Annals of Epidemiology 25 (2015) 329–335

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Comparing methods of measuring geographic patterns in temporal trends: an application to county-level heart disease mortality in the United States, 1973 to 2010

Adam S. Vaughan MPH, MS [a,b,*], Michael R. Kramer PhD [a,b], Lance A. Waller PhD [c], Linda J. Schieb MSPH [b], Sophia Greer MPH [b], Michele Casper PhD [b]

[a] Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA
[b] Division for Heart Disease and Stroke Prevention, Centers for Disease Control and Prevention, Atlanta, GA
[c] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA

ARTICLE INFO

Article history:
Received 30 October 2014
Accepted 16 February 2015
Available online 19 February 2015

Keywords:
Heart disease mortality
Spatiotemporal trends
Geographic patterns
Bayesian methods
Joinpoint regression
Generalized linear mixed models
Spatial analysis

ABSTRACT

Purpose: To demonstrate the implications of choosing analytical methods for quantifying spatiotemporal trends, we compare the assumptions, implementation, and outcomes of popular methods using county-level heart disease mortality in the United States between 1973 and 2010.
Methods: We applied four regression-based approaches (joinpoint regression, both aspatial and spatial generalized linear mixed models, and Bayesian space-time model) and compared resulting inferences for geographic patterns of local estimates of annual percent change and associated uncertainty.
Results: The average local percent change in heart disease mortality from each method was −4.5%, with the Bayesian model having the smallest range of values. The associated uncertainty in percent change differed markedly across the methods, with the Bayesian space-time model producing the narrowest range of variance (0.0–0.8). The geographic pattern of percent change was consistent across methods with smaller declines in the South Central United States and larger declines in the Northeast and Midwest. However, the geographic patterns of uncertainty differed markedly between methods.
Conclusions: The similarity of results, including geographic patterns, for magnitude of percent change across these methods validates the underlying spatial pattern of declines in heart disease mortality. However, marked differences in degree of uncertainty indicate that Bayesian modeling offers substantially more precise estimates.

Introduction

Analytical and computing advances have greatly increased the numbers of methods and tools available to quantify spatiotemporal trends of disease. Although these methods vary in their underlying assumptions, methodological and computational complexity, data requirements, and interpretability, a paucity of literature compares these methods to provide practical guidance to epidemiologists and public health practitioners.

Regression-based approaches commonly used to quantify spatiotemporal trends in local rates of chronic disease include joinpoint regression, generalized linear mixed models (GLMM), and Bayesian space-time models. Each approach quantifies temporal trends by estimating valid, succinct, and interpretable summaries of changing rates using software and methods that are accessible to many in public health. Comparing these methods' results within the context of their respective underlying assumptions provides important information for appropriate method selection in spatiotemporal studies of health outcomes.

We illustrate these methods using county-level heart disease death rates in the United States. Between 1950 and the turn of the 21st century, U.S. heart disease mortality decreased by roughly 60% [1,2]. Although studies have reported differential trends in heart disease mortality by larger geographic areas and urbanicity [3–10], few have used rigorous methods currently available for small-area trend estimation. In this study, we describe the key features of four regression-based spatiotemporal methods (joinpoint regression, aspatial GLMM, spatial GLMM, and Bayesian space-time models) and compare the resulting estimates of county-level percent change in heart disease death rates from 1973 to 2010.

* Corresponding author. Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322. Tel.: +1-505-301-8712; fax: 404-727-8661.
E-mail address: asvaugh@emory.edu (A.S. Vaughan).

http://dx.doi.org/10.1016/j.annepidem.2015.02.007
1047-2797/

Methods

Data sources

Annual age-specific counts of heart disease deaths in each U.S. county from 1973 to 2010 for people aged 35 years or more were obtained from the National Vital Statistics System. Over this continuous time period, a census, rather than a sample, of deaths in the United States was recorded. Heart disease deaths were defined based on underlying cause of death according to the following *International Classification of Diseases* (ICD) codes: ICD-8: 390–398, 402, 404, 410–429; ICD-9: 390–398, 402, 404–429; ICD-10: I00–I09, I11, I13, I20–I51. Comparability ratios between each ICD revision are approximately unity, indicating that temporal changes in ICD codes introduced little to no bias into the study and no adjustments were necessary [11,12]. County-level annual estimates of the population aged 35 years or more, produced by the U.S. Census Bureau in collaboration with the National Center for Health Statistics, were compiled [13]. To enhance rate stability, all data were aggregated into two-year intervals starting in 1973, resulting in 19 biennial intervals. Aggregated to a common set of 3099 counties in the contiguous United States, rates were age standardized using the 2000 U.S. standard population. The biennial age-standardized rate was multiplied by the biennial county population aged 35 years or more to produce a biennial age-standardized death count [14,15].

