010, 2015	

Table 2. Estimated coefficients from Bayesian generalized linear models, women ages 25-64, United States, 2000, 2010 and 2015.

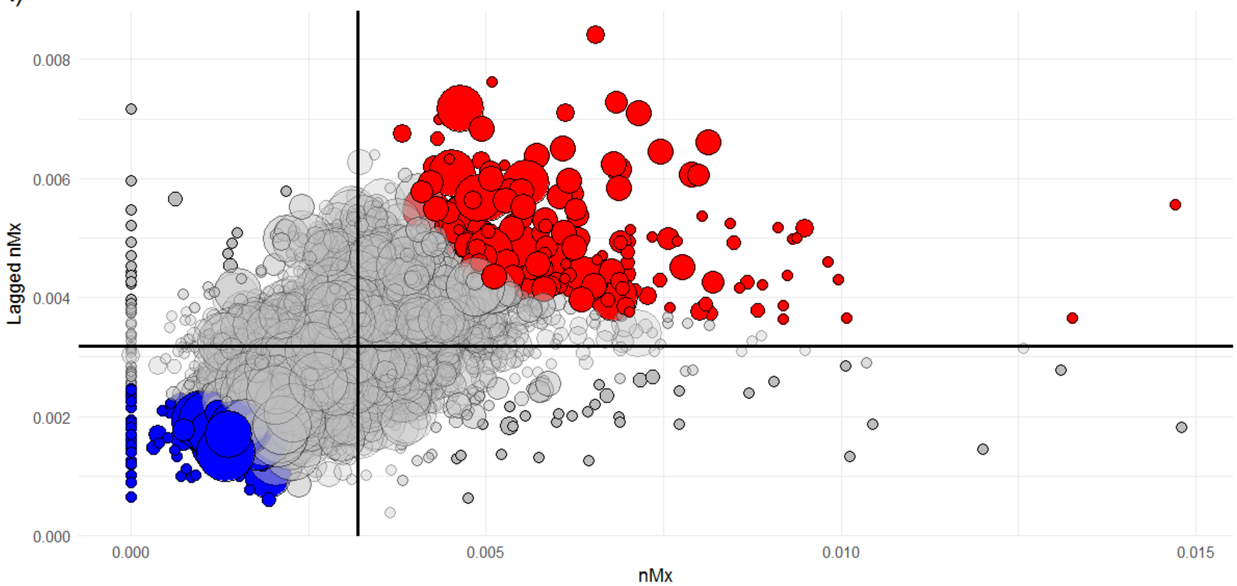
	Female			
	Model 1	Model 2	Model 3	Model 4
Year	1.00*	1.00*	1.00*	1.00*
Metro category				
Lg central metro	–	–	–	–
Lg fringe metro	0.95*	0.96*	0.99	0.87*
Md/Sm metro	1.12*	1.10*	0.95*	0.89*
Nonmetro	1.26*	1.21*	0.89*	0.85*
Region				
Pacific	–	–	–	–
Appalachia	–	1.22*	1.04*	1.07
East South Central	–	1.27*	1.11*	1.11
Mountain	–	1.01*	1.01*	1.11*
West South Central	–	1.21*	1.08*	1.09
New England	–	0.88*	0.97*	1.05
South Atlantic	–	1.10*	1.05*	1.09
East North Central	–	1.02*	0.98*	1.05
West North Central	–	0.93*	0.94*	1.07
Middle Atlantic	–	0.93*	0.95*	1.01
Social exposures				
Transfers	–	–	1.01*	1.01*
Unemployment	–	–	0.98*	0.98*
Poverty	–	–	1.02*	1.01*
Foreign-born	–	–	0.99*	0.98*
College	–	–	0.99*	0.99*
Working age	–	–	1.01*	1.00*
Global Moran's I	0.26	0.15	0.10	0.00
DIC	109406	100695	78846	67979
RMSE	0.00075	0.00069	0.00054	0.00039

[a] * indicate parameters with 95% of the posterior density above or below 1.

Model 1 includes metropolitan status, Model 2 adds regional indicators, Model 3 adds county-level predictors, and Model 4 adds a latent spatial random effects structure following the Besag distribution. The global Moran's I is calculated on the residuals from each model. The DIC and RMSE are reported to compare in-sample fit.

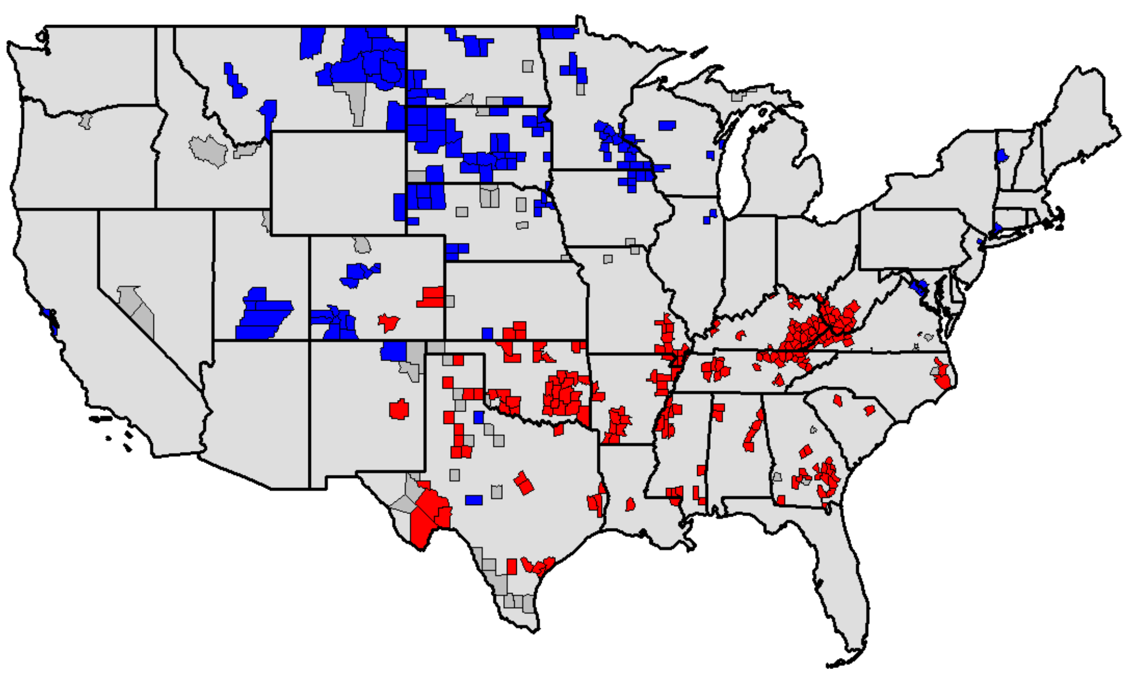
Figure 1. Global and LISA tests for spatial autocorrelation across female age-standardized mortality rates at ages 25-64, 2015.

A)



B)

Local Morans I Test (LISA)
Global Morans I: 0.27, p-value = 0



Panel A shows each county's mortality rate compared to the rates of surrounding counties defined by the Queens matrix. Dot size corresponds to county population. Red and blue indicate a significant LISA test for positive spatial autocorrelation (p<0.05), where red indicates high-mortality counties surrounded by similarly high-mortality counties and blue indicates low-mortality counties surrounded by similarly low-mortality counties. Dark grey indicates significant negative autocorrelation in the LISA test (either high-mortality surrounded by low-mortality of vice versa). Panel B visualizes the location of the counties surrounded by similarly high (red) or low (blue) mortality rates across the United States

Figure 2. The additive contribution of changes in the proportion of the population that is foreign-born to the changes in observed age-standardized (25-64, female) mortality rates (per 100,000).