Estimation of temporal trends

County-level average biennial percent change in heart disease mortality (hereafter referred to as "percent change") was estimated using four model-based methods: joinpoint regression, aspatial GLMM, spatial GLMM, and Bayesian space-time models. For each, we estimate percent change in heart disease death rates and its associated variance, where the variance represents uncertainty or precision in the estimated percent change. Appendix A contains additional details.

Joinpoint regression

Joinpoint regression, popularly used in examining temporal changes in cancer incidence, models rates of disease as piece-wise log-linear functions of time [16–19]. For each areal unit (e.g., county), this method finds inflection points (or joinpoints) representing the time at which the slope changes. As temporal trends may be represented by a single line or a series of linked segments, the trend across the entire study period is not required to be log linear.

The slopes of the segments adjacent to the joinpoint quantify change in the rate over time and define an annual percent change. Annual percent change of multiple segments can be summarized as average annual percent change (AAPC), or for our two-year pooled data, as average biennial percent change [20].

Using joinpoint regression, county-level age-standardized case counts were independently modeled as log-linear piece-wise functions of time with Poisson variance, log-population offset, uncorrelated errors, and a maximum of five joinpoints (Equation A.1) in Joinpoint Regression Program, version 4.0.4 (National Cancer Institute, Calverton, MD). Model fit and the numbers of joinpoints were assessed with the modified Bayesian information criterion.

Aspatial and spatial GLMMs

GLMMs quantify global and local temporal trends in disease by modeling counts and population at risk in a Poisson model or rates in a linear or log-linear model [21]. These models can be readily fit using standard statistical software packages (e.g., SAS, Stata). The assumption of log linearity in this method should be evaluated, with alternative specifications considered as needed.

For the GLMM analysis, Poisson regression modeled age-standardized county-level heart disease death rates as log-linear functions of a global intercept and temporal slope, and county-level random intercept and slope (Equation A.2) using PROC GLIMMIX in SAS, version 9.3 (SAS Institute, Cary, NC). Both aspatial GLMM (assuming spatial independence) and spatial GLMM (assuming distance-based spatially correlated rates within each two-year interval) were completed.

Bayesian space-time models

Bayesian space-time models are hierarchical mixed models where area-level rates are spatially correlated within and across time. Widely applied in small-area analysis and disease mapping applications, Bayesian models produce locally interpretable, statistically stable estimates while minimizing concerns for multiple testing [14,22–24].

The Bayesian space-time approach modeled age-standardized rates of heart disease mortality using Poisson regression (Equation A.5). Given our large number of counties and to ensure model convergence, spatially independent random effects were not used [25]. County-specific random intercepts and slopes were assigned a conditionally autoregressive normal prior which borrows information and statistical strength from adjacent counties using queen contiguity. Variance hyperpriors were assigned a uniform distribution between 0 and 1 [26]. Models were implemented in WinBUGS, version 1.4.3 (Imperial College and Medical Research Council, Cambridge, UK) using R package R2WinBUGS [27]. Each model was run with two chains for 30,000 iterations with thinning and the first half discarded. Chain convergence was evaluated through examination of trace plots of posterior parameter estimates and the Brooks-Gelman-Rubin statistic [28].

Calculating county-level percent change and its variance

The joinpoint software directly estimates biennial percent change (as AAPC) and its variance. For GLMM, the percent change and variance are calculated as a function of parameter estimates. For the Bayesian model, percent change is calculated as a function of parameter estimates, and its variance is estimated using the distribution of the posterior. See Appendix A for details.

Comparison of model results

To compare methods, descriptive statistics for estimated county-level percent changes and their variances were calculated nationally and by region. The mean percent change in heart disease death rates represents the central tendency of percent change across all counties, and the variance of the mean percent change represents the dispersion of percent change across counties. Similarly, the median variance of estimated percent change represents the central tendency of uncertainty in county-level estimated percent change, and the interquartile range (IQR) of the variance represents the dispersion of uncertainty in estimated percent change across counties.

To compare county-level geographic patterns, we mapped estimated percent change and its variance for each method using ArcMap, version 10.1 (ESRI, Mountain View, CA).

Results

For each two-year interval, age-standardized county-level heart disease death rates were approximately normally distributed (Fig.1). The mean, variance, and median death rate consistently decreased over the study period. Of 3099 counties, joinpoint regression modeled 2597 (84%) counties with a single declining segment and an additional 366 (12%) counties with multiple declining segments.