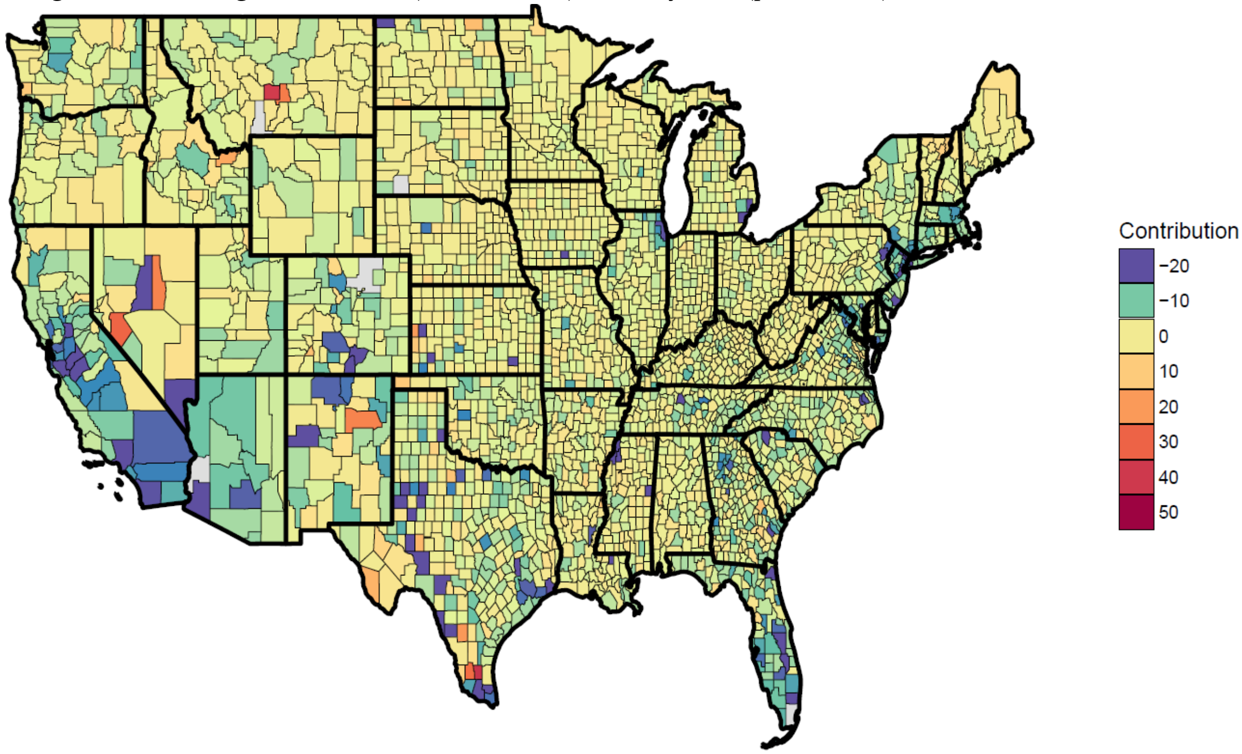


Figure 3. The additive contribution of changes in the proportion of personal income as government transfers to the changes in observed age-standardized (25-64, female) mortality rates (per 100,000).

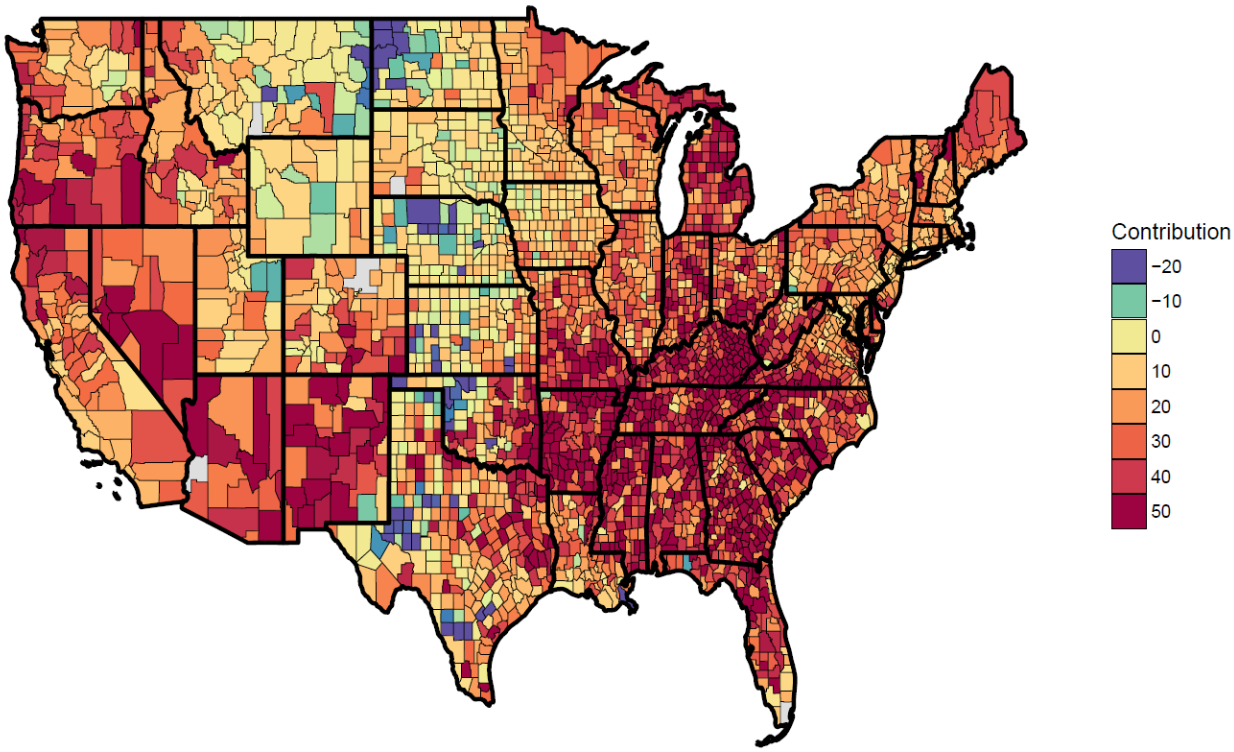
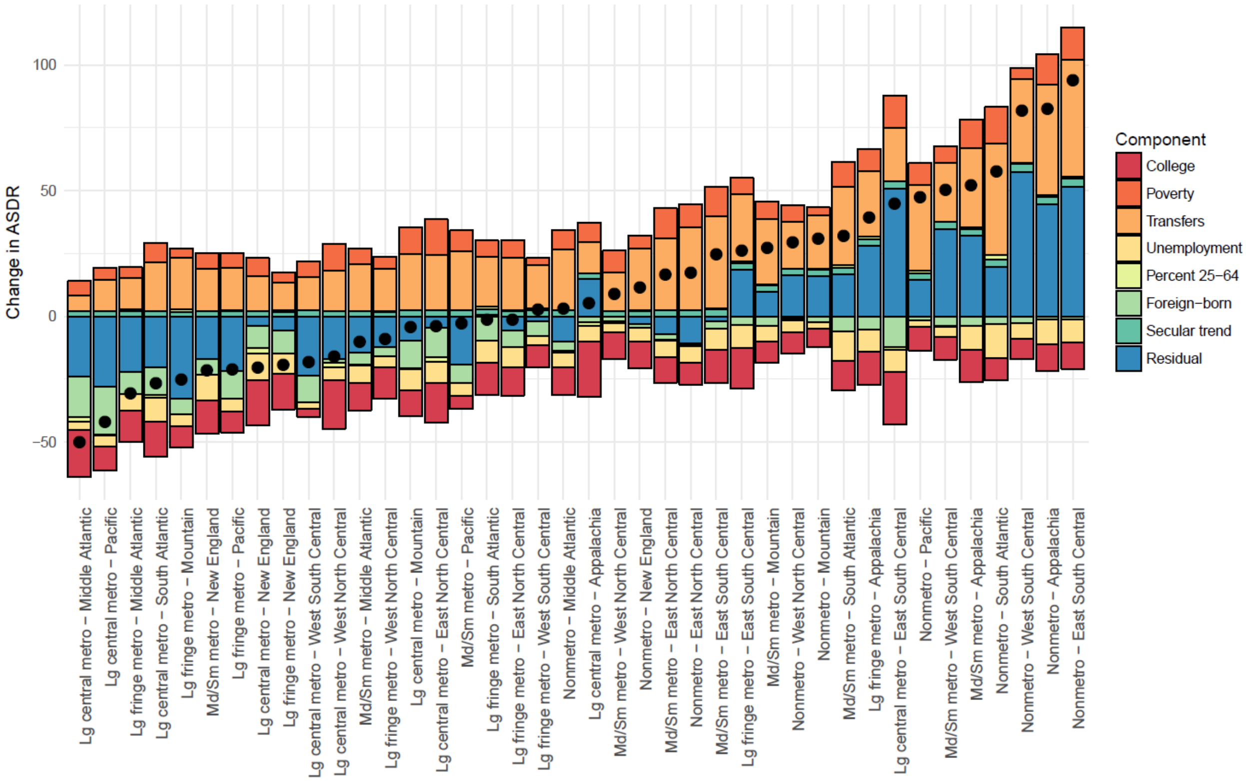


Figure 4. Changes in observed female age-standardized mortality rates at ages 25-64 between 2000 and 2015 by metro-nonmetro status and region.



The black dots represent the change in the ASCDR (per 100,000). The additive contributions of each factor from the Shapley decomposition to this observed mortality change are plotted as colored bars. The sum of all bars within a metro-region is equal to the observed mortality change.